Characteristics of Meaningful Chemistry Education

The case of water quality

Westbroek, Hanna Barbara
Characteristics of meaningful chemistry education - The case of water quality/ H.B.Westbroek. – Utrecht: CD-β Press, Centrum voor Didactiek van Wiskunde en Natuurwetenschappen, Universiteit Utrecht (CD-β Wetenschappelijke Bibliotheek, nr 52).
Proefschrift Universiteit Utrecht. Met literatuur opgave. Met samenvatting in het Nederlands.
ISBN: 90-73346-00-2

Trefw.: scheikundeonderwijs/ ontwikkelingsonderzoek/ context / need-to-know/ inbreng van leerlingen/ handelingspraktijken/ functionele inbedding/ interactie structuren.
Keywords: Chemistry education/ design research/ context/ need-to-know/ student input/practices/ functional embeddedness/ interaction structures.

Omslag en vormgeving: J.L. den Boesterd, AV-dienst, Universiteit Utrecht

CD-β Press, Utrecht

Characteristics of Meaningful Chemistry Education

The case of water quality

*

Kenmerken van betekenisvol scheikundeonderwijs

Ontwerp en evaluatie
van een lessenserie over het beoordelen van waterkwaliteit.

(met een samenvatting in het Nederlands)

Proefschrift

Ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de Rector Magnificus, Prof.dr. W.H. Gispen, ingevolge het besluit van het College voor Promoties in het openbaar te verdedigen op woensdag 21 september 2005 des namiddags te 14.30 uur

door

Hanna Barbara Westbroek
Geboren op 5 april 1967, te ‘s-Gravenhage

Promotor: Prof. Dr. A. Pilot
Co-promotores: Dr. C.W.J.M. Klaassen
Dr. Ir. A.M.W. Bulte

Centre for Science and Mathematics Education
Chemistry Department
Utrecht University
Utrecht, The Netherlands

Voor Hans
Emma en Mirjam

Table of contents

Chapter 1

Introduction

1.1 Motive for this study

This study is about the question of how to make chemistry education more meaningful to students in secondary education as it often fails to truly involve students in the learning of chemistry. This problem has long occupied and inspired teachers, curriculum developers, educational researchers and policymakers and has directed a variety of research programmes. In the Netherlands it has recently led to the initiative to reform upper secondary chemistry education, which shows that the problem is still a subject of current debate (Driessen & Meinema, 2003). With this study I hope to contribute to the discussion and to new ways to approach this problem.
In chapters 1 and 2 I will first explain what I think lies at the heart of the problem. In chapter 1, I will make a start by elaborating on my own experiences as a chemistry student and as a chemistry teacher which, I think, are exemplary for the experiences of many students and chemistry teachers and capture the problematic features of chemistry education well. In section 1.2, I will show that the problems and solution strategies sketched in section 1.1 are in fact acknowledged nationally and internationally. Apart from the idea of 'involving students in the process of learning chemistry' (on which this study focuses), two more issues play a role in the discussions. Chapter 1 ends with a brief overview of the chapters. In chapter 2 I will elaborate further on the problems and solution strategies that are sketched below, resulting in the formulation of the research question in this study.

To me, the question of how to make chemistry education more meaningful to students is rooted in my own experiences as a chemistry student. It became more and more a central question to me when I worked as a chemistry teacher. I never enjoyed chemistry that much at high school, being a somewhat strange and inaccessible subject to me in those days. Its language consisted of mysterious signs and models, where water became H_2O and could form things like 'hydrogen bonds' (which you could actually draw as a discontinuous line), and where something like 'chemical reactions' could be represented on the black board by series of letters en numbers. I remember the teacher spending lesson after lesson explaining to us how, for example, to equal chemical equations, using examples with codes like $C_6H_{12}O_6$, or H_2SO_4. To us it was like a kind of puzzle you could only solve if you stuck to the rules (never change the codes!). A lot of us didn't really grasp it. I'm sure at one time our teacher had explained why we had to do this, but I forgot.

Chemistry came to life for me, when, as a high-school student, my father showed me what life on earth looked like through the eyes of a chemist. Suddenly equating reactions became meaningful and $6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2$ were no longer just a dumb series of numbers and letters. It represented the chemistry of a process on which life depends: the process of capturing solar energy by plants, thus making it accessible to other forms of life. Intrigued, I decided not to be put off by the incomprehensibility of school chemistry. On the contrary, I was determined to get to the bottom of it. I think I was the only one in my chemistry class who wanted to continue having anything to do with chemistry.

How ironic that as a chemistry teacher, years later, I found myself spending lesson after lesson explaining to my students how to equal chemical equations as if nothing had changed. My students would ask me the same questions I struggled with, a lot of those questions essentially being: why are we doing this?

This story illustrates something that is now considered a serious problem in chemistry education: the distance between students and school chemistry. The content, character and aims of school chemistry seem to put teachers in the challenge of *really* involving students in the learning of chemistry, that is: to make chemistry education *meaningful to* students.

How did my father succeed in making the equalising of chemical equations meaningful to me? I think he just connected the abstract action of equalising representations of reactions (in this case) to a world that lies behind it. It was actually that world that appealed to me, and the equalising of equations became functional within this world. He showed me how it could help to represent a perspective of life on earth by pointing out the crucial role of the various metabolic reactions in the energy and element cycles on earth, thereby always addressing *my* questions. Of course my father could have picked other aspects of the world to connect 'the equalising of chemical equations' to, had my questions been different. The thing is, in the setting of the chemistry class, the abstract activity of equalising chemical reactions had been divorced from the aspect of the world it emerged from or plays an actual role in. In class, we had quickly drifted away from the historical context (Lavoisier), and never looked at, for example, the function of equalling chemical equations in chemical industries. For that matter, we barely covered the chemical reactions the equations represented. Despite using its formula (C_8H_{18}) in our calculations, no one had actually *smelled* or *seen* octane, or watch it burn.

The second story illustrates even more how important it is to be able to connect your actions (in this case doing and explaining an experiment) to a meaningful context. You need a reason for your actions (where does this all lead to?), in order to see the *point* of it all and make what you do meaningful. Wobbe de Vos (teacher and researcher of chemistry education at Utrecht University) taught a subject called 'orientation on the profession of chemistry teacher' for chemistry students who considered becoming a teacher. His aim was to let his students experience how difficult it is as an university chemistry student to put yourself in the position of a high school chemistry student with obviously little history in chemistry education, something you need to be able to do when teaching chemistry. The first thing he let his students do was a simple chemical experiment they all had seen and done before. They had to heat blue copper sulphate in a test tube and were to 'carefully write down their observations and explain them'. That was easy! So easy that every year students didn't see the point of doing it, but De Vos insisted. And every year all of them, without exception, explained their observations in terms of 'the evaporation of the water of crystallisation' (when I was his student it was as if I could see the water molecules leaving the heated crystal lattice of the blue copper sulphate). Of course this was a conclusion solely based on the student's previous knowledge and not on a single observation. They *knew* that the solid copper sulphate contained water molecules. They knew that this water would evaporate when the copper sulphate was heated. What had they observed? Only that some blue stuff lost its colour when heated leaving a white solid substance. Actually, none of them had even observed that *two* new substances had appeared, De Vos pointed out. Performing the experiment a second time, he demonstrated that, besides the white stuff a gas had emerged, which, when cooled down, resembled water. ('Of course, that would be the water of crystallisation' students tend to say, 'but are you sure?' de Vos asked). At this point, students started to become a bit irritated, actually considering De Vos's point (that the experiment had not contributed to their conclusions and that without their chemical background they would never have been able to 'explain' their 'observations') a trivial one. But De Vos, patient teacher as he was, proceeded quietly and undisturbed to the next experiment. Once more he wanted his students to explain what happened in this next experiment, solely using their observations. But this time he used a substance unknown to them. De Vos started heating the unknown white solid substance in a test tube at which the white stuff disappeared. At that moment a clear colourless liquid appeared. Then suddenly some bubbles emerged from the clear liquid. Finally the test tube was cooled down and the white stuff appeared again. That was it. This was magic!

In all the years De Vos was a teacher no one ever came up with a plausible explanation of what had happened in the test tube. Students didn't even agree on their observations. Wobbe refused to tell what the substance was. After a while, when students were so frustrated and irritated, getting nowhere, De Vos, would just smile and say 'well that's exactly how a lot of chemistry students probably feel when they are asked to do these kind of things at school. Like you, they often just have no sense of direction. How can they be expected to interpret an experiment? Remember that when you are a teacher.'

This second story shows how a lack of context for learning chemistry leaves students, even chemistry university students, without a sense of direction. It also shows how difficult it is for a chemist to put oneself in the position of a student with no chemical background and who doesn't possess the 'perspective and motivation of a chemist'.

Both my experiences as a chemistry student and as a chemistry teacher, of which these stories are examples, left me with the feeling that students weren't given a fair chance to appreciate what chemistry (or in a broader sense: science) could mean to them. That I wasn't the only one who considered the student's lack of interest in the learning of chemistry a more general problem of chemistry education became fully clear to me when the research project that is described in this thesis crossed my path. I jumped at the opportunity of studying how chemistry education could be made more *meaningful to* students. Obviously, I do not pretend to answer the question 'how to make chemistry education more meaningful to students' for once and for all. I merely hope to contribute to the discussion and to new approaches to this problem. In this chapter and in chapter 2, I will explain further how the concept of meaningful was narrowed down for this research project, leading to the formulation of the research question of this study.

As a start, the first story points at at least two solution strategies that could lead to meaningful chemistry education. One is connecting an abstract activity to a, for students, meaningful context in which it actually functions. The other is addressing and paying attention to the position and the questions of students. The second story illustrates explicitly how an unsuitable context can leave students with a lack direction.

In the next section I will show that the problems and solution strategies sketched above are in fact subject to national and international discussions on chemistry, (and broader: about physics and science) education.

1.2 Chemistry education discussed

The need to reconsider chemistry education and the way it is practiced is increasingly felt nationally and internationally. Next to the issue of 'how to involve students more in the process of learning chemistry, as chemistry education does this insufficiently, two more issues play a role in these discussions. The first issue is that it is felt that chemistry education does not give students an accurate and honest view of how chemistry is practiced in modern society and in scientific research and development. The second, connected, issue is that chemistry education is seen to insufficiently provide students with competencies they need as a citizen in our modern democratic society, which should lead to what is generally addressed as 'scientific literacy' (Graeber & Bolte, 1997; Laugksch, 2000; Shamos, 1995). Although in this research these last two issues will be taken into account as much as possible, the research questions will primarily focus on the issue of how to involve students in learning chemistry. The reason for this choice is that to truly involve students in the learning of chemistry and especially to secure their involvement over a longer period of time, rather than catching their attention for a short while, has proven to be difficult.

In order to show that this research touches a widely recognised problem, I will begin with briefly discussing the background of the Dutch initiative to reform upper secondary chemistry education.
Several early projects such as Salters' (chemistry, physics and science, UK), ChemCom (chemistry, USA) and PLON (physics, NL) have influenced the Dutch discussion on chemistry education (see also chapter 2). In 1996, a group of Dutch teachers, curriculum developers and chemistry education researchers were inspired by Salters' and tried to adapt this British chemistry method (see also section 2.3) for chemistry education in the Netherlands. The participants of this project aimed at addressing all three issues mentioned above: they wanted to involve students more in learning chemistry, to present them a more realistic view of chemistry in society and their daily lives and to educate them to become scientifically literate citizens. In line with the solution strategies that came to the fore in the previous section, they used 'for students relevant contexts' and 'for students relevant chemistry content in light of the contexts' as strategies to direct the design. Their aim was to produce examples of good practice. At the time this led to a number of modules (e.g. Lubeck *et al*. 1996) but the Dutch government did not further support the initiative.

Also at the chemistry education Department of Utrecht University the need was increasingly felt to contribute to the discussion and to focus research on the level of the chemistry curriculum. Up till then, the main focus had been on more or less isolated topics in the existing curriculum, with respect to students' learning problems, didactical outlining, teaching problems and so on (De Vos, 1985, Van Driel, 1990, De Jong, 1990, Goedhart, 1990, Van Keulen, 1995, Van Hoeve-Brouwer, 1996, Acampo, 1997). Inspired by these earlier research projects, the research began to focus in the first place on the analysis of problematic features of the entire curriculum. This resulted in a series of publications: De Vos & Verdonk, 1990, De Vos *et al.* 1993, Van Berkel *et al*. 2000 and De Vos & Pilot, 2001 (see also chapter 2). Recently the focus of research has shifted towards solution strategies at curriculum level. Van Aalsvoort was the first one to present a rationale for starting points of a possible new chemistry curriculum (Van Aalsvoort, 2000, see also chapter 2 and section 4.6). These starting points can be seen to include design strategies such as the use of contexts and the presentation of chemistry content on a need-to-know basis.

Recently, in the field of chemistry education, the discussion on the revision of the Dutch upper secondary chemistry curriculum was revived. In 1998 a group was formed, the 'Eenhoorn group', to initiate the analysis of the situation and start the discussion. This group consisted of chemistry teachers, teacher educators, curriculum developers and chemistry education researchers. In 1999 they produced a document in which they pointed to the isolated position school chemistry occupies with respect to society and students' daily life as the main problem, building on De Vos's (who took part in this Eenhoorn group) and Van Berkel's analyses of the chemistry curriculum (see chapter 2) (Bulte *et al*., 1999).
The Eenhoorn group proposed design directions for updating the chemistry curriculum and exemplary modules (Bulte *et al*., 2000). The two exemplary modules were designed and developed from the perspective of what the Eenhoorn group called 'Nieuwe Scheikunde' ('New Chemistry') (Jansen *et al*., 2002a, b, c, d). They also aimed at addressing all the three issues described above: education should give students insight into how chemistry functions in today's society. Learning goals should be formulated as 'competencies', which students would need as a citizen of modern society, rather then for example lists of 'need to know' goals. Moreover, they wanted to contribute to the discussion on how to involve students in the learning of chemistry in such a way that they experience the relevance of it and enjoy it.
The 'Nieuwe Scheikunde', which was developed in the exemplary modules, can be described as project-based chemistry. The designers of the Eenhoorn group used a specific model for the modules. The idea was to start with the design of 'well chosen, authentic and meaningful' thematic projects (the context) and from that they would derive relevant chemistry content (what students were expected to need-to-know). This was done to make sure that the thematic projects would appeal to students, would give them an insight into 'how chemistry functions in realistic practices' and would lead to relevant competencies. To involve students in the teaching-learning process, the chemistry content had to be *functional* to them, as a toolbox so to speak, that students would need to complete their projects.

The Eenhoorn group tried to address questions like: how to establish appropriate themes and thematic projects? What competencies are relevant? Do students enjoy it? Are they using the content of the 'basic knowledge and skills' part when conducting their thematic projects?

The group produced two modules for 16 year old pre-university students, of which 'Superslurpers', about super absorbents, is the most developed example until now (Jansen *et al*. 2002 a, b). It turned out that deriving *relevant* chemistry content (tools) from and coupling chemistry content functionally to theme projects was not easy. For one thing, the designers were restricted by the fixed chemistry examination programme and were thus not free to establish the chemistry content. This meant that students still had to learn certain chemistry content they would never use in their theme project. The trials with the modules showed that, although students enjoyed it, they had problems with relating the 'basic knowledge and skills' part to their thematic projects (Bulte *et al*., 2002). Based on their experiences, the teacher-designers produced a second version, but basically the problems remained.

In 2002 a Committee was installed by the Dutch Ministry of Education: the Committee on Renewal of Chemistry in Upper Secondary Education', chaired by Prof. Dr. G. van Koten. The Committee had the task to advise the Ministry of Education on the revision of chemistry education. They produced a report in October 2002 based on a series of interviews with teachers, students, chemistry education researchers, representatives of the chemical professional organization (KNCV), chemical industry (VNCI) and so on (Driessen & Meinema, 2003). In the report, which has the subtitle 'initiative for renewal', a few major causes of the decreasing interest of students are mentioned: the negative image of chemistry, and, in line with the above described issues: the underexposure of the importance of chemistry in today's society and the students finding it an abstract, difficult subject with little relevance. The Committee advocates a context-and-concept approach as a continuous learning line:

> Chemistry education starting from social, experimental, theoretical and professional contexts should appeal to a broad group of students. The contexts serve as a bridge between the real world and the chemical concepts that underpin the subject. The contexts serve as a framework for knowledge development in successive study years. New chemistry education should focus on gaining an understanding and interaction between contexts and concepts. This corresponds with recent international developments, for example in Germany and Great Britain.
>
> (Driessen & Meinema, 2003, p.7).

The Committee seems to address the problem of involving students as primarily a matter of appealing contexts (as these should serve as a bridge the real world and the chemical concepts, see citation above).

> Continuous renewal is an important prerequisite for a programme aimed at current and future challenges in chemistry and the associated questions from

> society. Repeatedly, new social, vocational, experimental or theoretical contexts should keep the programme updated. This is a necessary requirement in order to appeal to all students and to provide a good picture of the role of chemistry in society today.
>
> (Driessen & Meinema, 2003, p7)

In 2004/2005 groups of teachers with a coach were formed to produce exemplary materials that will be tried out at a selection of experimental schools.

As mentioned above, discussions like these are not limited to chemistry or to the Netherlands (see also chapter 2). In 1998 in the UK, Millar and Osborne produced the report 'Science 2000^{+}'. It reports the outcomes of a broad discussion on science education in the UK. They formulated the motive for this discussion as follows:

> This report is the product of a desire to provide a new vision of an education in science for our young people. It is driven by a sense of a growing disparity between the science education provided in our schools and the needs and interests of the young people who will be our future citizens.
>
> (Millar & Osborne, 1999, p.1)

In their foreword the editors point out that science education should aim at preparing young people for life in our society in the next century, rather then preparing them for being a scientist. The discussions, of which the above are examples, have on several occasions led to the development of new methods (see section 2.3.1). In Germany, for example, an innovation of upper secondary school chemistry was boosted by the low results on the international comparative results in science (PISA). This project, Chemie im Kontext, also influenced the Dutch initiatives (Parchmann *et al.* 2001, see also section 2.3.2).

To conclude, the problem of properly involving students in the learning of chemistry is considered nationally and internationally as a serious problem. Moreover, to really involve students in a teaching-learning process in secondary chemistry education that is meaningful to them and to secure their involvement throughout the process seems to be the difficult part (Lijnse, 2001, see also chapter 2). This research will therefore primarily focus on this particular problem, although I consider it as equally important to give students an accurate view of how chemistry is practiced in modern society and to sufficiently provide students with competencies they need as a citizen in our modern democratic society.

1.3 Thesis outline

The central question of this explorative study is how to make chemistry education more meaningful to students in secondary education. In this chapter I have sketched the problematic features of chemistry education and some possible directions for solutions that proved to be widely recognised, nationally and internationally. Basically, these directions, contexts and attention for students' questions (a need-to-know), imply that students need a sense of direction and attention for their questions in order to get involved. In chapter 2, I will elaborate on the problems and solution strategies sketched here. This results in the formulation of the research question for this study. The main point of chapter 2 is that, if the aim is to study meaningful chemistry education in the sense suggested above, the design and development of an exemplary teaching-learning process should be object of study in a process called developmental research. In chapter 3 this research strategy and accompanying methodological issues are discussed in detail. Next, in chapter 4, the first two research cycles with the first two versions of the design central to this study are described. In this chapter it is described how design decisions for a next version of the design are based on empirical evidence and theoretical considerations, resulting in the third version of the design. In chapter 5 the design of the third version, including the grounded expectations about the learning processes of the students, are presented in detail. Chapter 6 discusses the setting of the third research cycle with this third version and how the data was collected and analysed. This includes an account of how the teacher was prepared for his role. In chapter 7 the evaluation of the third version of the design is presented in detail. I analyse how and to what extent meaningful chemistry education was realised in this case. Finally, in chapter 8 the research question is answered. Conclusions are drawn about design directions for meaningful chemistry education at the level of one module and how this might lead to hypotheses at the level of the curriculum.

Chapter 2

Characteristics of meaningful chemistry education

2.1 Introduction

In chapter 1, two stories illustrated how chemistry education can alienate students and give them the impression of meaningless chemistry education. They also suggested how learning chemistry *meaningfully* is connected to both being able to identify with the questions raised and, partly related to this, to having a sense of direction of 'where it all leads to'. The stories pointed towards two potential strategies to bring this about. The first one was the use of contexts that are relevant and recognisable for students in order to provide a setting in which they experience the actual function of the abstract chemical actions involved. The second one was attention for the students' own position and questions. It was also shown that both the problem of properly involving students in the learning of chemistry and the sketched solutions strategies are actually nationally and internationally acknowledged in discussions on chemistry (and science) education. Besides two other but related issues that play a role in these discussions, the issue of really involving students in a teaching-learning process in secondary chemistry education that is meaningful to them and to secure their involvement throughout the process seems to be the difficult part.

In chapter 2, I will elaborate on the problems outlined and strategies to solve them. I will first describe different analyses of chemistry and science education. These analyses reveal three main problematic features of chemistry education that contribute to the alienation and lack of sense of direction of students: 'rhetoric of conclusions', 'incoherent sequencing' and 'lack of student input'. They play a role on the curriculum level, at the level of one module, and at the level of one lesson (section 2.2).
Next I will describe in section 2.3 how several projects in the field of chemistry, physics and science education all embody three intertwined characteristics of what I have called 'characteristics of meaningful': *context*, *need-to–know* and *attention for student input*. They are interpreted as key elements of solution strategies to address the problematic features described in section 2.2. This framework of solution strategies is considered to provide for useful design directions for meaningful chemistry, and broader: physics and science education. Nevertheless, none of these projects addresses the question if, how and to what extent the applied solution strategies actually work as intended and result in meaningful chemistry, science or physics education. In other words: the teaching-learning process and the quality with which these characteristics are given shape in the design are not made object of study. I therefore conclude that this question about the design is a neglected and underestimated field of research. Next, in section 2.4 it is argued that if the design is made object of study, a specific research strategy called developmental research is needed to address the question if, how and to what extent it gives rise to meaningful chemistry education. Finally, the research question of this study is formulated in this perspective.

2.2 Problematic features of current chemistry education

In this section I will elaborate on three main problematic features of school chemistry (and broader: school science) which each contributes to the alienation and the lack of sense of direction of students. This results in the conclusion that if chemistry education is to be made more meaningful to students, these problematic features have to be addressed.

2.2.1 Rhetoric of conclusions

De Vos *et al*. (1990, 2001) describe the first problematic feature of chemistry education as chemistry education being a presentation of selected results of chemical research. A success story, which leaves the process of how people came to these results (or worse: the fact that *people* came to these results), and why they bothered, unanswered. Schwab already referred to this phenomenon as 'rhetoric of conclusions' (Schwab, 1962). At the present time school chemistry is deprived of motives of the people involved. It is wholly deprived of the emotional side of the practice: the excitement, frustration, fascination and adventure. It is hardly surprising that this leads to alienation and a lack of sense of direction of students. In school chemistry, where ideas, models and theories are introduced 'as such' and without a clear purpose, it is impossible for students to have even a remote idea of 'what comes next'. That automatically leaves the chemistry teacher completely in charge of providing the motives for the content-related progression. S/he is the only one with an overview. Wandersee (1994) showed how this leads to the creation of so-called 'misconceptions' among students. This is not surprising. If students do not have a content-related sense of purpose, they will create their own purpose for learning the concepts involved and as a consequence there will be considerable room for misconceptions.

What happened with De Vos' students in the second anecdote of chapter 1 could also be explained from this perspective. As Joling *et al.* (1988) pointed out scientists' observations are always guided by hypotheses, which emerge from a specific theoretical context. This provides them a specific purpose for experimentation and observation. In school chemistry, students have no clear view of these theoretical contexts. As a result they often don't even know which observations are relevant; see also Hart (2000).

As I mentioned, this problematic feature is not limited to chemistry education, it is a widely recognised, shared problem of chemistry, physics and science education (Bencze & Hodson, 1999; Fensham, 1999; Millar & Osborne, 1999; Rennie *et al*., 2001). Osborne & Collins (2000) for example analysed how the 'growing gulf between science-as-it-is-practised (e.g. 'making observations' is guided by hypotheses that are embedded in theoretical contexts) and science-as-it-is-being-taught-in-schools' (rhetoric of conclusions) alienates students. They found that this growing gulf could result in students expressing their antipathy to topics such the periodic table. Students experienced learning difficulties with this topic and failed to see its relevance:

> Edward: It [the periodic table] doesn't mean anything to me. I'm never going to use that. It's never going to come into anything; it's just boring.
>
> (Osborne & Collins, 2000 p.17)

2.2.2 Incoherent sequencing

The second problematic feature of school chemistry is the incoherent sequencing of the content. Two related types of incoherent sequencing have been described: at the level of concepts and at the level of curriculum emphases. Firstly, De Vos *et al.* (2001) describe a 'layer-like structure' as a specific problem of school chemistry, with implicit shifts between different meanings of the same words. This makes it very difficult for students to understand and correctly use different concepts and relations. Secondly, Roberts describes incoherent sequencing at the level of 'implicit shifts between different curriculum emphases', which lead to incoherent messages about the purpose of learning certain content (Roberts, 1982; 1988; 1995). This type of inconsistency concerns science education and physics education as well as chemistry education.

Incoherent sequencing at concept level

De Vos & Verdonk (1990) and De Vos & Pilot (2001) argued that the modernisations of school chemistry content in the 20th century, being successive and separate decisions, have basically led to the addition rather than the integration of new topics. This way, every modernisation created its own new 'layer'.

De Vos and Pilot analysed chapters in textbooks and recognised six different 'layers' for the topic of acids and bases: a craft, synthesis, analytical, Arrhenius, Brønsted and application layer (De Vos & Pilot, 2001). They showed that the definition of acid is not consistent throughout all the layers. How these different definitions of acid relate to each other and maybe emerged from each other is not made explicit in school chemistry books. Moreover, such incoherent shifts between layers and (as a result) between definitions appear at curriculum level, within one chapter and even within a learning task.

Not only De Vos and colleagues acknowledged this problematic feature. Kaper and Ten Voorde referred to this phenomenon as 'stealthy [Dutch: sluipend] context changes' (Kaper & Ten Voorde, 1991; Kaper 1997) and Joling speaks of 'changing contexts' [Dutch: wisselende contexten] (Joling, 1993).

De Vos & Pilot (2001) argue that this incoherent sequencing of the chemistry content at the level of concepts must lead to learning problems.

Incoherent sequencing at the level of curriculum emphases

Roberts' concept of curriculum emphasis showed a different, but overlapping, level of incoherence in school chemistry. He developed this concept analysing North American science textbooks (1982, 1988, 1995), but the concept can be applied to chemistry education as well (Van Berkel *et al.*, 2000). According to Roberts a curriculum emphasis gives rise to a coherent set of meta-messages about the *purpose* of learning certain topics.

> Students not only learn a subject, they also learn some fringe-benefit lessons at the same time- including the reason the teacher or professor (or curriculum policy maker) believes students should learn the material.
>
> (Roberts, 1988, p.32)

The curriculum emphasis reveals itself in the way the subject is the taught by the teacher, in the materials and methods and in the manner of assessment. Roberts distinguished emphases like: ‘Solid Foundation’, ‘Self as Explainer’, ‘Science/Technology Decisions’, ‘and Correct Explanations’. Fensham (1999) gives examples of how different projects were centred on different emphases:

> For example, Personal Coping is present in the rationale for the sequence of the Salters’ Chemistry project, Science/Technology Decisions was central in Logical Reasoning in Science and Technology, in a secondary text used in some central Canadian provinces and in some PLON units. Self as Explainer was quite explicit in The Science Framework in Victoria, Australia and in the Science Plus materials from the Atlantic science project in eastern Canada.
>
> (Fensham, 1999, p.5)

Roberts argued that different emphases have to be developed *consistently* in *separate* units and not as isolated pieces of information, if students are to become aware and confident that their learning of science does have a coherent and meaningful purpose. Thus, the choice of an emphasis should influence the choice of content, as they have to be consistent with each other (Roberts 1988, Fensham, 1999).

> …the subject matter topics in a unit have to flow logically, of course, but so does the emphasis.
>
> (Roberts, 1982, p.251)

How does *incoherent* sequencing at this level occur and how does this influence the way students perceive school chemistry? The following typical example shows how implicit shifts between different emphases in (in this case) a textbook produce confusing messages to the students about ‘Why are we learning this?’. Try to put yourself in the position of a student and we will go through the first paragraph of the chapter where ‘acidity’ is introduced in a widely used Dutch chemistry method for lower level secondary education (Pieren *et al*., 1995). It starts with a story about the way different foods taste different (sour, bitter, salty). Next, students are to *test* different daily life liquids like ammonia, vinegar and lemon juice. They use something like an ‘indicator’, and are supposed to categorise the tested fluids into the categories ‘acid’, ‘neutral’ and ‘basic’ based on their test results. Immediately after these experiments the concept of acidity (pH) is introduced, as a refinement of the categories. The reason for this refinement and what it has to do with different tastes is not made explicit for students. This is the point were you, a student, might ask ‘Why am I doing this?’

In terms of curriculum emphases this example can bee seen as an implicit shift between the emphases ‘Everyday Coping’ (taste different foods) and ‘Solid Foundation’ (categorise solutions using pH).

2.2.3 Lack of attention for student input

Besides the features of the school chemistry content, there is another, but related, problematic feature of chemistry education, which contributes to the lack of sense of direction of students and their alienation from school chemistry. The limited attention for student input in the teaching-learning process is also an, again widely, recognised problem in physics and science education

In a large - scale study that investigated the quality of science teaching and learning in Australian schools, for example, secondary students were asked about 'student input'. The authors report that:

> For many secondary students, the teaching-learning process is teacher-directed and lessons are of two main types: practical activities where students follow the directions of the teacher to complete an experiment, and the chalk and talk lesson in which learning is centred on teacher explanation, copying notes and working from an expository text.
>
> (Rennie *et al*., 2001, p.475)

What results in:

> Forty percent of the students indicate they never or only sometimes have to think about what they are doing in science.
>
> (Rennie *et al*., 2001, p.474)

Studies that address the role of students' and teacher's questions (to students) indicate that students need room to ask their own questions and they need to be listened to carefully by the teacher. Such a student-centred approach provides them with a sense of direction and the feeling that their input matters (Lemke, 1990, Rop, 2003, Watts *et al*., 1997; Zee *et al*., 2001). Moreover, Watts *et al*. (1997) and Marbach-Ad & Sokolove (2000) also found how students' questions could be indicators of their interest, conceptual engagement and conceptual change.

Lemke (1990) studied this problematic feature in detail. He observed and analysed classroom dialogue in traditional science lessons. He found how these are dominated by whole class interaction and the dialogue structure he called 'triadic dialogue structure'. Basically this structure shows the following pattern: teacher-question student-answer teacher-evaluation. The triadic dialogue pattern prevents students from asking questions that could indicate self-directed learning:

> The teacher begins each sequence by asking a question that either targets the whole class or singles out one particular student hoping that student will provide a good answer. Sometimes, as Lemke explains, a student breaks the discursive sequence by asking a related or tangential question and interrupts the smooth flow of the lesson. These student questions stimulate certain teacher reactions that Lemke suggests relate to maintaining the classroom control. However, these questions may represent student curiosity and self-directed learning. They may represent an interested student's efforts to conceptualise a topic or make it relevant to life.
>
> (described by Rop, 2003, p.14)

Obviously this problematic feature exceeds chemistry, physics and science education. It might very well also play a role in history education, or modern languages. However, the feature can be expected to be more acute in science subjects as these subjects in particular consist of for students' inaccessible and abstract bodies of knowledge, leaving the teacher largely in charge of the content-related progression.

2.2.4 Concluding remarks

Where do these problematic features originate? De Vos *et al*. (1993, 2001) argued that the first two characteristics 'rhetoric of conclusions' and 'incoherent sequencing' are a result of two different but connected forces, which from the beginning played a role in the evolution of school chemistry. First, there is the desire to create an overview curriculum, which should give students an adequate impression of the state-of-the-art in chemistry, as they are to be our future chemists. Second, there is the aim to avoid an overloaded curriculum, but that proved to be impossible within the decisions based on the criterion 'overview curriculum'. These two forces mostly drove decisions about the school chemistry content. They were separate and successive decisions.

The third feature (the lack of attention for the input of students) is largely the result of the feature 'rhetoric of conclusions'. This promotes rather than prevents a teacher centred teaching-learning process: only the teacher has an overview of 'what comes next' and is therefore in charge of the content-related progress, leaving little room or rather opportunity for student input.
But also the overloaded curriculum has contributed to this third feature. Teachers do not encourage student input, because of time pressure.

> Teacher focus group discussions confirmed the heavy content burden at secondary level with teachers and students rushing superficially through the content so that it is covered for "the test".
>
> (Rennie *et al*., 2001, p.474)

Newton *et al*. 1999 confirm these findings. They found that teachers feel the pressure to cover the content, and as a result they cannot afford to 'waste time' on so called meaningful discussions, thereby restricting student intellectual involvement.

The question 'How can chemistry education become more meaningful?', posed in chapter 1, might now be interpreted as: 'What will contribute to the diminishment of these three problematic features?'. Before describing the research strategy followed in this study, I will first summarise how others have attempted to address this question and discuss their results.

2.3 Potential solutions and characteristics of meaningful

There have been several attempts to diminish the problematic features of chemistry (and science) education described in section 2.2 by designing more 'meaningful' lessons (e.g. Bennett & Holman, 2002; Campbell *et al*., 1994; ChemCom, 2002; Eijkelhof & Wiersma 1986; Kraijcek *et al.*1999; Kresner *et al*., 1997; Rivet *et al*., 2000; Rivet, 2003; Roth 1998, 2003; Schwartz, 1999; Van Aalsvoort 2000). They range from isolated projects, designed as small-scale experiments or welcome additions to the existing curriculum, to projects with the aim to reform the curriculum. Although not explicitly labelled as such, in both the small-scale and large-scale projects three intertwined 'characteristics of meaningful chemistry education' can generally be seen to play a role. In section 2.3.1 I will give a brief tentative description of these three characteristics and how they can be interpreted as solution strategies for the problematic features that were described in section 2.2. In section 2.3.2 I will discuss six projects and indicate in what sense they all can be seen as using these three characteristics of meaningful in their design and teaching. I will also summarise what kind of evaluation studies were carried out with regard to these projects and what the main findings were. This discussion is based on what was reported on the projects in the literature. In section 2.3.4 I will summarise the findings of this section. I conclude that a thorough underpinning and evaluation of the designed teaching-learning process in terms of the three characteristics is generally lacking in the discussed projects, although they all more or less aim to address the same problems and can all be seen to embody the three characteristics. None of them evaluate whether the design lived up to its aims. This points to a neglected research area, on which I will elaborate in section 2.4.

2.3.1 A first outline of the three characteristics of meaningful

Before discussing the different projects, I will briefly describe what I consider to be three characteristics of meaningful. This description is based on a general analysis of the literature about a selection of projects (see section 2.3.2). At this point I will only very tentatively indicate how the characteristics can be considered a solution strategy for the problems described in section 2.2.

First characteristic: context

A well defined and for students recognisable *context* for concepts provides the use of these concepts with a distinct function, and thereby makes students' use of the concepts meaningful and motivating.

The 'rhetoric of conclusions' feature can be avoided when the emphasis is shifted from 'getting an overview of the conceptual products of chemistry' to the 'functional use of concepts in relation to a certain relevant, recognisable *context*'. A *consistent development of concepts* can be achieved in this way (as they are to be used with a 'distinct function and meaning').

Second characteristic: need-to-know

Addressing students' questions on a need to know basis, which also implies properly building on their existing knowledge, provides for an increasing involvement of students in the teaching-learning process as they will see the point of what they learn every step of the way.

This characteristic, together with the first characteristic of a well-defined *context*, can provide for the development of a *coherent emphasis*. The question 'why are we learning this?' is captured in the need-to-know approach. The first and second characteristic both contribute to the diminishment of the incoherence.

Third characteristic: attention for student input

The third characteristic is closely related to the second characteristic: if one aims at *really* incorporating a need-to-know approach in the design of a teaching-learning process, then 'real attention for student input' is inevitable. In a successful *need-to-know* approach students have more insight into and experience the functionality of 'what comes next'. Thus the teacher has more opportunity to pay real attention to their input, which now could become a driving force of the content-related progression. Consequently, students will *feel that their input matters*.

Obviously this characteristic addresses the problem of 'lack of students input'.

2.3.2 The three characteristics of meaningful in different projects

I will describe in broad outlines six different projects that were aiming at a diminishing of the problematic features of science education and that I selected as well known, well documented, striking examples of many more projects. I will show that all six can be seen as embodying the three characteristics of meaningful described in section 2.3.1. I realise that in discussing these projects from this sole perspective I leave out all other perspectives and in that sense do not do justice to the scope and depth of these projects. The people involved were asked to approve of the descriptions.
In their reports, the projects all aim to address problems similar to the ones I described in section 2.2 (although some projects put different accents). The projects described here are selected because they are, in general, well known projects. They are referred to frequently, and are seen as representative examples of efforts to create more meaningful chemistry, science or physics education, at different times, in different places. This shows that the problems and solutions are generally accepted, but not new. For each project I will briefly go into its rationale and how the three characteristics can be considered as forming the main outlines of the design and how they are accounted for. Finally, I will briefly describe how the projects were evaluated, especially concerning the three characteristics of meaningful. As I mentioned before, the teaching-learning process is hardly ever evaluated in detail. In the text I will indicate between brackets which characteristic I recognise in the accounts.

PLON: Physics Curriculum Development Project

PLON refers to a Dutch project that developed context-based curricula for secondary physics education (Kortland, 2002). This large-scale project ran from 1972 to 1986. It embodied and gave rise to ideas on meaningful physics education that seem to be still very much alive and in line with the three characteristics of meaningful.

Rationale

The motive behind PLON was to make physics education more attractive to students and to shift its emphasis towards how physics functions in daily life situations (technology) and in explaining well-known phenomena. The people involved in PLON refer to the same type of problematic features of physics education as discussed in chapter 1 and section 2.2 with respect to chemistry education: through school books and examination programmes students are presented with a picture of an old fashioned, estranged and almighty science. Almighty in the sense that for every single problem there is one neat and correct answer that solves the problem (Kortland and Van der Loo referring to Wijker, 1986, p.66-67). Considering that most students did not continue their physics study after secondary school, the idea was that physics education should contribute to 'personal development' of *all* students in terms of a growing awareness of the cultural importance of science and an increasing ability to 'think scientifically' (Kortland, 2002, p2). This resulted in a broadening of content and skills to be covered. The traditional content was to be broadened with knowledge about science/technology related social issues and the traditional skills with skills like issue related decision-making. The aim was to educate students in dealing with daily life situations that involved technology and in understanding natural phenomena (Eijkelhof & Kortland, 1988). The expectation was that the promotion of activity-based teaching and learning in relevant life world contexts would make students experience the content taught as more relevant and useful (Lijnse *et al.* 1990, Lijnse 1995).

The three characteristics in the design

On a unit level PLON used the following design features: each unit was to start with an orientation on a theme, introducing a basic question taken from the society students live in, and regarded as relevant to them with respect to their (future) life roles as a consumer and member of society [characteristic 1]. Such a central question concerns a daily life situation or controversial societal situation in which a decision must be made. The question defines the boundaries of the theme and can be used as a selection criterion for what would be relevant physics content and skills for students to learn. 'How does the use of a helmet and of safety belts contribute to safety in traffic?' was one of the central questions used in PLON (Kortland & Van der Loo, 1986, p73). The next part was to address the physics which was expected to be relevant for answering the basic question [characteristic 2]. This way the 'basic question' was supposed to help the designers to select content. It was supposed to help to avoid the abundance of aspects and to strengthen the coherence of a unit. After students had addressed the relevant physics, this part was followed by a number of options in which groups of students could independently do some further work on aspects encountered in the unit's previous part [characteristic 2], and report their findings to other groups in class [characteristic 3]. Finally, the basic question was addressed [characteristic 1] and the students were to apply what they learned in order to answer the question. The students were to reflect on their learning process with questions like: does the physics taught help in finding answers? [characteristic 2]. Because students were working quite a lot in small groups, the teachers were expected to be less involved in frontal classroom teaching and more in giving guidance to these small groups of students. During the reporting sessions students were to completely overtake the teaching role of the teacher, with the teacher playing the part of the observer with the task of giving adequate feedback on their presentations. This was expected to contribute to a diminishment of the distance between teacher and students, and can be interpreted as promoting the importance of and attention for student input and therefore their feeling that their input matters [characteristic 3].

Evaluation

PLON was a large-scale project. The first version of units was mainly evaluated in a general sense, to eliminate 'infant diseases', using instruments like evaluative meetings with teachers, questionnaires for students and very few (because time consuming) observations at school. One outcome was that students appreciated the selected topics, the student activities and the applied character (Kortland, 2002).
The evaluation of the second version of the materials was more directed at the aims of the project. The impact on students' learning, in terms of learning outcomes and learning difficulties but also changes in their attitudes towards physics learning as a result of PLON-units were studied for example, mostly using questionnaires (Wiersma, 1984; Eijkelhof & Wiersma, 1986; Eijkelhof & Kortland, 1988). These studies showed for example that students' cognitive outcomes (including learning difficulties) did not differ from the traditional learning outcomes. They also showed that it is difficult to have students use their acquired conceptual science knowledge in decision-making situations related to applications such as ionising radiation, especially in those situations in which students (might) have already formed an opinion. Apparently students did not need to know the physics involved and based their decisions on other arguments. How this related to the outlining of the teaching-learning process was not an object of research back then. In his reflection on PLON, Lijnse argues that, for one thing, PLON revealed the importance of further research to make abstract ideas such as 'physical knowledge should function in daily life situations' concrete and testable in order to improve physics education (Lijnse, 1986, p308-321).

According to Kortland (2004), students themselves did not recognize the need-to-know approach in PLON. The modules were therefore less motivating and effective for students than they could have been. External pressure also played a role, a lot of content was still selected for the sake of tradition and not because students 'needed to know it'.

ChemCom: Chemistry in the Community

The American Chemical Society developed ChemCom in the early 1980's as a response to criticisms on the nature of science, physics and chemistry education in the US. These criticisms were in line with the problematic characteristics described in 2.2: traditional chemistry courses were said to present science concepts as a series of unconnected facts and to leave out the real-world connections entirely (ChemCom, 2002).

Rationale

ChemCom is in more than one way similar to the PLON project. Like PLON, ChemCom made a shift from an emphasis on 'solid foundation' towards an emphasis on the application of scientific knowledge, everyday life and the social and environmental implications of scientific and technological developments. The aim of the course was shifted accordingly from more traditional aims to "the preparation of students for informed, effective citizenship through stimulating and engaging in decision-making activities" (ChemCom, 2002, p.1). Similar to PLON, ChemCom is characterised by the authors as "A student–centred, activity-based, issues-oriented chemistry course that encourages small-group learning" (ChemCom, 2002, p.1). It was expected that ChemCom would lead to students experiencing the relevance of the chemistry involved.

The three characteristics in the design

ChemCom units are structured as follows. Each unit starts with a social/technological issue: a problem or challenge for the students to deal with [characteristic 1]. Students study topics that the designers expected to relate to this issue on a need to know basis: the societal problems require certain chemistry knowledge, and students, having to study the problems, were thought to subsequently recognise the need to understand the relevant chemistry. The expectation was that 'students experience the use and application of their chemistry learning, leading to a greater sense of motivation and a feeling of 'ownership' of their new knowledge' (ChemCom, 2002, p.3) [characteristic 2]. The course promotes class discussions and co-operative learning techniques as most of the problem solving is expected to be done in small groups in which the 'sharing of students thought and experiences helps shape the learning process" (ChemCom, 2002, p.8). As I interpret it, this would require 'more attention for the input of students' [characteristic 3].

Evaluation

Like PLON, the evaluation of ChemCom did not involve in depth studies of the teaching-learning process and how the design had contributed to the realisation of the three characteristics. Sutman and Bruce reported a five-year evaluation of ChemCom (Sutman & Bruce, 1992). One of their conclusions was that students responded very positively to the context-based chemistry materials, willing to get engaged (Bennett & Holman, 2002). Further evaluation consisted mainly of feedback from teachers, which led to revision. Questions such as the following were not object of research: Have students really experienced the relevance of the chemistry involved as a result of a proper worked out need-to know characteristic? And as a result, did they use the chemistry in their decision-making? Did the design lead to the students sharing their thought and experiences? And did this help shape the teaching-learning process?

The Salters' approach

A British initiative in 1983 to discuss how science, physics and chemistry could be made more 'attractive to students' has resulted, twenty years later, in a 'whole family of science, physics and chemistry courses' (Campbell, 1994).

Rationale

The idea was that a new kind of school science, more appealing to ordinary students, was called for. It needed to be "more relevant to students' interests and everyday lives, and would involve them in a wide range of activities in which they could actively engage" (p.6). The aim was that "the ideas and concepts selected, and the contexts within which they are studied, should enhance young people's appreciation of how

chemistry contributes to their lives or helps them to acquire a better understanding of the natural environment” (Campbell, 1994, p.418-419).
The Salters’ approach was designed from the concept of ‘narrative cognition’ (Bruner, 1962).

The three characteristics in the design

The Salter’s approach uses the idea of story lines as contexts. The story lines are to motivate and involve students in the learning of chemistry and to provide students with a more authentic picture of it (Campbell *et al*., 1994) [characteristic 1].
The need-to-know approach was addressed as follows: the unfolding of the story line would, in a natural way, provide for motives at certain points to get a deeper insight in scientific ideas and concepts. Or at least, the continuing story would provide for a context in which the use of the concepts and ideas has a clear role and function (and therefore is meaningful). In other words, students would experience a need for relevant scientific concepts and ideas to be able to ‘follow’ the story line, or would at least experience the functionality of the concepts and ideas in the light of the story line [characteristic 2]:

> Units of the course should start with aspects of the students’ lives, which they have experienced either personally or via the media, and should introduce ideas and concepts only as they are needed.
>
> (Campbell *et al*., 1994, p.419)

Moreover, the developers were well aware of the fact that science teaching was still ‘heavily relying on the transmission of content knowledge’. To address this problem, they decided that the course should include a wide range of learning activities, which would actively involve students [characterisitic 3]. They based their choice on the work of Barnes, Davies & Greene, Lemke and Sutton, who emphasise the importance of the use of language by learners in the learning of science (see also section 2.2: lack of student input). Many student activities were included to promote the use of student discussion in lessons (Campbell, 1994, p419). Bennett and Holman point out that the adoption of active-learning approaches is an important feature of the Salters’ approach. Students are more actively involved in activities such as discussions, presentations, decision-making than in traditional courses (Bennett & Holman, 2002). This can be interpreted as contributing to the third characteristic.

Evaluation

Evaluation studies of the project cover a variety of fields ranging from ‘comparing the understanding of certain chemical concepts of students using a context approach with students using a traditional approach’ (Ramsden, 1997; Barker & Millar, 1996; Barker, 1999) to ‘teachers’ change as a result of teaching a context-based course’ (Ramsden, 1994). It was not studied, though, in any detail how and to what extent the design of the teaching-learning process had contributed to the three characteristics of meaningful: had students experienced a need to know with the unfolding of the stories? And did they, as a result, experience the chemical content as relevant and useful?

Chemie im Kontext

Chemie im Kontext (ChiK) is a German project that was initiated in 1999 by the universities of Dortmund, Oldenburg and Saarbrücken and the Leibniz-Institute for Science Education in Kiel (Eilks *et al*., 2004). The project was influenced by projects like ChemCom and Salters' Advanced Chemistry (Nentwig *et al*., 2002; Eilks *et al*., 2004).

Rationale

The motive for this initiative was created by specific studies on German students' performances in science education and their attitudes towards science education. They showed that German students are very much like other students: they have little interest in learning science, their attitude towards learning science when doing it is at best 'ambivalent' and finally, students in Germany are better at reproducing facts than applying knowledge outside the classroom situation (Parchmann *et al*., 2001). The people involved in ChiK concluded that certain problematic features of the German science (and chemistry) education embodied the problem. It needed "more emphasis on a limited number of science concepts for the enlightenment of issues that are relevant in the learner's world". Moreover, the aim of science and chemistry education should be that learners learn to "use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity" (p.3).
ChiK is presented as a 'framework for teachers', which they can adapt to their specific practice, rather than a detailed prescription. The authors describe ChiK as being 'context oriented'. ChiK has a slightly different approach than the other projects described here in the sense that it pays explicit attention to what is seen as a knowledge transfer problem: 'situated learning' leads to 'situated knowledge and competencies'. The aim of ChiK therefore is to develop among students "de-contextualised basic scientific concepts which students can apply in different situations relevant to them". In spite of the slightly different emphasis, this has led to a rather similar design model to, for example, PLON and ChemCom.

The three characteristics in the design

Each unit starts with a question or an issue that is considered relevant for students [characteristic 1]. Next, students elaborate on the chemical knowledge they are expected to need in order to address the questions or issues raised. This way the learning processes are "embedded into authentic problems". The idea is that learners acquire knowledge and competence in dealing with an issue of interest and relevance. Therefore, learning activities aim at "enriching knowledge and competence of the learners, so that the initial problems can be solved more efficiently and to the learners satisfaction" [characteristic 2]. The problem of creating situated knowledge rather than de-contextualised knowledge is addressed with the use of 'multiple contexts'. That means that elaboration (on a need to know basis) is done in a sequence of different themes in which the knowledge is constructed and reflected upon (de-contextualized). To increase student input, the teaching methodology promotes self-directed and co-operative forms of work [characteristic 3].

Evaluation

Evaluation of ChiK focussed until now for a large extent on teachers and the implementation process (Eilks *et al*., 2004; Parchmann *et al.,* 2001). Some evaluations involved studies on effects on motivation of students, changes in cognitive learning outcomes and the changes of the learners' perception of the quality of instruction. Most of these evaluations were large-scale studies in which questionnaires and pre- and post-tests were used to measure effects. Again, a detailed study of the realised teaching-learning process in relation to the designed teaching-learning process was not an object of research so far. The studies showed for example 'significant positive effects for the perception of relevance of chemical contents and quality of instruction', but not how this related to specific characteristics of the design, which is implemented very differently in the modules being described.

Chemie in producten (chemistry in products)

This small-scale project evolved from the research of Van Aalsvoort and her colleagues. She designed a chemistry course for 14-15 year old students, which is now used at a growing number of schools in the Netherlands.

Rationale

The course is meant to be an exemplary solution for the problems described in section 2.2. It aims to make a shift from 'science orientedness (which led to the problematic characteristic 'rhetoric of conclusions')' to 'activity oriented'. The aim was that students should explore the different roles that people fulfil in real life practices (Van Aalsvoort, 2000).

The three characteristics in the design

Inspired by the ideas of Leont'ev, Van Aalsvoort interpreted context as being an activity, derived from a recognisable and relevant practice and with different roles connected to it [characteristic 1] (Van Aalsvoort, 2000). The units were designed around practices, in which products are produced, such as: 'producing and selling apple compote' and 'producing drinking water'. The aim of such a practice, for example in the case of 'producing and selling apple compote' the aim would be to produce apple compote for a consumer market, helps to establish what roles are involved and how these roles contribute to the aim of the practice. Learners are expected to adopt and identify with the roles of the practice as much as possible. The expectation is that when learners identify with such roles, they will adopt the 'motives for action' connected to these roles. Van Aalsvoort presents the following example: producers have to find solutions for production problems, such as: how can we produce apple compote, which will keep for a while, and doesn't turn brown? As students identify with the role of producer, they will adopt the motive to find out: 'How can we produce apple compote that will keep for a while and doesn't turn brown?. This, according to Van Aalsvoort, defines what chemistry content students will then feel that they need to know [characteristic 2].

Both the practice and the roles students fulfil are expected provide them with a clear sense of direction (purposeness) and involvement, because the teaching-learning process depends in part on their contributions as players of roles in the practice. In this sense, this feature can be interpreted as promoting that attention for student input [characteristic 3].

Evaluation

Van Aalsvoort presents a thorough rationale for the designed teaching-learning process. The design itself, however, is not evaluated in detail. Questions such as: 'Was the process of identification of students with their intended roles realised as intended and did it give rise to intended motives for learning chemistry content?', were not addressed. In that sense, how and to what extent, for example, the design gave rise to the general motivation of students, a sense of purpose and a 'need-to-know-experience' among students, was not object of research (Van Aalsvoort, 2000, see also section 4.6: 'The roles of the students in the instructional version of the practice').

The Evolution of Water and the Helmet project

The Evolution of Water project is part of a research programme that studies contextualized instruction through project–based science and aims "to help students meet content standards while also making meaning of science through inquiry" (Rivet, 2003, p1; Kraijcek *et al*., 1999). Rivet reports her findings with a similar project that is called 'the Helmet project', a project that is also part of this research programme (Rivet, 2003).

Rationale

The aim of 'contextualizing instruction' is that through inquiry students engage in investigations from which they learn scientific processes and how these work to generate new information. It is thought that this approach stimulates the 'active involvement' of students in the process of knowledge development and provides room and opportunity for students' own interests and questions (Rivet, 2003). This could be interpreted as an emphasis shift from 'rhetoric of conclusions' to a more process oriented emphasis such as 'science as a human activity'.

In her thesis, Rivet defines features of 'contextualized instruction' that she expects will contribute to the above described aim (Rivet, 2003, p.3, 29 and 30), such as the following:

A. The use of for students' meaningful problems and situations, in the sense that they have implications for them outside of school.
B. A meaningful problem provides a need-to-know situation to learn specific scientific ideas and concepts.

> Although this [*need-to-know*] is similar to the role provided by application problems commonly used at the end of an instructional unit, contextualizing instruction differs in that it drives the need for understanding and thus the instruction, rather than simply as a way to integrate and apply knowledge after it has been gained in a more abstract setting.
>
> (Rivet, 2003, p.30)

Rivet explains how she thinks that project-based learning will contribute to a need to know:

> These instructional models are similar in that there is a single overarching problem or setting that drives the instruction and provides a purpose for learning.
> (Rivet, 2003, p.22)

Within this framework Rivet refers to concepts such as "learning needs" (p23) and "reasons [for students] to engage in learning activities" (p24). From this I get the impression that Rivet's aim is for students to experience this need *before* getting involved in the respective activity in which such a need is satisfied, and not afterwards.

The three characteristics in the design

Rivet presents the following design features that could be interpreted in terms of the three characteristics:

1. A driving question to introduce and organise the contextualizing theme of the project [characteristic 1]. Sub-questions are used to structure the driving question into more manageable chunks, and help to maintain the connection between the individual ideas presented in the unit and the overall context of the project [characteristic 2]. To motivate students to both take ownership of the problems and thoughtfully pursue a solution, the driving question is designed to encompass real-world problems that students find meaningful and that relate to students' prior knowledge and experiences (Rivet, 2003, p34-36).
2. Project activities linked and woven in with the driving question and contextualizing theme. The activities, investigations and discussions in the project relate back to the issues and themes introduced through the driving question and anchoring events, emphasising the relevance of what is being learned to understanding, and then developing solutions to, the driving question. They should accurately reflect content ideas [characteristic 2].

According to Rivet, the driving question is expected to 'contextualise scientific ideas using students' experiences' and, as a result, to motivate students. Sub-questions introduce scientific knowledge the students are expected to 'need-to-know' in order to develop a greater understanding of the scope and the depth of the driving question. The set of successive sub-questions is carefully derived from the driving question by putting oneself in the place of the students (Rivet *et al*., 2000). The driving question and related sub-questions form the framework for learning [characteristic 2].
In the Helmet project the driving question was 'Why do I need to wear a helmet when I ride my bike?" (Rivet, 2003, p.46). Rivet expected that the contextualizing theme, bike riding, bike accidents, and helmets would provide a purpose for developing understanding of scientific ideas such as motion, velocity and force.
In the Evolution of water project the driving was formulated as 'What is the water like in our river?'
The project is designed to foster students' collaboration within a 'learning community' as students are seen as novices who have to learn to participate in a scientific community. This, of course, requires the use of language. Students

communicate with each other, teachers, community members and scientists to find information and solutions to *their* questions and to discuss *their* findings and understandings [characteristic 3]. The website is used as an (innovative) platform for this learning community.

Evaluation

The Evolution of Water project was initially evaluated in more detail than the other projects. In two pilot studies intensive observations, students' artefacts and students and teacher interviews were analysed and used to establish "where teachers and students had difficulty and what parts of the curriculum were missing" (Rivet *et al*., 2000, p.11). The authors of 'the evolution of water, designing and developing effective curricula' report that both teachers and students lose sight of the 'driving question' as the project evolves. Their activities become 'de-contextualised' as they call it:

> Data from students' artefacts and interviews indicated that students did not see a connection between the content presented in the class and the context of the water quality project.
>
> (Rivet *et al*., 2000, p.15)

The developers adjusted the content and content sequence and introduced what they called anchoring events and experiences, thus attempting to strengthen the connection between content and *context* from the students' point of view. They do not report if and to what extent this has been successful.

The evaluation of the Helmet–project focussed on selected contextualizing lessons from the curriculum materials.

> Contextualizing lessons were those in which the contextualizing features of instruction played a dominant role. Such lessons included the introduction of the project and the anchoring events; the set-up to benchmark lessons and inquiring activities that established a need to know situation for student learning; the wrap-up to these lessons where students integrated what they learned with what they already knew and related back to the contextualizing problem or situation; and the culminating event of the project where students further integrated and applied their knowledge to address the overall contextualizing problem through artefacts and presentations.
>
> (Rivet, 2003, p.57).

Also in this case the analyses of what happened in the classroom during these lessons were not specifically related, in detail, to the design and the aim to, for example, create a need to know. Class observations, interviews and pre- and post assessments were used to answer evaluative questions such as: "What does contextualizing instruction in middle-school project based science classrooms look like? What are students' uses of the contextualizing aspects during instruction? What is the relationship between students' utilization of the contextualizing aspects of the project and their science learning?" (Rivet, 2003, p.64).

2.3.3 Concluding remarks

The study of the six projects shows two things. Firstly, that all these projects can be interpreted as trying to give content to the three characteristics of meaningful: *context*, *need to know* and *attention for student input*. Moreover, in these projects, the *need to know* and *attention for student input* characteristics seem to be inspired by ideas which are rooted in learning theories. In fact, all projects refer to such ideas, which are considered neglected (according to the developers) in traditional science and chemistry education, such as: students should be actively involved in the teaching-learning process and there should be more attention for the student's position, ideas and motives for learning. The *context* characteristic is generally rooted in the idea that a more authentic picture of science should be created, and is sometimes also (according to the developers) based on learning theories. The way the *context* characteristic is given content ranges from 'elaborating on the stories of the major ideas in science' (Salters' approach, with reference to Bruner) to 'imitating roles of an existing chemistry practice' (Chemistry in Products, with reference to Vygotsky and Leont'ev).
Secondly, the study of these six projects shows that none of them evaluated whether the design lived up to its aims: did it give rise to the intended teaching-learning process?

2.4 The research question

I will now adopt the three characteristics, *context*, *need-to-know* and *attention for student input*, at least tentatively, as essential ingredients of a *general framework for design of meaningful chemistry (and science) education.* By giving the three characteristics of meaningful the status of essential ingredients of a general framework for design of meaningful science education, I do not suggest that those characteristics are new or revolutionary. In fact, most of them are almost obvious. In adopting the framework, I merely explicitly underwrite the sorts of potential solutions for some problematic features of chemistry (and science) education. The general framework as such or the learning theories they may be rooted in, therefore, is not *primarily* my object of research. In the first instance, I think, the main problems primarily concern the quality with which the framework is given content.
Firstly, the framework does not provide for distinct design principles. It merely provides for rather general directions for designing a teaching-learning process. Secondly, none of the innovative projects evaluate to what extent meaningful chemistry, physics or science education has actually been realised *in relation* to the detailed designed teaching-learning process. Research concerning projects like the ones described in 2.3.2 generally focuses on the evaluation of larger-scale characteristics of the designs. They address questions such as: 'What are suitable contexts?' (which, for example, both lead to certain content and appeal to students), 'What are the learning effects of the new approach?' or questions that refer to implementation: 'How to involve teachers?'. These types of questions are mostly concerned with what can be described as quantitative research, using for example pre- and post tests, involving control groups, large scale interviews and means as questionnaires (e.g. Bennett, Hogarth & Lubben, 2002). Some projects do focus on the

teaching-learning process. However, their object of research is rarely the relationship between the realised teaching-learning process and *the design*. They focus on issues such as ‘motivation’ or ‘quality of argumentation’. The designs involved are merely described in broad outlines and play a subservient role (e.g.: Banet & Núñez, 1997; Patronis, 1999; Roth 1993).
In the very few cases in which what happens in the classroom is evaluated in relation to the design, it merely results in an acknowledgement that characteristics of meaningful are difficult to *really* establish. As described in the previous section, in the Evolution of Water project it was initially found for example that students lose track of the driving question. This led to the notion that a need to know approach is difficult to realise, and although ‘anchor lessons’ were implemented, it did not lead to a fundamental rethinking of the *need to know* characteristic in the design. To conclude: giving content to the three characteristics in a design with sufficient quality and depth is an important, but rather neglected and underestimated field of research. It is the kind of research, moreover, that in my opinion ought to precede or at least inform further theorizing, and could lead to a better or sharper formulation of the general framework (see also Lijnse, 2001).

In the previous sections I argued that if the aim is to involve students in the learning of chemistry, the problematic features of chemistry education should be addressed: rhetoric of conclusions, incoherent sequencing and a lack of attention for student input. This led to the identification of three solution strategies, the three characteristics of meaningful: *context*, *need to know* and *attention for student input*. Although these characteristics can be seen as widely applied in a variety of projects, the quality with which they are given content is rarely object of study. To comprehend the scope and depth of ‘properly’ in the following question a detailed intensive study of the teaching-learning process is necessary.

What is an adequate teaching-learning process that properly embodies the three characteristics of meaningful: a motivating context, a proper need to know and a proper attention for student input?

This research question leads to different types of questions than the ones that are usually addressed. The following questions reflect the criteria for a proper implementation of the characteristics of meaningful based on which it can be decided to what extent meaningful chemistry education has been achieved:

- *Characteristic 1*: Do students relate the context to their daily lives and are they as a result broadly motivated to get involved in the teaching-learning process?
- *Characteristic 2*: Do students really know what they do and why they do it every step of the way and how can this be related to the design?
- *Characteristic 3*: Do they feel that their input matters and how can this be related to the design?

Because of the intensive and detailed character of this type of research one module is chosen as the object of study. Inspired by projects like Globe, the Evolution of Water project, Green and several others, it was decided that the theme of the module would

be ‘water quality’ and specifically ‘testing and judging the quality of surface water in the neighbourhood’ (Becker *et al.* 1997**;** Howland *et al.*, 2002; Rivet *et al.*, 2000). Water quality is a well known theme in chemistry (and science) education. It is a rich theme, in the sense that it embodies a variety of topics such as concentration and accuracy of the test method. It can be expected to appeal to students, because it affects them personally. Also, the aim to let students develop an insight in what is involved in testing and judging water quality fits the purpose that students should learn about how chemistry or science actually functions in society.
Thus, the situated question is as follows:

What is an adequate teaching-learning process in a module about judging water quality for initial chemistry education (students: 14-15) that properly embodies the three characteristics of meaningful: a motivating context, a proper need to know and a proper attention for student input?

Of course, if one module is object of study, conclusions on how and to what extent meaningful chemistry education has been realised are situated, and primarily refer to the design and development of one particular module, which is put into practice in a limited number of classes involving a limited number of teachers. Still, the aim of this research is to get a deeper insight in meaningful chemistry education in general by combining the situated findings of this one module with findings from different but similar projects and broader insights derived from the literature. It is considered to be possible to formulate potential outlines for contexts, the embedding of content in these contexts in a need to know approach and the implementation of attention for student input based on this specific case of meaningful chemistry education.

To conclude, if based on this case research results about meaningful chemistry education can be established, this will contribute to more specific design directions for other modules additional to the directions that were derived from the projects described in section 2.3.2. I also expect that the situated conclusions and ideas will lead to hypotheses (rather than conclusions) at curriculum level.

Next, in chapter 3, I will discuss the research strategy and some methodological issues in more detail.

Chapter 3

Methodology: developmental research and the overall design of this study

3.1 Introduction

In Chapter 2, I argued that if the question is:

What is an adequate teaching-learning process in a module about judging water quality for initial chemistry education (students: 14-15) that properly embodies the three characteristics of meaningful: a motivating context, a proper need to know and a proper attention for student input?

this demands a specific research approach in which the teaching-learning process is object of study. In this chapter I will argue that scenario-based developmental research' is suitable for answering this research question. It leads to a specific kind of theory, called 'domain specific didactical theory'. Central to developmental research is the design and evaluation of a teaching-learning process. Its principle aim is to find out what teaching-learning process actually does the job, in the intended way, of effectively arriving at what we want the students to learn. To achieve this aim, the teaching-learning process to be designed and studied should consist of both a detailed description of a sequence of teaching and learning activities *and* a detailed description of argued expectations of the unfolding of each activity in what is called the scenario (Lijnse, 1995, 2003). Such expectations form the basis for evaluation. They are the hypotheses, which are tested when the designed teaching-learning process is evaluated by putting it into practice. Only then it becomes possible to evaluate the argumentation of the design and formulate design directions which are informed by empirical evidence.
Applied to this situation specific study, its expected main contribution to didactical theory is the spectrum of didactical considerations that will come to the fore and will be evaluated in the attempt to properly design a teaching-learning process that embodies the three characteristics of meaningful. These didactical considerations and the way they are connected provide for design heuristics for *other* modules of meaningful chemistry education in the sense that they provide a framework for making specific, balanced choices (see Chapter 8). This is but one view of developmental research or design research. In section 3.2 I will clarify my position by discussing a number of design studies that each present a different view of design research.
Finally, classroom discourse is interpreted in relation to the designed teaching-learning process. This means that students' remarks, answers, questions and classroom discourse or discussions must be compared to what was expected: Did the teaching-learning process unfold as intended? In section 3.3, I will further discuss this issue, specifically the role of the scenario, as conclusions on the quality of the design and reflections on possible improvements depend on a proper interpretation of the unfolded teaching-learning process.
In section 3.4, I will reflect on the tension between the role of the teacher and the role of the researcher in this type of study. My main point here is that a teacher is a practitioner and therefore expert in *teaching*. Given the fact that teachers do not have time to get involved in this type of research, they can not be expected to be an expert

in designing and evaluating an experimental teaching-learning process with enough research-oriented didactical quality. This, I will argue, must be left to the researcher, although it enlarges or intensifies the problem of how to involve a teacher in a role he or she did not design, i.e. how to nevertheless achieve teacher co-ownership and especially how to bring about that the teacher fulfils his/her role as intended by the researcher.

In section 3.5, I will discuss the characteristics of the successive research cycles of this study. Each phase builds on the preceding and serves a different function in answering the research question, as the result of an increasing understanding of the scope and depth of this question. Developmental research therefore is a spiral rather than a cyclic process. I will address each cycle in the light of its function in answering the research question, including the design of the trials, the type of data collected, methods of analysis and how this led to an increasing understanding of the research question.

3.2 Developmental research and domain specific didactical theory

An increasing number of researchers considers developmental research[1] as a promising strategy to improve educational practice by bridging the gap between theory and practice (e.g. Brown, 1992; Cobb, 2002; The Design-Based Research Collective, 2003; Gravemeijer, 1994; Leach & Scott, 2002, 2003; Lijnse, 1995, 2003; Meheut & Chromat, 1997; Meheut & Psilllos, 2004; Shavelson *et al*. 2003; Tiberghien, 1997, 2000; Viennot & Rainson, 1999). However, their perception of developmental research or design research differs from Lijnse's perception of this type of research, which I underwrite, and in various ways also from each other. They differ with respect to what is to be designed, what informed the design, what is to be evaluated and how and perhaps also with respect to what is to count as a domain specific didactical theory. The differences are, I think, of both width and depth. Some studies address only aspects of the teaching-learning process (width). None of the discussed studies compare detailed and grounded expectations of every teaching and learning activity with what is actually realised in the classroom (depth).

To clarify my position, I will first briefly discuss Lijnse's view of developmental research in section 3.2.1. Next, I will discuss three design studies that each have a different perception of design research in sections 3.2.2-3.2.4. For each of these three projects it will be discussed what was done and to what type of outcomes it led, using what was reported in the literature on these studies. These discussions are from the perspective of Lijnse's view of developmental research. I discuss questions such as: How is the design presented? To what extent (if at all) was the design evaluated?, and: To what extent (if at all) did this lead to conclusions on the design? I want to emphasise here that by doing this I realise that I leave out all other possible perspectives on these projects. It only serves to clarify my position before I describe the design of this study. This section ends with conclusions in section 3.2.5.

[1] Also referred to as 'design experiments' (Cobb 2003) or 'educational design studies' (The Design Based Research Collective, 2003)

3.2.1 Design research according to Lijnse

Before discussing some design studies by Cobb, Meheut & Chromat and Leach & Scott, I will first briefly discuss Lijnse's perception of developmental research.

Lijnse argues that the aim of developmental research should be to understand and improve science education at a very concrete level, by bridging the gap between theory and practice:

> Our "theories" should, in the first place not so much aim to contribute to general ideas about teaching and learning, though we may and should draw on them, but to understanding and improving science-teaching practice.
>
> (Lijnse, 1995, p.190)

According to Lijnse, developmental research deals essentially with questions like 'How to teach X?', or 'How to teach X better?' (Lijnse, 2003). To address such questions and contribute to didactical theory exemplary teaching practices should be developed and studied in detail. Only by reflecting on such practice can it take shape as a method (Gravemeijer, 1994).

Central to developmental research is the scenario (Klaassen, 1995, see also section 3.3). In the scenario the expected teaching-learning process is predicted and justified, which is more than just a detailed description of the teaching-learning activities (Lijnse, 2003). In the scenario it is argued how abstract ideas such as context, need-to-know and attention for student input, are given content in the design. Every teaching-learning activity is accounted for in terms of argued assumptions and expectations of what will happen. These are the set of hypotheses to be tested. In the scenario insight is given into how such an argumentation is based on common sense, empirical evidence derived from trials with previous versions of the design, and on theoretical considerations (see also 3.3 and table 3.1).

> It [the scenario] forces the researcher to make his didactical knowledge, expectations and theoretical perspective explicit in detail and thereby empirically testable.
>
> (Lijnse, 2003)

The scenario becomes increasingly informed by empirical evidence in the process of developmental research in which the design is tested, evaluated and adapted in successive research cycles (see section 3.3 and 3.5). The contribution to didactical theory, as I mentioned before, emerges from this process and from the increasingly fitting scenario. In this case this contribution will be, on the one hand, the set of design directions for modules exemplary for meaningful chemistry education and, on the other hand, it is the wide spectrum of didactical considerations that will come to the fore in the attempt to properly design a module as a case of meaningful chemistry education (see sections 8.2 and 8.3).

From this perspective, Lijnse has two fundamental criticisms of current science education research. First of all, Lijnse argues that such research often refers to abstract ideas rooted in learning theories but does not give any information on *how exactly* such ideas should be implemented in science teaching practice. In other words, they do not address the question of how science teaching is to be concretely informed by an idea such as “the teacher must have a good idea of what concepts the pupils might already have and then engage pupils in activities that would help them construct the desired understanding” (referring to Duit *et al*. 1992). Because they are so abstract, ideas like this may or may not improve the learning results as compared to those of traditional teaching, but the scope of such improvements is and will remain limited (Lijnse, 1995, p.192). They do not provide for any useful directions for improving science teaching practice and in that sense “place too much of the essential burden on the teacher and students, and too little on the researcher” (Lijnse, 1995, p.191).

Secondly, if students are *really* to understand and use what they are taught, they should be engaged in what Lijnse calls a “bottom-up” teaching-learning process in which teaching tasks carefully build on what students know. Lijnse argues that not only traditional science curricula, but also STS curricula such as PLON (see chapter 2) that were aimed at actively involving students in relevant daily life issues, actually teach the basic concepts of science reflecting its basic “logical” structure. But this is *not* necessarily logical from the students’ perspective. You cannot conclude whether students experienced the relevance and logic of the designed teaching-learning process if you have not studied the teaching-learning process in detail[2].

According to Lijnse, the challenge of developmental research is to arrive at a design that solves the tension between the aim of letting students achieve top-down goals on the one hand and making them see the point of every step in the teaching-learning process on the other hand. The only way of achieving this is through a strategy in which the teaching-learning process is be evaluated in detail:

> The design of such teaching is therefore necessarily an empirical process of closely interconnected research and development, that we call “developmental research”. It concerns a cyclic process of theoretical reflection, conceptual analysis, small-scale curriculum development, and classroom research of the interaction of teaching-learning processes. The final, empirically based, description and justification of these interrelated processes and activities constitutes what we call a possible “didactical structure” for the topic under consideration
>
> (Lijnse, 1995, p.192)

Lijnse considers such didactical structures, the didactical considerations that come to the fore and are evaluated in the process of developmental research and the design heuristics that emerge from all this as the contribution of developmental research to

[2] Illustrative of this is the shift in perspective on the teaching-learning process after the first research cycle with the first version of the design in this study (table 3.1).

didactical theory. Of course for every new design based on such design heuristics a new case study of developmental research is needed to secure that a true bottom-up approach is realised (Lijnse, 2003).

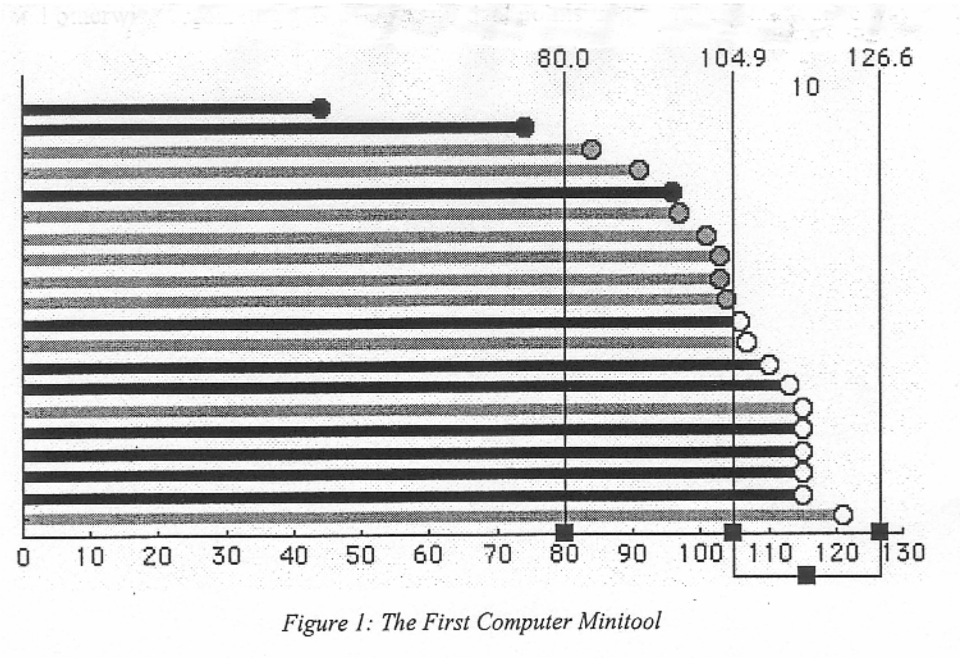
3.2.2 A design study by Cobb

I will first of all discuss a seventh-grade classroom teaching experiment that focused on statistical data analysis (Cobb, 2002). The instructional design that is to support students' mathematical learning is presented in broad outlines. Its quality was not object of study, as I will show below. Instead Cobb and associates studied two episodes (one occurs in the fifth lesson and one in the twenty-second lesson) of classroom discourse. They analyse the episodes of discourse as 'emerging mathematical practices', but without relating this to teaching and learning activities.

What is presented?

The study involves a teaching sequence in a class of 29 seventh-grade students, which consisted of 34 lessons over a ten-week period. The teaching sequence addressed statistical data analysis and is described only in a general sense. The successive phases of the teaching sequence are presented in broad outlines. There is an introduction phase in which the teacher introduces the data students are to analyse. Cobb emphasises that the teacher discusses the creation process of the data with the students to "provide the data with a history" (Cobb, 2002, p.176). In the two successive phases the students work successively with two different computer mini-tools. Cobb describes the characteristics of the two computer mini-tools, which were used by students to manipulate data and describes in broad outlines the possible actions students could perform with each computer mini-tool. The data students are to analyse with the first computer mini-tool concern the lasting time of batteries of two different battery brands. The lasting-times of ten batteries of each brand have been measured and are all presented by this computer mini-tool. The first computer mini-tool, which is introduced in the fifth lesson, presents the data points as horizontal bars, black for one battery brand, grey for the other (figure 3.1).

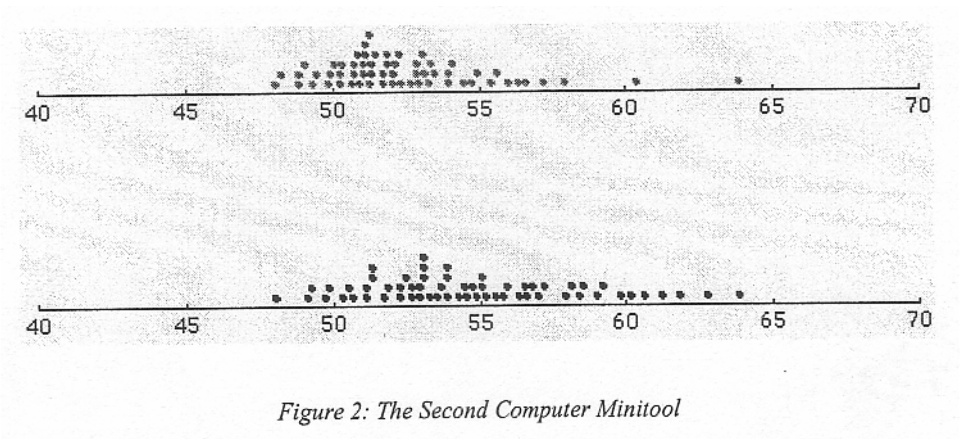
Figure 3.1 Representation of the first computer mini-tool



(Cobb, 2002, p177)

The second computer mini-tool, introduced in the twenty-second lesson, presents the data points as dots on an axis plot (figure 3.2). Students can structure data in various ways with this mini-tool. They can partition the data set into groups of points by dragging vertical bars along the axis, they can partition the data into groups of a specified size, they can create a precursor of a histogram by partitioning the data groups with a specified interval width, they can partition the data in two or in four equal groups.

Figure 3.2 Representation of the second computer mini-tool*.*



(Cobb, 2002, p179)

Cobb describes the introduction of this second computer mini-tool in the twenty-second lesson as follows: first a set of [not further specified by Cobb] data was shown as horizontal bars, similar to the representation of data by the first computer mini-tool. Next, the bars were removed to leave only the dots, which were finally transposed onto the horizontal axis (Cobb, 2002, p179).

Cobb highlights two episodes of classroom discourse, one occurred in the fifth and one occurred in the twenty-second lesson, which emerged just after the students worked with each computer mini-tool. These two moments of classroom discourse are represented by some exemplary protocols, in which one or two students explain how they analysed the data using the computer mini-tools. Cobb identifies different types of student reasoning, one type occurring after the students worked with the first computer mini-tool and the other type occurring after students worked with the second computer mini-tool. The two different types of reasoning are labelled as 'different emerging mathematical practices'. The first 'mathematical practice' is represented by the protocols of three students making comments in response to the teacher or to each other. They discuss the data on the two different battery brands after working with the first computer mini-tool (see figure 1). The discussion is about how the data of the two battery brands should be compared: how to decide which brand is better? Cobb concludes that students discuss this problem, as a result of working with the first computer mini-tool in an additive manner, i.e. in part-whole terms instead of the proportion. One student for example argues that 'seven of the highest data values of figure 1 are of one brand'. He concludes from the first mathematical practice that students reason in an additive way about the data, instead of a multiplicative way. Based on this it is concluded that a practice emerged in which students 'explore qualitative characteristics of collections of data points' (Cobb, 2002, p179).

The characterization of the second mathematical practice, in the twenty-second lesson, is done in the same way. Cobb presents two episodes from two whole class discussions that, according to Cobb, illustrate the second 'emerging practice' (Cobb, 2002, p180). The protocols are of two students who explain how they addressed the question of whether the introduction of a police speed trap on a road with 50 miles per hour speed limit had slowed down the traffic speed and thus reduced accidents. The speed measurement points are presented by dots on an axis. The first student, for example, speaks of 'hills' when she refers to the distribution of the dots. The second student speaks of 'most people': she used 'a fixed interval width of five' to determine how fast 'most people' were driving before the speed trap and after the speed trap. Cobb sees this as a significant shift in the nature of their reasoning about the data, namely multiplicative, instead of additive. The second illustrative (according to Cobb) episode, apparently of the same practice, occurs a week later after the students had worked with the second computer mini-tool on a different data set. In this episode three students say something about the organization of the data while the teacher directs the discussion. One student speaks about 'ranges' where 'the majority of numbers were', and another student explains this to a third student as 'where the point is where it [the distribution of data points] goes up.' (Cobb, 2002, p181). Based on these comments, Cobb concludes that students are now definitely talking about data organised multiplicatively as qualitative proportions (Cobb, 2002, p181). Cobb finally concludes from these two episodes that a practice has emerged in which students 'explore qualitative characteristics of distributions'. Students, according to Cobb, reasoned about data multiplicatively, using the second computer mini-tool to identify global patterns and to describe them in quantitative terms (Cobb, 2002, p183).

Outcomes

The main outcomes of this study of Cobb are 1. The identification of the two mathematical practices that emerged and 2. a description of a 'chain of signification' as a sort of design heuristic.

I will first address the emerging of the mathematical practices. Again, I will do this by viewing the outcomes from the perspective of my view of design research, solely to clarify my position. First of all, it is not clear why Cobb uses the term 'practices' for what seems to be a type of reasoning. It is not clear, for example, whether the specific protocols were exemplary for the majority of the students' explanations and whether the students generally felt *participants* of the practice Cobb sees emerging. Were students aware of some practice, its purpose and their roles in it? In short, Cobb does not present evidence that the students experienced their actions as functional in the light of the purpose of the practice involved. And if this is not he case, what is the advantage of labelling 'types of reasoning' as 'practices'? Secondly, it is not clear whether Cobb (and associates) *expected* the emerging of the two types of explanations of students. In terms of my interpretation of domain specific instructional theory, which should inform future designs, the question that comes up in me is: what *did* Cobb and associates expect and based on what arguments? The comments of the students (the 'emerging mathematical practices') are presented 'as such' without relating them to actual learning and supportive teaching *activities.* As Cobb does not present argued expectations concerning the unfolding of the teaching-learning process, it is not clear at all whether the mathematical practices were the actual goal. There is also no information on what happened in the remaining lessons. How were the two presented episodes embedded in those remaining lessons? Of course, I realise that it is impossible to give a detailed argued description of *all* the activities and their interrelationships, especially when such a long teaching sequence is involved. However if the design is object of study, a general outlining of the different phases of the teaching-learning process, their function and a detailed description of key activities and striking examples of these activities should be presented in order to do justice to the detailed scale of design research. In that way it should at least be made clear what the aims were, how the teaching sequence *as a whole* was expected to guide students towards the goals, and how such hypotheses were tested.

With respect to the second outcome, Cobb presents a 'chain of significance' (Cobb 2002, p.185), which combines the phasing of the teaching-learning process as described in the introduction, with the moments at which the different mathematical practices emerged (figure 3.3).

Figure 3.3 The chain of signification constituted during the design experiment.

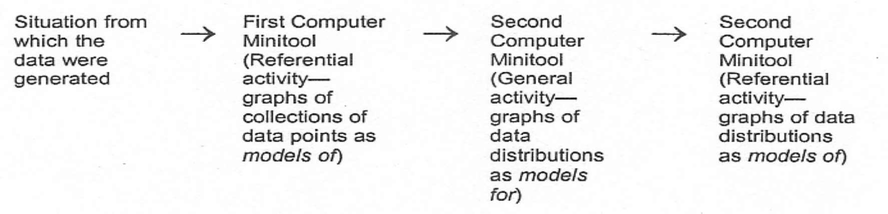


Figure 6: The chain of signification constituted during the design experiment

(Cobb, 2002, p185)

However, the status of the chain of signification is unclear, as Cobb does not explicitly present the chain of signification as an outcome of the experiment. If it represents some hypothetical outlining for future designs, then this raises questions like: how does this direct such a design? It seems to be based on the analysis of the two episodes of classroom discourse, which were not related specifically to the teaching and learning activities and happened far apart, as I described above. An argumentation as to why this can be considered a plausible hypothesis based on empirical findings and (domain specific) didactical theory is lacking.

To conclude, Cobb’s way of performing a design experiment is not adequate as a means to answer my research question: How to adequately give content, in a specific design, to the three characteristics of meaningful with enough quality. In order to answer this question the designed teaching-learning process as a whole should be object of study: does it give rise to the intended teaching-learning process? Cobb and his colleagues seem to be more interested in the emergence of certain types of reasoning (which they call mathematical practices). Cobb and associates do not seem to be interested in how exactly, and why, these types of reasoning emerge from a sequence of teaching and learning activities.

3.2.3 A design study by Meheut and Chromat

The second design study I will describe is by Meheut & Chromat (1997). In contrast to Cobb, Meheut and Chromat describe the learning activities in more detail. The learning activities they describe consist of a series of questions students are to answer about a certain phenomenon shown by the teacher. The questions are to guide students to a ‘particulate model of matter’-type of explanation of these phenomena. Meheut and Chromat also place the study they present in the field of design research (they do this more explicitly in later publications). As I will argue below, Meheut and Chromat focus their study on how students score on the series of open questions. In this sense, like Cobb, Meheut and Chromat highlight an aspect of a possible teaching sequence albeit a different one. Similar to Cobb, Meheut and Chromat do not present a thoroughly underpinned set of expectations of their series of questions, i.e. hypotheses to be tested when the design is put into practice in the classroom, which can lead to a richer theory.

What is presented?

First, Meheut and Chromat describe some difficulties students experience when learning a particular model of matter, which includes organization and motion of particles. Their basic idea is that pupils can be brought to build on a basic particle model, which proposes particles as objects with invariable properties. Meheut and Chromat presume that when students are asked to explain certain specific phenomena, this basic model will direct their explanations towards explanations in terms of organization and motion of particles (being the only possible type of explanation, when particles have invariant properties).

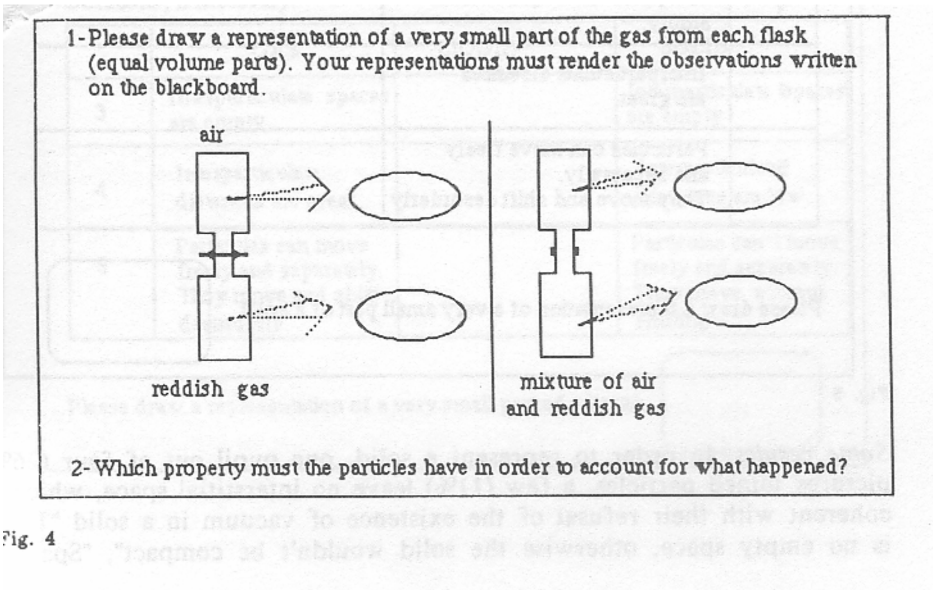
> The invariance of particle properties is supposed to lead the pupils to establish relationships between observable changes and variations in the organization and motion of the particles.
>
> (Meheut & Chromat, 1997, p267)

Meheut and Chromat present how they worked out this idea in a teaching sequence as follows. They describe in broad outlines the phasing of the teaching sequence. For each phase they present in detail the worksheets students were to answer. Meheut and Chromat distinguish three phases A: 'characterization of the system and description of the phenomenon', B: ' building a model' and C: 'making the rules of correspondence between variables of the model and variables descriptive of the system explicit'. In phase C the students are to extend the basic particle model of invariant particles with rules of organization and motion of those particles:

> C. Making the rules of correspondence between variables of the model and variables descriptive of the system explicit. The pupils must now discuss in small groups the compatibility of a few representations with the previous description of the phenomenon [compression of a gas]. Among these representations some have incompatibilities with the description of the phenomenon. Do pupils grasp these incompatibilities? Does this lead them to express clearly the connection between variables of the model and variables descriptive of the modelled system?
>
> (Meheut & Chromat, 1997, p270)

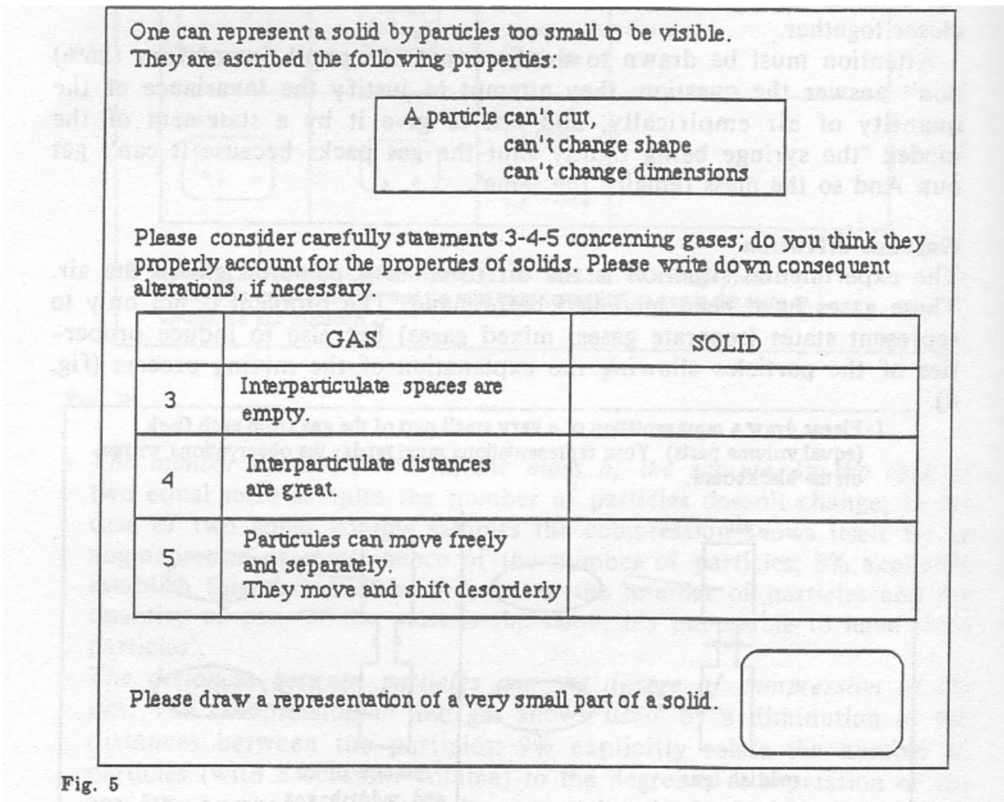
At one point in this phase, Meheut and Chromat present three worksheets on 'gaseous diffusion' and on 'interpretation of properties of the solid state'. These are the very worksheets based on which the students are to explicitly extend their particle model with rules for organization and motion (figures 3.4, 3.5 and 3.6).

Figure 3.4 The first worksheet



(Meheut & Chromat, 1997, p275)

Figure 3.5 The second worksheet



One can represent a solid by particles too small to be visible.
They are ascribed the following properties:

A particle can't cut,
can't change shape
can't change dimensions

Please consider carefully statements 3-4-5 concerning gases; do you think they properly account for the properties of solids. Please write down consequent alterations, if necessary.

	GAS	SOLID
3	Interparticulate spaces are empty.	
4	Interparticulate distances are great.	
5	Particules can move freely and separately. They move and shift desorderly	

Please draw a representation of a very small part of a solid.

Fig. 5

(Meheut & Chromat, 1997, p 276)

Figure 3.6 The third worksheet

One can represent a liquid by particles too small to be visible.
They are ascribed the following properties:

A particle can't cut,
can't change shape
can't change dimensions

Please consider carefully statements 3-4-5 concerning gases; do you think they properly account for the properties of liquids. Please write down consequent alterations, if necessary.

	GAS	LIQUID	SOLID
3	Interparticulate spaces are empty.		Interparticulate spaces are empty.
4	Interparticulate distances are great.		Interparticulate distances are low.
5	Particules can move freely and separately. They move and shift desorderly.		Particules can't move freely and separately. They move without shifting.

Please draw a representation of a very small part of a liquid.

Fig. 6

(Meheut & Chromat, 1997, p277)

The description of the teaching sequence typically stops at this level (throughout all the phases). Next, Meheut and Chromat present how the students answered each question. The percentages of these scores are presented, sometimes together with remarks of students, which show for example that at least one out of two pupils think that to account for the properties of a solid, the particles must be motionless (Meheut & Chromat, 1997, p277).

I will explain next why Meheut and Chromat's approach does not provide the means to answer my research question. Meheut and Chromat do not present an argumentation as to why they selected these specific questions in this order and they do not make clear what their expectations were. In the absence of such expectations, and this is my main point, Meheut and Chromat cannot draw conclusions, which might lead to directions for future designs, as they did not test such hypothetical expectations. They do not argue, for example, how, with the idea of 'invariance of particle properties' as sole input, the worksheets can be expected to lead students to the relationships between observable changes and variations in the organization and motion of the particles. In fact, looking at the three worksheets it comes as no surprise to me that students do *not* propose a particle model of a solid in which particles are moving.
Apart from the fact that the representations are two dimensional and static, whereas the questions refer to a three dimensional dynamic phenomenon and ask for three-dimensional dynamic explanations, it is difficult to see how (this concerns the second worksheet) a discussion on 'the validity of the propositions formulated in order to model the gaseous state' will lead to 'students suggesting consequential modifications so that the model can take into account the properties of matter in the solid state' (Meheut & Chromat, 1997, p276). I wonder what Meheut and Chromat mean by 'consequential'. I can easily imagine that pupils consider the behaviour of the particles of the solid state as distinctly different from the behaviour of the same particles in a gaseous state. In line with Vollebregt, my opinion is that the problem lies in the fact that Meheut and Chromat do not distinguish between phenomena which *can* be directly connected to a model of particle behaviour by students, and phenomena which can*not* be connected directly with such a particle model (Vollebregt, 1998). It is for example possible to image the compression of a gas as particles being pushed into a smaller volume'. In this image the particle model 'covers' the observed phenomenon. But it is different with phenomena which cannot be directly connected to a model of particle behaviour, like 'the characteristics of a solid'. The image of particles organised in a rigid structure is enough to 'cover' the observed characteristics of a solid and from the perspective of the pupils; there is no *need* for an image in which those particles move. This is why Vollebregt argues in her thesis that those obvious connections, in which macroscopic behaviour is directly related to particle behaviour, must be distinguished from less obvious connections, such as 'mean speed of particles increases with rising temperature'. Such less obvious connections are too complicated for students to discover themselves, with the model of 'invariance of particles' as only means, and must be made explicit from the very beginning of a teaching sequence. In view of this, I wonder what Meheut and Chromat expected students would answer, and why?

Outcomes

Meheut's and Chromat's discussion and conclusions are in terms of 'students find it difficult to', based on how students scored on the questions. They for example conclude that students have difficulty to develop a particle model that involves explanation of phenomena in terms of organization and motion of particles. Meheut and Chromat speculate that the reason must be that the representations they used are static, and propose the use of kinetic computer simulations.
To conclude, Meheut and Chromat's way of performing a design study is not adequate for solving my research question either. Although they present their design (the series of questions) in more detail than Cobb and his colleagues, they do not evaluate the teaching-learning process and draw conclusions on the quality of their proposed flow of questions. In Meheut (2004) in which the same study is presented, Meheut seems to acknowledge this problem:

> Experimenting with the first sequence, we did not make our hypotheses about the effects of such or such situation explicit and did not anticipate precisely the possible learning pathways of students through these situations. We could mainly evaluate its global effectiveness with regard to our explicit objectives.
>
> (Meheut, 2004, p617)

3.2.4 A design study by Leach and Scott

Finally I will discuss an experimental study by Leach and Scott, which they explicitly position as design research (Leach & Scott, 2002, 2003; Leach, Hind, Lewis & Scott, 2002). Leach & Scott focus on yet a different aspect of the design, namely on the teacher role and how the teacher might support student learning by promoting specific types of discourse. Similar to Cobb, Leach & Scott present the teaching sequence in a rather superficial way. They do not account in detail for the choices they made in the design.

What is presented?

Leach and Scott's main concern is that the teacher should support "the nature of the classroom talk", they consider as "appropriate at different points of the teaching sequence" (Leach & Scott, 2002, p130). They differentiate between points were "ideas are introduced to learners", which call for, in their view, "an authoritative type of classroom discourse", and points were "teaching should support", what they call "internalisation of those ideas". This implies, according to Leach and Scott, "a dialogic type of classroom discourse", involving "exploration of meanings" (Leach & Scott, p.137 and p.122). As far as I can tell, they do not provide a more detailed description of the role of the teacher in the teaching sequence (Leach & Scott 2002, 2003; Leach, Hind, Lewis and Scott, 2002).

In their search for what a teacher role should look like, Leach and Scott introduce the concept of 'learning demand' to establish at what points learning difficulties can be expected and an authoritative type of classroom discourse is needed. The idea of a 'learning demand' is a kind of operationalisation of how they view learning science. Basically, they consider, referring to Bakhtin (Leach & Scott, 2002; 2003), learning school science to be like learning a social language used by specific communities of people for particular purposes, which differs from the everyday social language students use. A learning demand, according to Leach and Scott, emerges when such *differences* are specifically addressed: differences related to conceptual tools, epistemological underpinning of those conceptual tools used and differences in the ontology on which those conceptual tools are based (Leach & Scott, 2002, p.126). Leach and Scott give several examples of these differences, for example:

> There is evidence that many lower secondary school students recognise the logical implications of specific pieces of evidence in relation to different models of simple series electrical circuits, but resolve logical inconsistencies by selecting different models to explain behaviour of different circuits (ref. to Leach 1999).
>
> (Leach & Scott, 2002, p.127)

They conclude from this example that:

> They [the students] do not draw upon the epistemological principle of consistency that is an important feature of school science.
>
> (Leach & Scott, 2002, p.127)

Similar to Cobb, Leach & Scott describe in a very general sense the flow of the teaching sequence, which was designed together with the teachers involved, without going into specific learning activities. They do not present an argumentation underlying their choices for certain learning pathways. They, for example, propose two possibilities for introducing a scientific model of an electric current consisting of a flow of charge. The first one Leach & Scott call 'the inductive approach', which would involve students making observations and measurements of simple circuits and then working with students to develop a theoretical model consistent with those data. The second one is a deductive approach, which involves "the teacher introducing a simple model and checking out the 'fit' of this model with observations and measurements" (Leach & Scott, 2002, p.134). Leach & Scott think that both approaches are equally possible. They choose the second approach, but might as well have chosen the first one, as they do not present a more distinct argumentation than the following:

> The teacher can help students to develop an understanding of the concepts charge/current/resistance/energy as the model is introduced and these understandings can then be further developed through working empirically with the model.
>
> (Leach & Scott, 2002, p.134)

Leach and Scott do not present a description of the key features of their design, and how and why they thought their design would guide students towards the learning goals. There is no presentation of exemplary detailed striking examples of what happened in the classroom when students were involved in certain key activities. This would, for example, show the reader how to Leach and Scott's expectations turned out. No hypotheses on the design were tested, except to some extent for the teacher role.

Outcomes

Leach and Scott evaluated whether the teacher 'staged the activities in the intended way'. I presume, as no specific examples are presented, that this means that they observe whether the teacher indeed stages an authoritative classroom discourse when learning demands are involved, and dialogical classroom discourse when students need to practice and how this affects learning outcomes:

> The staging of the teaching would be evaluated in terms of the extent to which classroom discourse had followed the pattern proposed in the design of the teaching sequence. For example, classroom data would be collected to evaluate how the teacher changed the balance of authoritative and dialogic classroom talk in introducing new scientific ideas, in supporting student internalisation, in responding to student questions and so on.
>
> (Leach & Scott, 2002, p.138)

In Leach, Hind, Lewis & Scott (2002) the authors reported in more detail about the data collected in this project and the type of research outcomes they expected, as findings were still tentative at that moment. The authors were primarily interested in the relation between learning effects (through pre and post tests) and 'how the teacher set the stage' (through video recordings of all lessons). They, for example, tentatively concluded that students who followed the teaching sequence perform better than their peers in the same school who did not (based on the tentative findings Leach *et al.* present in their report). They speculate about possible causes: 'this [effect] may partly be explained by the fact that the designed sequence emphasised understanding of the concepts tested.' (Leach *et al.*, 2002). Other findings mostly concern the teacher's role. Leach and colleagues, for example, found strong indications that teachers who tend to use dialogic discourse in their usual teaching use significantly more dialogic talk in the teaching sequence, and that a coherent presentation through authoritative talk alone is not particularly successful in promoting pupil learning (Leach *et al.*, 2002).

To conclude, although Leach and Scott did pay attention to the teacher role, in contrast to Cobb and Meheut and Chromat, their way of performing a design study is also inadequate for solving my research question. Apart from some general guidelines for the teacher role, a learning demand does not direct the detailed design of a teaching-learning process. It does not answer the question of how certain learning aims should be embedded in a sequence of learning activities, or what appropriate supporting teaching activities might be. Only at the level of a detailed and empirically tested teaching sequence can an 'inductive approach' be rationally compared to a 'deductive approach'. Until then it is just a choice.

This last point can be further illustrated by Leach and Scott's critique on Tiberghien. According to Leach and Scott, the idea of learning demands informs the design of a teaching sequence more adequately than 'the body of theory in didactics' Tiberghien uses to inform the design of teaching learning sequences on 'sound' and 'energy' (Leach & Scott, 2002, p.137):

> The teaching activities designed by Tiberghien and her associates on the basis of a body of theory in didactics could equally have been designed using the notion of learning demand. However, there are key differences in emphasis between how those teaching activities are 'staged' (mise-en-scene, Tiberghien, 2000, p.109) on the basis of theory in didactics, and how they would be 'staged' on the basis of the identification of learning demands and the adoption of social constructivist perspectives. This can be illustrated by the example of students working to 'construct a symbolic representation, in term of the model [of energy], of the experimental setting [a battery operating a motor to lift an object].' (Tiberghien, 2000, p109). The teaching approach was designed to require students to work on this problem in pairs, without any intervention of the teacher, with the intention of supporting 'devolution'[3]. However, in our view, the learning demand to be addressed in this case involves enabling students to use parts of a new social language (focussing on modelling devices in terms of their role in energy transformation). In order to do this the teacher's role would involve promoting dialogic discourse, with the dual functions of supporting internalisation on the part of the students, and allowing the teacher to listen to student talk and assess their learning.
>
> (Leach & Scott, 2002, p.137)

I think it cannot be concluded that the abstract idea of a learning demand (in the sense as described above) informs the design of a teaching sequence *better* than 'the body of didactical theory Tiberghien uses'. This discussion takes place on such an abstract level that it cannot be resolved. Tiberghien might disagree with Leach and Scott, and that is it. To provide a contribution to didactical theory, as envisioned by Lijnse, one should focus on the level of what was *actually* designed and how it *actually* was realised. At this level, teaching sequences and their grounded expectations (the scenario) can profitably be compared. The discussion would then focus on didactical issues such as the phasing of the design and the functions of these phases, in light of the aims, the design of key activities, etc. The next step, which might resolve such a discussion, would be to empirically test the two teaching sequences.

3.2.5 Conclusions

I used extensive descriptions of three design studies to clarify my position and explain my choices in methodology, because a short description of my arguments on methodological issues of developmental research would not have been sufficient. The need for such an extensive explanation is illustrated by the fact that Meheut positions Lijnse as a radical constructivist, which shows how Lijnse's approach is sometimes seriously misunderstood (Meheut, 2004, p.606).

[3] A concept which refers to the teacher handing over responsibility for the learning process to students, and the students taking this responsibility

To summarise, the discussed researchers differ in their perceptions of design research in two main ways, which appears to be exemplary for this field of research.
The first one is that each study highlights *aspects* of the teaching-learning process in depth: the teacher's role (Leach & Scott), the learning activities (Meheut) or classroom discourse (Cobb). However, according to Lijnse, as I described in section 3.2.1, *all* these aspects and their interrelatedness should be studied. Didactics concerns the organisation of the content to be learned both in a sequence of successive learning activities and in supportive teaching activities, in such a way that it supports the learning process of the students and the learning goals are sufficiently met (Lijnse, 1995, 2003).
The second difference in interpretation is that no *expectations* are formulated, to compare with what happened in the classroom, let alone on the detailed scale of every activity. An exception is the design study of Leach and Scott. Although they focused on one aspect of the design, the teacher role, they did compare the intended teacher role with the realised teacher role. When the design is object of study such detailed expectations are necessary to be able to evaluate the quality of the design properly. Of course I realise, as I mentioned before, that it is impossible to present a design experiment in full detail, but it must at least be made clear that it *was done* in detail. It must be clear what characterized the design, including an argued general outlining and some striking detailed examples of key activities. And it should be made clear how conclusions were drawn based on empirical evidence by presenting some exemplary detailed striking examples.

To conclude, in my study the didactical quality of the design will be object of study to contribute to the formulation of a domain specific didactical theory as I described in the introduction of this chapter (see also chapter 8). This implies that each activity has to be accounted for. As I will show in chapters 4-8, this interpretation of a design study leads to a much more detailed study of how to organize the content properly in successive learning activities so that they guide students to the learning goals than in the cases described in sections 3.2.2-3.2.4. I will also show how a much more detailed approach of the teacher's 'stage setting' is possible (and necessary) using 'interaction structures'. This leads to much more detailed relations between 'stage setting' and the type of content-related progression than Leach and Scott propose (see also section 8.3). Interaction structures describe teacher activities on the detailed level of how each teaching activity supports what students are supposed to do in *each* learning activity. It involves detailed descriptions of how, with respect to the content and the type of discourse, the teacher introduces a learning activity (or a series of learning activities), guides students in performing the learning activity and evaluates the learning activity. The way the teacher does this should support the aims of the learning activity involved (see chapters 4 and 5).

Such a detailed approach of the teaching sequence has, of course, implications for the teacher role in the design research, as I will explain in section 3.4: 'role of the teacher'. Finally, classroom discourse is interpreted in relation to the designed teaching-learning process. This means that students' remarks, answers, questions and classroom discourse or discussions must be compared to what was expected: Did the teaching-learning process unfold as intended? In section 3.3, I will further discuss this issue: the scenario for interpretation and reflection, as conclusions on the quality of the design and reflections on possible improvements depend on a proper interpretation of the unfolded teaching-learning process.

3.3 Using the scenario for interpretation and reflection.

In this section it is discussed how the scenario is used for directing classroom observations and for reflecting on discrepancies between what happened in the classroom and what was expected to happen (e.g. Klaassen 1995; Vollebregt 1998; Kortland, 2001; Lijnse, 1995, 2003).

First of all, when designing the scenario the researcher/designer should place him/herself in the position of the students (Klaassen, 1995, 2003; Lijnse & Klaassen 2004). To achieve that students see the point of every activity to be designed the researcher/designer should empathise with the students' thoughts and considerations as much as possible by asking him/herself constantly: how would I (being in the position of the student) interpret this activity? How would I act? What conclusion would I draw? These questions together with the top-down established learning goals guide the design of the scenario. That it is possible to place oneself in the position of 'the Student' and design a scenario this way assumes that people have shared thoughts (common ground) and that a reasonable continuation of these shared thoughts will be shared. This point of departure is based on Klaassen[4], who adapted the work of the philosopher Davidson to the field of science education (Klaassen 1995, 2003).

Secondly, as described in section 3.2.1, the scenario is used as a reference for interpreting the realised teaching-learning process. The data on the realised teaching-learning process to be collected are largely qualitative and consist for example of transcribed protocols of video and audiotapes, worksheets, transcribed interviews etc. (see chapter 6). Because the scenario consists of detailed expectations of what will happen and why this is expected to happen, it provides for crucial guidance in how to select and how to interpret the selected data:

> It enables us to focus in particular on those moments where something unexpected or something crucial happens, or where we felt uncertain about what to expect.
>
> (Lijnse, 2003)

Or, as Kortland puts it:

> The scenario presents an assumption for each successive task about the outcomes of the task under certain conditions. The classroom observations [and later, in more detail the analysis of the other data] can therefore focus on

[4] It is a different point of departure than the idea that student's pre-knowledge is often conflicting with (school-) scientific concepts, an idea which emerged from a constructivist view of learning and resulted in an extensive number of studies on students' 'misconceptions' and 'alternative conceptions' (e.g. Fensham 1999; Duit, 1999; Driver, 1985).

> the question whether or not these conditions are fulfilled to a sufficient degree, and if so, whether or not the outcomes roughly turn out to be as assumed.
>
> (Kortland, 2001, p51)

Of course, the nature and seriousness of discrepancies between the intended and realised teaching-learning process depends on the quality of the scenario. The scenario helps to reflect on such discrepancies: were certain conditions as described in the scenario not met in classroom practice, were the assumptions about the tasks' outcomes asking too much or too little of the students, did the teacher forget something important or was the scenario unclear about what s/he was to do or say? (Kortland, 2001, p52).

To properly interpret those moments of discrepancies between the intended and realised teaching-learning process as much as possible, data sources are used to reconstruct the realised teaching-learning process and answer these questions: field notes from class observations, protocols from class discussions, worksheets etc (see chapter 6 for a detailed discussion on how such a reconstructing process was set up for the third research cycle). While reconstructing the realised teaching-learning process this way one should (obviously) place oneself in the students' position, their interests and purposes and from their perspective try to unravel what is *actually* communicated (by the teacher/books). Students' remarks, answers etc. should be interpreted as part of a communication process in light of their interests and purposes. This means that in the classroom one should look for logic and consistent patterns in students' remarks.
Thus, if answers of students are not consistent with our top-down interpretation of the question, then we must first of all wonder whether this interpretation of ours is correct. Perhaps, the students actually answered a *different* question. That means we have to try to find out what question the students may have been answering. Such information informs the 'bottom-up' evaluation and design as described already in section 3.2.1. Klaassen calls this 'finding common ground' on which the design and the teacher should build.

In the case of water quality, the issue of interpreting students' remarks and finding common ground is relatively easy, as the teaching-learning process does not involve very abstract and theoretical content. However, what always should be taken into account when interpreting students' remarks, is the school setting in which the teaching-learning process takes place. The questions "What is the student saying?" and "What is the teacher actually communicating?' should be placed in this school setting as it provides for unwritten rules for communication and the intentions of performing speech acts (Grice, 1975). A very common example of this is the way the teachers ask the students a question. It is not that the teacher does not know the answer. The school setting ensures that both the teacher and the student know that the teacher *knows* the answer and is (probably) checking the student's knowledge (in other words: the teacher's intention of performing the speech act of asking the question). It might also be the case that the particular student was not paying attention, and that the teacher asked the question with the intention of making the student pay attention. In the latter case a teacher does not wait long for an actual answer from the student, as this is not the intention. S/he only wants the student to be quiet and pay attention. So part of the actual speech content was: “Be quiet and pay attention’. And the student knows it.

3.4 The role of the teacher

The role of the teacher in a design experiment is partly determined by pragmatic considerations, and partly by more fundamental issues. The pragmatic considerations concern the facilitation of the teachers: time in their lessons and in their own time to prepare for experimental teaching. In the Netherlands lessons prepare for an overloaded examination programme and do not leave much room for experiments. Moreover, experiments with new content and methods are at this moment not supported by an alternative examination programme for the students involved, nor are there enough financial means available to facilitate the teachers involved. As a result, researchers are happy with an interested teacher, even when the teacher’s preparation time is scarce.

With this in mind the dilemma is as follows: do we leave a great part of the *detailed* design of the teaching sequence up to the teachers, or do we leave the design to the researchers?

In the first case, the teachers will feel more involved, and are to a larger extent the owners of the teaching sequence. This has the result that more time can be spent on preparing them, but not enough for a proper research-oriented didactical underpinning of the detailed design.

In the second case, this will lead to a detailed design of a teaching sequence, embedded in a detailed and systematic research-oriented didactical underpinning of that teaching sequence. But then the teachers involved will be presented with this detailed teaching sequence, which they do not ‘own’. They will have to get involved (really understand) in the ideas, arguments and often a way of teaching that is new to them.

In this dilemma, I have chosen for the second option, which is in line with developmental research done at the Centre for Science and Mathematics Education at Utrecht University (UU) (e.g. Klaassen, 1995; Kortland, 2001; Vollebregt, 1998; Knippels, 2002; Verhoeff, 2003) for the following more fundamental reason. As I argued before, a sequence of teaching learning activities must be worked out in detail and with enough quality in order to yield a testable didactical theory, which is our research goal. Such a properly worked out sequence of teaching-learning activities, in which each activity builds on he prior one and directs the next, lies at the core of the teacher role. And in my opinion, teachers, not being researchers and not having time or facilities to become researchers, cannot be expected to possess the necessary competencies for designing and reflecting upon a teacher sequence in such detail with such quality. Experiences in our group, which consists of many former teachers, show that it typically takes a couple of years, involving at least two research cycles, before a teaching sequence of sufficient quality can be achieved, based upon which a didactical theory can be formulated.

Moreover, for novices to get involved in this type of research and *really* understand what it means to look at a teaching-learning process from a research –oriented didactical perspective takes time and effort. For me, being a former teacher myself, it took three years with two research cycles to fully understand what it means to design a teaching-learning process with enough didactical quality (see section 3.5). In general, teachers are experts in practising teaching a certain subject, and their (explicit and often implicit) knowledge, competencies and expertise emerge from their experiences in that practice. Designing a qualitatively good teaching sequence within the scope of developmental research is a *different* practice with different purposes, which asks for different reflections, leading to different knowledge, competencies and expertise. Teachers cannot be expected to easily gain the specific expertise necessary in this particular research field, when there is no time and no facilities for that goal (as is the case in the Netherlands). The same goes for researchers with no experience in teaching: they cannot be expected to easily gain expertise in teaching. In general[5], the expertise of the teacher is complementary rather than similar to the expertise of the researcher (being practitioners in different, but related practices):

> Schon refers to this kind of reflection as reflection-in-action. Reflection-in-action involves a kind of reflective conversation with unique situations. It does not depend on the categories of established theory and technique. Reflection-in-action is more concerned with 'problem setting' than with problem solving. Problem setting involves selecting those aspects of the situation that will be paid attention to, and framing them in the content in which this will occur. In doing this, a practitioner draws on a repertoire of exemplars or generative metaphors to frame and reframe the problematic situation. The adequacy of various ways of framing the situation can then be tested by actions that also function as probes to shape and to explore the situation.
>
> (Schon 1983, cited by Geddis, 1991, p175)

The problem of 'How to involve the teacher?' remains. Involving a teacher, who is interested in participating, in a general sense in such ideas as the framework of meaningful chemistry education and the problems it should address, is not the hardest part. Involving him or her in the specific didactical implementation of these ideas in an example is much more difficult. Furthermore, even if teachers seem to be 'on the same level', a new teaching sequence often means a new teaching style and can involve new content. Roth showed in a case study in which he used video to reflect on his own practice, how difficult it is to *change* practice, even when reflecting on one's own practice leads to new insights (Roth, 2003a).

Illustrative is the fact that in this study the intended role of the teacher and the question of how to involve the teacher in this role has only been specifically addressed

[5] Of course there are science teachers who are also science education researchers, but this is generally not the case (not in this study anyway).

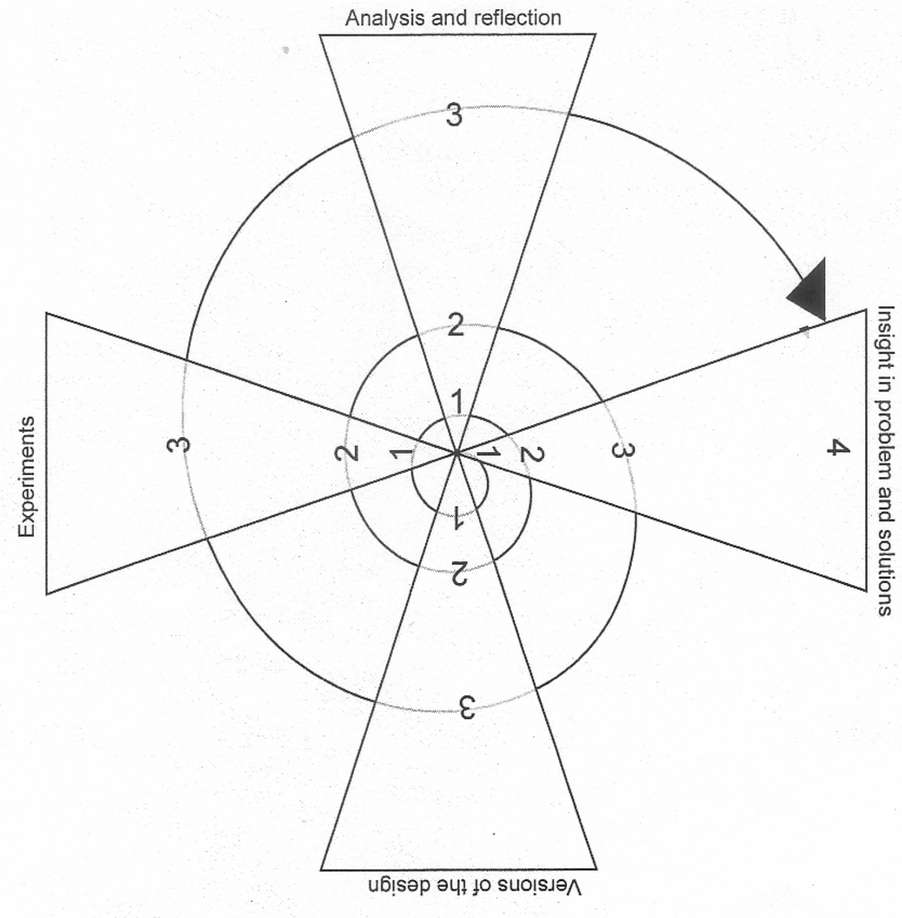
(in detail) in the third experiment. I did not fully grasp the impact of the teacher's role and especially the detailed scale of his/her role on the teaching-learning process earlier (although I thought I did). The first version of the teaching sequence was of such quality that the main focus back then was on the question 'How to drastically revise the teaching sequence in order to improve it?' This resulted in a second, thoroughly revised, version which included such general ideas on the teacher's role as: 'the teacher should pay real attention to student input' and somewhat more specific guidelines as to how this should be done at specific points of the teaching sequence. But I still had not worked out a more systematic approach of describing the teacher's role and how to involve the teachers in such a role. Not surprisingly, as it turned out, the teachers in this second experiment were not properly prepared for their role (and their role was not sufficiently worked out yet). The second experiment helped to clarify the problem for me. That is: it revealed to me in what sense, and at what level of interaction between teacher and students, the teacher's role was productive or counterproductive in 'giving students the feeling that their input matters' and 'letting them experience a need-to-know' (see Chapter 4).

3.5 The overall design of this study

In this section I will describe the successive research cycles of the study with an increasingly meaningful teaching sequence.

This study consisted of four successive stages. An orientation stage by means of a literature study and a pilot trial with the first version of the design (research cycle 1). Two successive research cycles with two successive versions of the design (2&3) and a final reflection, leading to conclusions, discussions and recommendations in the final report (4). Figure 3.7 represents the spiral character, typical for a design study, of these successive stages.

Figure 3.7 The typical spiral character of a design study.
The numbers 1 to 4 indicate each of the four stages. The four triangles represent the four main activities. The increasing insight in the problem and solutions and the increasing quality of the design is symbolized by the widening of the spiral.



The spiral starts with a certain insight in the problem in stage 1, leading to the first version of the design, and so on. The figure is meant to show how the design activities bridge theory and practice through analysis and reflection. The design activities are represented in the lower triangle, the theory in the right triangle (insight in problem and solutions, which includes the creation of a domain specific theory), practice in the left triangle (the trial) and analysis and reflection in the upper triangle. In the first research cycle a teaching-learning process is designed, based on what the insight in the problem and solutions is at that moment. The design is tested in a trial and outcomes are analysed and reflected upon. Reflection is based on both empirical evidence and theoretical considerations. It leads to an increasing insight in the scope and depth of the problem and solutions. The solutions remain, of course, hypothetical to a certain extent. Based on this increasing insight a second, improved version of the design is produced. And so on. In this case, the study ends (for now) with the fourth stage, which is the writing of this final report. I placed this fourth stage in the triangle of 'insight in problem and solutions', because it involves a final reflection on the produced domain specific didactical theory and the study as a whole.

Table 3.1 gives an overview of the different research cycles and of the contributions of each cycle to answering the research question. It is to be noted that this is an overview in retrospect. Interim reflections on the process and planning have led to adjustments in directions taken, details of analysis etc. For example, already at an early stage of the first research cycle it became clear that the first version of the design needed a thorough and fundamental revision. The realised teaching-learning process differed so much from the intended teaching-learning process that a detailed analysis of protocols would have been pointless. Such a detailed analysis of the realised teaching-learning process became only relevant in the second research cycle, when the *need-to-know* approach was worked out in a much better and detailed way.
Table 3.1 also shows how the focus of the study broadened with each research cycle. In retrospect it can be seen that the focus of the design had at first *in fact* been very much on the content. In the second version of the design, with a more proper need-to-know approach, the *students* were given more attention in the design of the successive learning activities. As a result, in the second cycle, the role of the *teacher* in giving students the feeling that their input matters by giving them proper attention became clear and at the same time problematic and also needed attention.
Section 4.2 discusses in broad outlines the first two research cycles. The third research cycle is described in detail in chapters 5 ('the scenario') and 6 (third trial, data processing and presentation') and 7.

Table 3.1 An overview of the different stages of this study

Research cycle:	**1**	**2**	**3**	**4**
Focus	Orientation on problem and solutions	More detailed orientation on *need-to-know* and *context* Orientation on role of the teacher	Proper didactical outlining of the social practice of testing and judging water quality	Drawing conclusions on results concerning the domain specific didactical theory and the process of development of this theory.
Activities	Literature study Design of trial with first version and evaluation	Design of trial with second version and evaluation	Design of trial with third version and evaluation	Overall analysis of data, drawing conclusions and final reporting.
Type of data analysed	Field notes of observations Some protocols of classroom discourse and parts of interviews with the teacher and students illustrative of observations	Detailed analysis of protocols of classroom discourse, of worksheets. Observations and interviews serve as secondary information source Test and questionnaire	Detailed analysis of protocols of classroom discourse, of worksheets. Observations and interviews serve as secondary information source Test and questionnaire	Analysis of the results of research cycles 1-3
Emerging insights about	*Need-to-know-* structure of the design	Relation *context* and *need-to know* structure of the design Proper *attention for students input*: Interaction between teacher and students	Details of a proper didactical outlining of the social practice of testing and judging water quality	A. Design heuristics for meaningful chemistry education B. Didactical considerations that came to the fore in the process of developmental research
Perspective on teaching-learning process	Content *and student oriented*	Content, student *and teacher oriented*	Content, student and teacher oriented	Content, student and teacher oriented

Chapter 4

Giving content to the characteristics of meaningful

4.1 Introduction

In chapter 2 three characteristics of meaningful chemistry education emerged, which can be seen as more or less generally accepted solution strategies for problematic features of chemistry (and science) education: a *context*, a *need-to-know* and *an attention for students input* characteristic. It was argued, however, that to incorporate these characteristics of meaningful with enough didactical quality seems to be a neglected field of research. It leads to specific research questions to establish to what extent meaningful chemistry education has been achieved:

- *Characteristic 1*: Do students relate the context to their daily lives and are they as a result broadly motivated to get involved in the teaching-learning process?
- *Characteristic 2*: Do students really know what they do and why they do it every step of the way and how can this be related to the design?
- *Characteristic 3*: Do they feel that their input matters and how can this be related to the design?

This asks for a specific research strategy: scenario-based developmental research on which I elaborated in chapter 3.

Before presenting the detailed scenario of the third version (chapter 5), and the findings with this version (chapter 7), I will first make clear in this chapter how the third version of the design is based on its previous versions. My aim is also to give the reader insight into how the operationalisation of the characteristics of meaningful and subsequently their interrelatedness evolved in the process of developmental research. I will do this by elaborating on the first two research cycles. For the reader's sake I will describe this process at the level of 'the Characteristics of meaningful' and of 'the Didactical structure & Phases' (see table 4.1). The most concrete levels of description are only addressed in an illustrative manner, when functional for clarifying certain lines of thought. Thus, I will not *present* the extensive detailed argumentation underlying the interpretations and conclusions of these first two research cycles here. For an insight in the kind of argumentation, which is characteristic for developmental research, I refer to chapters 6 and 7, in which the third research cycle *is* presented in detail.

Table 4.1 gives an overview of the different levels of description of the design. These levels range from the more general perspective, the somewhat abstract description of the characteristics of meaningful, followed by the didactical structure with its different steps (see figures 4.4, 4.6A and B), to the most concrete, detailed perspective, which concerns the actual activities. The description of the learning activities of course involves answering questions, reading texts, listening to a specific explanation of the teacher, discussing a certain issue, doing experimental work etc.

To conclude, this chapter addresses the levels of 'Characteristics of meaningful' and 'Didactical structure & phases' in a structural way, whereas the Episodes & interaction structures' and 'Activity' levels are mostly addressed in an illustrative manner (chapter 5, the scenario, and chapters 6 and 7 structurally address *all* levels).

Table 4.1 The design: levels of description

Description level	
Characteristics of meaningful	Most abstract level
Didactical structure & phases	
Episodes & interaction structures	
Activity	Most concrete level

Chapter 4 is structured as follows. First, in section 4.2, I will present an account for the choice of the theme, content and learning goals of the module. In section 4.3, the set ups of the first two research cycles are described in broad outlines. Next, in sections 4.4 and 4.5, I will elaborate upon those specific findings with respect to the first two versions of the design, which led to the most significant design changes resulting in the third version. I have to note that for the reader's sake I present this as a more or less consistent, linear process. In reality however, it is a much more iterative and less straightforward process. The decision which findings are relevant to present in more detail is for a large part determined by their implications for the characteristics of meaningful, that is, for the operationalisation of the characteristics. In section 4.6 the product of the first two research cycles is presented: a schematic general didactical structure of the third version of the design. Such a structure provides for a general overview of the outlining of the design, and in its generality might serve as a heuristic for analogous designs (see chapter 8). The final operationalisation of the characteristics of meaningful and their interrelatedness is made visible in this general outlining, as far as it concerns the steps of content-related progression students are supposed to take.

4.2 Content and learning goals

Inspired by projects like Globe, the Evolution of Water project, Green and several others, it was decided that the theme of the module would be 'water quality' and specifically 'testing and judging the quality of surface water in the neighbourhood' (Becker *et al.* 1997; Howland et al., 2002; Rivet *et al*., 2000).
As mentioned in chapter 2, this decision was based on the following arguments. Water quality is a well-known theme in chemistry (and science) education. It is a rich theme, in the sense that it embodies a variety of topics such as concentration and accuracy of the test method. I expected it to be appealing to students, because it affects them personally. They might swim or fish at certain water sites, or they might be interested in the quality of the water simply because a water site is in their neighbourhood. The reports of the above mentioned projects confirm this expectation. I also thought 'testing and judging water quality' to be a type of activity in which students could work in groups to investigate their own research object (in this case a water site), thus

providing for an opportunity to address student input. Also, the aim to let students develop an insight in what is involved in testing and judging water quality fits the purpose that students should learn about how chemistry or science actually functions in society.

The scope and depth of the learning aim ‘students are to develop an insight in what is involved in testing and judging water quality’, thereby considering that the module consists of five lessons of 50 minutes, was established as follows. First, a procedure for environmental quality research was derived from a standard textbook on environmental research (Haselager, 2001). The procedure described in the standard textbook involved the following basic steps:

1. Formulating a research question
2. Operationalising the research question in testable objects.
3. Establishing a reference list to compare the test results with.
4. Establish a suitable test strategy, including methods
5. Producing test results
6. Comparing the test results to the reference list.
7. Answering the research question.

In the case of water quality research, step 1 involves the formulation of 1) a research question, such as: ‘Is the water quality clean enough for swimming?’ and 2) criteria for, in this case, swimming water quality. Such criteria should be formulated at the level of, for example: ‘Swimming water should be *safe* to swim in’. The second and third step involves making these types of criteria operational, that is: translate them properly into testable parameters and accompanying criteria to compare the test results with (Burns, 2000, p.108). In this case the criterion ‘safe for swimming’, for example, should be operationalised into the proper (qualitative) parameters (for swimming water) to be tested and (quantitative) reference-norms to compare the test results to. Step 4 involves such questions as ‘What is a suitable, that is: representative, sample site?’, ‘What sample frequency is necessary for a representative view of water quality?’ and ‘What are suitable test methods, considering for example that the range of these methods should be within the range of the norms?’. This leads to steps 5 and 6. Based on this the research question is answered in the final step 7. A presentation of the conclusions should include an account of the reference parameters and norms, and the test strategy.

The general outline of this procedure served as a format for establishing what would be relevant steps for students to learn about this theme in order to be attainable for 14-15 year old students in 5 lessons of 50 minutes. The procedure as described in Haselager was obviously somewhat simplified for this purpose but without losing its essence. That is: judging water quality involves the translation of questions such as ‘Is the water clean enough for some purpose?’ into representative and *testable* parameters and reference norms, which always involves doubts at different levels which should be made explicit.

The steps of Haselager's procedure were adapted for the design as follows. Haselager's steps 2 and 3, 'Operationalising the research question in testable objects' and 'Establishing a reference list to compare the test results to', are, according to Haselager, based on a complex argumentation, which is a complete study in itself. These steps were simplified to 'Determine the water function (and maybe possible special circumstances). This determines what are relevant parameters and norms to be tested'. The students were to look up the appropriate legal list of standard parameters (based on the water function). They could find this information on a special website. These legally established lists of parameters and norms would then serve as a reference, like in reality, to compare the test results to.

Haselager's step 4, 'To establish a strategy and the proper test methods' was also considered to be a much too complicated step for students. It was decided that students would be given the test methods. Students were to think about what might be a representative sample site, but to figure out what might be a representative 'frequency of the sampling' was expected to be too complicated and much too time consuming (as there was only time for one sample activity). This last issue of step 4 was therefore skipped, although I could imagine that students might come up with this issue themselves. Also, Haselager's last step involves in reality a complex and detailed reflection on the accuracy and reliability of used test methods, which should also be comparable to the officially used methods, and on how the reference list of values was established. This I expected also to be a much too complex step for the students. It was simplified in two ways. First by the fact that step 3 was simplified (to 'look up the reference list of parameters and norms for the concerning water function'); the reflection on the procedure of this step was simplified as a consequence. Next, the accuracy and reliability issues were reduced to a general reflection on the accuracy of the test method in terms of 'The variability of the test results (that is: range in the test results) and their position with respect to the norm', simply formulated as 'Variability and position'. Students were expected to develop a notion that the variability in their test results reflects their accuracy. Of course this only reflects the extreme test results (Burns, 2000, p.48). However, I considered it too complex to involve concepts such as standard deviation. I did expect that students could question extreme test results as clear 'mistakes that should be skipped'. They were to discuss for each of the parameters they tested whether their results, in spite of the variability, were sufficiently within the norm. That is: whether the position of their results, including their variability, was such that they would conclude their results to be within the norm.

The second notion that students were to develop with respect to the test results was that their performance might influence the test results and therefore their judgement.
All this I expected to result in the notion among students that test results *give an impression* of the true values and how this might impact their judgement. As I mentioned, I expected that all this would make the procedure attainable for students in the given time, without losing its essence: the research question should be translated into a measurement question and, in turn, the research question should be answered based on the test results. This always involves certain doubts.

Figure 4.1 The overview of the procedural steps of testing and judging water quality

Judging surface water quality
Determine the water function *(swimming water)*, this determines the demands
Determine the demands *(safe)*, this determines the criteria
Determine the criteria: what parameters and norms should be tested?
Determine the accuracy *(doubts)* of the test methods
Measure parameters
Compare test results with criteria (bearing the accuracy of the test method in mind), this determines:
Your judgment *(including an account of the accuracy of the test method)*

4.3 The set-up of the trials

In this section I will describe briefly how the trials with the first and the second version of the design were set up. Issues like the type of schools involved, teacher preparation, classes and numbers of students, data types and data analysis are addressed.

4.3.1 Trials with the first version of the design

The schools

The trials with the first version of the teaching sequence were conducted at two different schools.

School A was a protestant school, situated in a small village nearby Utrecht. The general teaching approach was traditional, that is: it could be for a large part characterised as teacher centred, whole class instruction. Students typically sat in rows, facing the black board.
In this trial two teachers participated, A1 and his class of 19 students, and A2 and his class of 23 students. The classes consisted of 14-15 year old pre-A/lower secondary-level students. The trials took place in April-May, 2001.

Shortly afterwards, May-June, 2001, the third trial took place at School B in Amsterdam. School B bases its pedagogy on the Montessori vision:

> Maria Montessori focused on the abilities of the students, and on what they are ready for. With astonishing results. She created a stimulating environment and good education, and organised individual guidance for the children. All this is found in our school. A school where teachers and students are considered equal. That determines how they treat each other.
> As a student you have a certain amount of freedom. You learn to independently discover, organize, decide and do as much as possible. Therefore, you partly decide for yourself what you are going to work on and when. You can work at your own pace, as long as the work is finished within a certain period. Teachers will guide and help you. Cooperation and discussion are important features of our pedagogy. Therefore a lot of work is done in groups, in which students of different ages are often working together.
>
> From the school website: http://hosting.alias.nl/msa/mla/

The general teaching approach was, not surprisingly, in line with these ideas. Students worked at their own pace, individually or in small groups. The teacher guided these processes, giving students much individual attention. Whole class instruction was rather exception than rule, as students obviously scarcely worked on the same tasks at the same time. Also, the tables were not positioned in rows but in groups, to support this way of teaching.
In this trial one teacher, B1, and his class of 7 students participated. This class consisted of 15-16 year old students who redid their pre A-level year.

Teacher preparation
Looking back, the teachers were prepared for their roles in a rather superficial way. In fact, the role of the teacher was not explicitly worked out in the first version of the design. In two successive meetings I discussed with the teachers the actual learning materials (a syllabus) and the teacher guide. This guide gave an overview of the general ideas behind the design, the learning goals, a possible planning for each lesson, a time indication and a proposal for an instructional format (class discussion, group work, individual work etc.) for each learning activity. Also expected outcomes were discussed. Before every lesson, I discussed with the teacher his planning and together we decided in broad outlines how he could ‘set the stage’ for the learning activities (which we evaluated afterwards).

Data collection
All lessons were observed and audio taped. All the group discussions between students were audio taped. Students were asked questions during activities about what they were doing and thinking. All students’ work sheets were collected. To establish whether these answers were not just the answers ‘indicated by the teachers as the right ones’, explicit attention was paid to whether the work sheets were collected before or after an intervention by the teacher. All students did a final test and were asked to fill in a questionnaire at the end of the lesson series. A selection of students was

interviewed after teaching. All meetings with the teachers, also those meetings just before and after a lesson, were audio taped.

Data analysis
Very soon it became clear that the didactical structure and outlining of the teaching learning activities did not meet the criterion for the *need-to-know* characteristic and were of such quality that a thorough revision was needed (see section 4.4.2). A detailed analysis of the data was therefore pointless. Clues for revision were based on comparing more general observations of the realised teaching-learning process (also using information from the evaluative interviews with teachers and students) to more general expectations of the sequence of activities. The audio taped classroom discourse was therefore only transcribed in detailed protocols as illustrative for such clues for revision.

4.3.2 Trials with the second version of the design

The schools
The five trials with the second version of the teaching sequence took place at the same two schools A and B, and with the same three teachers, A 1, A2 and B1.
The teaching sequence was first put into practice at school B, in March-April 2002. Three classes of respectively 25, 25 and 28 students participated in respectively three trials. Shortly afterwards, in May-June 2002, the teaching sequence was put into practice in two trials at school A. The first trial took place in the class of teacher A1, with 27 students, the second in the class of teacher A2 involving 28 students.

Teacher preparation
Looking back, I have to conclude that the preparation trajectory of the teachers for this second trial was again rather superficial. It involved two meetings. In the first one the conclusions of the first trial and revisions in the design were discussed. The purpose of this meeting was to involve the teachers in the rationale behind the revisions and to check if they were on the same wavelength. Although it differed considerably with the first version I hoped and expected it would be relatively easy to involve the teachers in the rationale behind this second version, as they had participated in the first trial and I could refer to their experiences. In the second meeting the teaching sequence was discussed in detail. With every teacher, each lesson was discussed before and after teaching.

Data collection
The data were collected in the same way as described for the first trials (section 4.2.1).

Data analysis
This time the didactical outlining of the design was much more adequate (see section 4.5.2), that is: the criteria for meaningful chemistry education were increasingly met, especially based on first observations. Therefore, the realised teaching-learning process was now compared to the detailed expectations and justifications for each learning activity (the scenario). Class observations provided first impressions but the primary information sources were the audio taped class discussions and discussions

between the students. The students' work sheets (which were collected only at those moments where it was established that the students had not adjusted the answers after a class discussion) and the observations were used to reconstruct the realised teaching-learning process. Interviews mainly helped to clarify some of the students' remarks and to verify interpretations.

4.4 The first research cycle: design, findings and implications

In this section I will present in broad outlines the first research cycle. This section is structured as follows. In section 4.4.1 the first operationalisation of the characteristics of meaningful is presented. Next, in section 4.4.2, the design is presented in broad outlines, including argued expectations as to how the criteria for meaningful chemistry education were expected to be met. The main findings of the evaluation are presented in 4.4.3 and finally, in section 4.4.4 the most important implications for the next version of the design are discussed.

4.4.1 The characteristics of meaningful in the first version

In this section I will elaborate on how the three characteristics of meaningful were operationalised in the first version of the design. For each characteristic its related specific criterion for evaluation is formulated.

Context

In the first version of the module, the *context* characteristic was operationalised as follows: a relevant and recognisable driving question that was expected to appeal to students. In other words: the use of the driving question 'Is the water of our neighbourhood clean enough?' was expected to provide for a motivating recognisable *context* for the content to be learned. The choice for this driving question was inspired by projects like Evolution of water, GREEN and Globe (Becker *et al*., 1997; Howland *et al*., 2002; Rivet *et al*., 2000). Similar to these projects, students were to test and judge water in their neighbourhood in thematic projects (see below, *need-to-know*, part III in figure 4.2). The projects all reported that this *context* generally motivated the students.

The *context* characteristic was evaluated based on the criterion: 'does the driving question appeal to students in the sense that they relate the driving question to their daily life? And are they motivated to answer it?'

Attention for student input

This characteristic was operationalised in two ways:

a. Students experience certain autonomy of choice, in particular which water they want to test.

Each of the different groups were to choose themselves which surface water in their neighbourhood they wanted to test. They were to decide themselves what they would test their water on and were to produce their own list of test results.

b. Students present their own findings to their classmates. In a final reflective class discussion, guided by the teacher, the conclusion is drawn that they all basically followed the same procedure.

The insight that each group basically followed the same procedure and the realisation that all groups contributed to this insight should add to the students feeling that their input matters.

The *attention for student input* characteristic was evaluated based on the following criteria: Did students feel that their input matters in the sense that:

a. they experienced and appreciated that they could make their own choices?
b. their input (their findings and presentations) contributed to achieving the learning goal of making the common procedure explicit?

Need-to-know

This characteristic as operationalised as follows: A set of sub-questions that students were expected to need for answering the driving question. The sub-questions were derived from the driving question by putting oneself in the position of the students keeping in mind the learning goals.

Because it was a learning goal that students mastered the procedure for judging water quality, some of the sub-questions could, in retrospect, be seen as somewhat reflecting the steps of this procedure (see figure 4.1 for an overview of the steps), although at the time this was not the explicit aim. See table 4.2 for a general outline of the first version of the design.

The *need-to-know* characteristic was evaluated based on the criterion: did students experience what they learned by answering the sub-questions as contributions to answering the driving question?

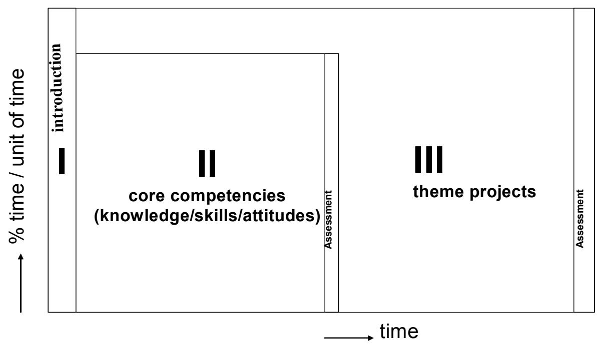
4.4.2 Design and expectations

In this section I will present in broad outlines how the characteristics of meaningful are elaborated upon in the first version of the design and discuss the accompanying expectations. I will first discuss the general structure of the module. Next, each of the successive sub-questions is discussed. For each sub-question I will argue how I expected it to meet the *need-to-know* criterion ‘students experience what they learn by answering the sub-questions as contributions to answering the driving question’

The structure of the module

The module was designed according to a thematic/basic knowledge and skills model. The format was based on project-based curricula in Engineering both in Aalborg (Denmark) and at the University of Twente (The Netherlands) (Kjersdam & Enemark, 1994, Peters and Powell, 1999). At the time, the general format was described as follows.

Figure 4.2 Schematic representation of a module based on the thematic/basic knowledge and skills model.



Each module consists of three major parts (figure 4.2):

- I: the introduction to the theme of the module starting by posing a problem related to the theme of the module,
- II: a part in which students systematically learn core competencies (chemical key concepts, skills and attitudes) including an assessment.
- III: a part in which students work in groups on projects within the common theme of the module.

Assessment in part III takes place through a final group product of the project (Powell & Weenk, 2003)

The idea is that in part I the theme of the module is introduced, together with the driving question, which is the central project-question. Students are to work on the driving question in part III. Students are expected to orientate themselves in part I on the product they are going to produce in part III and on relevant, in light of the driving question, thematic knowledge. All this is expected to motivate students in a general way and, more specifically, they are expected to see the point of getting a deeper insight of the issues raised in part II, in light of the driving question. Part II concerns the basic knowledge and skills part. It involves theme-independent, general knowledge, of which the functionality in light of answering the driving question is expected to be clear now to the students. The content of part II, because it must serve the driving question, might very well differ from traditional chemistry content. In part II students are to extend their knowledge in the direction of these basic issues.
In part III students are to work in groups on their projects and are expected to apply the knowledge learned in part II. This part is to conclude with a class discussion on the extent to which the different groups adequately applied this knowledge in answering the driving question. It is expected to lead to a reflection by the students on what they have learned.

The idea of functionality was further elaborated upon in the design of the *need-to-know* characteristic. A framework of the driving question and functional sub-questions was used to give content to this idea as follows. As I mentioned above, the driving question was introduced in part I, in the first lesson, in such a way that students were expected to realise that this question was the *driving* question of the module, and that every step of the module is to serve to answer this question. To emphasise this, the learning activities were explicitly structured around a framework of sub-questions, each indicating such a step. At the end of part I this involved (for example) the sub-question: ‘What do we need to know in order to formulate a *reliable* judgment of the water quality?’ (See table 4.2). This question precedes the introduction, in part II, of new, basic content, in particular about the principle and accuracy of the colorimetric test method to be used in part III.
The main idea (and expectation) was that students would naturally see the point of learning about the accuracy of the test method in light of being able to reach an adequate judgment (the need-to-know criterion). From this they in turn were expected to see the point in light of being able to answer questions like ‘Is the water clean enough to swim in?’, and to make such important decisions as ‘Are we going to swim in this water?’ (the need-to-know criterion). The sub-questions, which in a sense reflected the procedure described in section 4.3, were supposed to follow logically, from the perspective of students, from the driving question, thereby building on ‘what students already knew’ to arrive at ‘what students did not yet know’.
As a result the students were expected to experience that what they learned by answering the sub-questions contributed to answering the driving question and (with respect to the last question) to their feeling that their input contributes to achieving the learning goal of making a common procedure explicit. Table 4.2 gives an overview of the sub-questions of the three parts.

Table 4.2 ***A general outline of the first version of the design***.
The need-to-know characteristic is embodied by the framework of sub-questions, which should logically follow from each other and in relation to the driving question for students.

Part	Sub-question	Lesson
I. Orientation and motivation	1. What are we going to do these five lessons?	1
	2. Judging water quality: what steps are involved?	
	3. What information do you need to be able to judge the water quality?	
	3A. What things should you take into account when you want to judge the water quality?	
	3B. What is a suitable sample site?	
	3C. What are relevant parameters and norms you should test your water on?	
	4. What do we need to know in order to formulate a *reliable* judgment of the water quality?	
II. Extending knowledge	5. How and how accurately can we test the parameters?	2
III. Applying knowledge	6. Can we apply our knowledge to an example?	3
	7. Is 'our' water clean enough for its function?	4
Reflecting	8 Did we adequately judge the water quality of the water in our neighbourhood? 9 What is the common procedure for testing and judging water quality?	5

The sub-questions and expectations

Next I will give an impression of how I addressed the idea of functionality for each of the successive sub-questions. For each sub-question I will argue how I expected it to meet the *need-to-know* criterion: students experience what they learn by answering the sub-questions as contributions to answering the driving question.

1. What are we going to do these five lessons?

This sub-question was to give students a clear overview of the purpose of the lessons (answering the driving questions in a satisfactory manner) and their set up (testing and judging surface water in the neighbourhood themselves and as a result learning the knowledge and skills involved). The idea was that the teacher would 'enthusiastically introduce' the driving question and the set up of the lessons thereby involving students and their experiences with this theme as much as possible. The learning materials included a similar, short introduction of the theme, driving question and an overview of the lessons. I expected students to be motivated by this purpose (answering the driving question) and by this set up (The *context*-criterion). I also expected that students would experience what they learned by answering the sub-questions as contributions to answering the driving question (the *need-to-know* criterion).

2. Judging water quality: what steps are involved?
This sub-question was to give students an overview of the procedural steps involved in order to activate their pre-knowledge about some steps and to put them on the right track, if necessary ('This is what we are going to do', as a sort of advance organiser). The teacher was to explicitly point out to the students the schematic overview of the steps involved in 'judging surface water quality' in the syllabus (see figure 4.1). I expected their pre-knowledge to be activated in the sense that they would recognise and appreciate the logic of many steps (*need-to-know* criterion). I also expected that this overview would help them to structure beforehand the activities throughout the lessons, in the sense that they could place each activity in the procedure (*need-to-know* and *attention for student input* criterion-b).

3. What information do you need to be able to judge the water quality?
This sub-question was to functionally introduce the first three steps of the procedure in light of the driving question. The sub-questions 3A, B and C addressed information you *need* in order to answer the driving question, such as the relation between water function and water quality criteria, factors that can influence the water quality and the relevant parameters and norm to test the water on (see below) (*need-to-know* criterion).
The question was used as the heading of the section in which sub-questions 3A, B and C were addressed. In the introductory-text of this section the following example was put to the students:

Imagine this water site in your neighbourhood. Perhaps you would like to swim in it in the summer. How would you know it is safe?

The idea was that the teacher would use this example to make the sub-question more clear.

3A. What things should you take into account when you want to judge the water quality?
This question was to make students realise that these water quality criteria need to be translated into a testable list of parameters and norms by using the intuitive notions of students of the relation between water function and quality criteria, and the notion that water might contain all kinds of stuff which influences the water quality (*need-to-know* criterion).
In the learning activity, which had this sub-question as heading, students were to answer different questions. I expected that these questions would activate their intuitive knowledge of the relation between water function and water quality criteria. The students had to look up different water functions on a special website. These water functions included lists of water quality criteria (parameters and norms). It was explained in the text why these lists differed and as an example of this students had to look up and compare the chloride-norm for different water functions and to explain

the different criteria for chloride. I expected that students would realise in this way that water quality criteria depend on water use and that these criteria are expressed in testable lists of parameters and norms. The activity was to be done in groups of two or three, and to be evaluated in a classroom discussion.

3B. What is a suitable sample site?

This sub-question was to make students think about the different uses of the surface water sites in *their neighbourhood.* Students were to realise that they should also consider the surroundings of the water site, including factors and circumstances which might influence the water quality (*need-to-know* criterion). They were expected, again, to use their intuitive notion that 'water might contain stuff, which influences the water quality'. The question also served to get the student-groups involved in their projects.

The sub-question was the title of the learning activity in which student-groups (which were formed just before) had to decide which water site they were going to investigate, and which sample site would be suitable. Newspaper clippings on 'real life situations' were used. I expected these to draw the attention of the students to situations in which the water quality is actually influenced by 'stuff in the water' because of the factors in the surroundings. Students were asked to read the newspaper clippings and then to discuss with their group-mates the following question 'What might make a water site interesting?', before actually deciding on a water site. Next they had to discuss 'What would be a suitable, because representative but also reachable, sample site?'. Every group had to conclude this activity by marking their sample site on a map on the classroom wall for everyone to see throughout the whole lesson series.

3C. What are relevant parameters and norms you should test your water site on?

This sub-question was to let students explicitly structure their pre-knowledge concerning 'What do you have to take into account when you are to establish relevant and testable parameters and norms'.

Students were to find relevant parameters and norms on the special web site using two keywords. A newspaper clipping was expected to help students to formulate the two search-keywords they were to use. They were to fiund out that (*need-to-know* criterion):

-you need to know what the water of your water site is used for (by which they could find a standard list on the website).

-you need to consider specific circumstances (of which descriptions and suggestions for relevant parameters were given on the website).

The expectation was that the groups would produce a list of relevant parameters to be tested together with an argumentation (based on water function and specific circumstances).

4. What do you need to know in order to form a reliable judgment of the water quality?
This sub-question could be considered as an intermediate question between parts II and I. Its function was to motivate students to get a deeper insight of the accuracy of the test methods.
I expected that students would easily see the point of a *reliable* judgment, before they actually tested their water (*need-to-know* criterion). This sub-question was addressed in the introduction of the section, which also addressed the next sub-question ('How and how accurate can we test these parameters?', see below). The teacher was to draw the attention of the students to the issue of a reliable judgment and how this relates to the accuracy of the test methods. He was to use an actual example of a situation in which the accuracy of some test result was questioned and at the same time decisive:

Imagine that you have chosen to test the water quality of a water site in a nature reserve. The norm for oxygen in this case is 5,0 mg /l. Imagine that your test result for oxygen with the Merck kit is 0,45 mg/L. Can you now conclude that the water quality meets the oxygen level criterion?

After briefly discussing the example with the class, I expected that the students to realise the importance and relevance of the issue of accuracy.

5. How and how accurately can we test these parameters?
This question was to satisfy the *need-to-know* students were expected to have experienced in light of the previous sub-question: 'What do you need to know in order to form a reliable judgment of the water quality'.

This sub-question was addressed by letting students in groups do some colorimetric analyses, with their eyes as instruments, comparable with the tests methods they were to apply in their projects. The idea was that students would *experience* the inaccuracy of this method and the doubts, which came along. I expected this to have a much stronger impact on them than when they are simply told about it.
Students were for example to establish the concentrations of two unknown copper sulphate solutions using a calibration sequence. They had to write down the results and hand it over to the teacher who, only when everyone was finished, would give an overview of all the estimated concentrations on the black board. Obviously a certain distribution of the results could be expected. And I expected the students to ask the teacher for the 'right answer'. The teacher was to tell the students that there is no 'right answer'. The only information available about how much copper sulphate there was in the water would be the list of their test results. The idea was that the students realise that test results are not 'true values' of what is in the water, but they give an impression of what the true value might be. I expected students to realise that the distribution of the results reflects the accuracy of the tests. The students were asked to give an opinion on the accuracy of this type of tests.
In a second part of this section, students were to trial with so–called 'indirect determination tests'. In these types of tests, the water sample has to be treated first, because the actual parameter cannot be determined directly. It has to be made 'visible

through a reaction and some reaction product is measured. I thought this to be relevant, because the tests in the Merck kit were all indirect. In this section students were to determine the hardness of tap water (calcium/magnesium) using soap and the fact that the more calcium or magnesium was in the water, the less foam development would be visible.

6. Can we apply our knowledge to an example?
This sub-question had two functions: students were to gain experience with the steps involved in judging water quality and with the actual tests.
The teacher was to give the students water samples, without telling them from what site they are from. Students were to answer the question 'Is this water clean enough?'. The idea was that this way:

1. The students would go over and actually apply the procedure, as a sort of test if they really understand what is involved. They had to, for example, realise that they need to know where the water came from, what its function is etc. I expected that students would realise this and ask the teacher.
2. The students were to gain experience with doing the tests. As a result I expected them to become more confident when doing the tests with 'their own' water.

7. Is our water clean enough for its function?
This was actually the driving question in which students needed to apply what they had learned.
Students were to test and judge their water, working in groups and using their experience from the previous sub-questions. They were to produce a report for the web site and prepare a presentation. In this way I expected students to think about the quality of their judgment, including the accuracy of the tests methods.

8. Did we adequately judge the water quality of the water in our neighbourhood?
With this sub-question students needed to reflect on 'What makes a judgment about water quality adequate?'. The different groups were to present their findings and judgments to their classmates, which were to be discussed in a class discussion. I expected the class discussion to focus on the adequacy of the judgments and to lead to the explicit expression of the followed procedure, as this would serve the quality of their answers to the driving question (*need-to-know* criterion). Different groups had tested and judged different water sites with different functions. I expected students to see that to evaluate the quality of the judgments (see the previous question) means to evaluate the steps of the procedure. I therefore expected the procedures they had used to be part of the presentations. As a result, in the class discussion students were expected to ask such questions as: Did you establish the water function correctly? Did you establish the relevant parameters and norms correctly? How did you test these parameters? I expected the students to br able for a large part to ask questions like this to their classmates, having done a similar project and having thought about all the issues themselves. Also I expected that students would be motivated to be critical because sometimes the judgments might influence 'big decisions', such as: will we swim in this water or not?

9. What is the common procedure for testing and judging water quality?
This question was to make students explicitly realise that they all actually applied a common procedure for testing and judging water quality, in other words: they were to make explicit what they had learned.
This question was actually expected to emerge rather naturally from the classroom discussion. I expected that it would be necessary for the teacher to direct the discussion explicitly in the direction of 'a common procedure'. I also expect that the students would easily follow this line of thought, as it had proven to be a *functional* procedure for judging all kinds of water quality.
This last question was expected to contribute to the *attention for student input* characteristic (b): the input of all groups was necessary to achieve that the common procedure for judging water quality is made explicit. The insight that each group basically followed the same procedure and the realisation that all groups contributed to this insight was expected to make the students' feel that their input mattered.

4.4.3 Evaluation: the main findings

The main findings of the first trial were:

There were no indications that the *context* and *attention for student input* characteristics needed rethinking. However, the *need-to-know* characteristic had been designed inadequately. The sub-questions, their content and sometimes timing, did not reflect what students felt they needed to know for answering the driving question. That is: students mostly did not experience what they learned by answering the sub-questions as contributions to answering the driving question. Moreover, several of the sub-questions turned out to be much too complex. As a result, the first version had not achieved that students considered 'arriving at a common procedure for judging water quality' a learning goal. Apparently, the procedure had not been a well-integrated part of the learning process.

In the next section, I will discuss the findings with the first version in more detail. I will focus thereby on the main problem of the inadequate *need-to-know* by discussing the unfolding of each of the sub-questions.

Main findings with each of the sub-questions

In this section, I will briefly evaluate the main findings with each of the sub-questions, and how I interpreted these findings from the perspective of the *need-to-know* characteristic**:** did students experience what they learned by answering the sub-questions contributed to answering the driving question? The evaluation is structured along the sub-questions, which are used as headings in the descriptions below.

1) What are we going to do?
This sub-question was to give students a clear overview of the purpose of the lessons (answering the driving question in a satisfactory manner) and their set up (testing and judging surface water in the neighbourhood themselves and as a result gaining the knowledge and skills involved). This was only partly achieved. In general, after the

teachers introduced the driving question and the set up of the lessons, students reacted very enthusiastically. Most of them instantly began to discuss within their just established groups of what water site they were going to investigate the water quality. They tended to use arguments like: 'We always swim here'; 'It is nearby where I live'; 'They' are going to build there and the water looks funny'; and so on. At this moment, much earlier than was intended, almost all groups were also very interested in the question 'What should we test our water on?'

In general, the students seemed impatient to move on and address the question of what to test their water on, apparently considering the answer to this sub-question to contribute to answering the driving question. As a result, the students seemed to experience the learning activities, that addressed the sub-questions 'What things should you take into account' and 'what are suitable sample sites' as a hindrance, they obviously did not experience what they learned by answering *these* sub-questions as contributions to answering the driving question.
They had already made a choice and wanted to move on to establishing relevant parameters and norms.

2) Judging water quality: what steps are involved?
This sub-question was to give students an overview of the procedural steps and activate their pre knowledge in the sense that they would recognise and appreciate the logic of a lot of the steps. This was not achieved. Students seemed to take the overview of the steps (see figure 4.1) for granted. The teacher introduced it and students glanced at it. They did not ask questions about it, they gave no indications that they did not understand when asked by the teacher ('Is this clear for everyone?') and they never referred to it again (I did not catch students referring to it later on). Also, in the final class discussion (see below), in none of the classes students would refer to a 'common procedure'. Based on this it was concluded that students had not experienced answering this sub-question to contribute to answering the driving question.

3) What information do you need to be able to judge the water quality?
This sub-question was to functionally introduce the first three steps of the procedure in light of the driving question. Based on the findings with the first sub-question, which showed that students immediately addressed the questions 'What is a suitable sample site?' and 'What should we test the water on?' (referring to relevant parameters and norms) and on the findings with the sub-questions 3 A, B and C (see below) I concluded that this sub-question and the way it was given content in the design did not meet the *need-to-know* criterion: students did not experience what they learned by answering this sub-question as a contribution to answering the driving question.

3A) What things should you take into account?
This sub-question was to use the following intuitive notions of students to make them realise that water quality criteria need to be translated into a testable list of parameters

and norms: the notion that water use and quality criteria are related and the notion that water can contain all kinds of stuff that influence the water quality.
This was achieved to the extent that the intuitive notions of students of the relation between water function and water quality criteria were made explicit. It was obvious to the students that water quality criteria depend on its use. They in general could easily explain in terms of different water quality criteria why the different water functions had different norms for the parameter chloride. However, as I described above, students were impatient to move on to relevant parameters and norms. They just did not acknowledge this sub-question to be a step in a strategy to arrive at what might be relevant parameters and norms, as intended, from which it could be concluded that it did not meet the *need-to-know* criterion.

3B) What is a suitable sample site?
This sub-question was to make students think about the different functions of the surface water sites in *their neighbourhood* and to realise that they should also consider the surroundings of the water site, including factors and circumstances, which might influence the water quality.
As I mentioned above, this sub-question emerged naturally among students much earlier and in a different way than intended. Students chose a water site for totally different reasons than was aimed for. They did not discuss the newspaper clippings the way I intended (see above); in general they did not seem to understand what these clippings were doing in this activity. For a large part, the students had already chosen their water sites based on their own experiences: because they swam in it or because it was situated next to their home etc. Students did not think about *suitable* sample sites either. They just pinned a spot without discussing if that spot was representative (that is: for the water quality of the whole water site) and practical (that is: reachable). Also, I had expected that this sub-question would involve the groups in their projects, as they had to pick their object of research for the first time operating as a group. This function was not relevant at this moment either, as the groups had already been formed and the water sites had already been picked. I concluded that the way this sub-question was addressed and the moment it was raised had not been adequate: again, students had not experienced what they learned by answering this sub-question as a contribution to answering the driving question.

3C) What are relevant parameters and norms you should test your water site on?
As mentioned before, students almost immediately found this a very relevant question. When it was finally addressed, it proved to be too difficult. First of all, most students did realise that the water function would determine what would be relevant parameters and norms to test. But in many cases it proved to be not so easy to define the function of a specific water site. When students tried to find clues on the website, they asked a lot of questions, concerning the water function, factors that might influence the water quality, and so on.
Secondly, establishing relevant parameters based on specific circumstances also proved to be too difficult. Sometimes students had such questions as for example: 'The water site we want to investigate has a strange yellowish colour, what should we test?', questions they apparently experienced at that moment as contributing to

answering the driving question (but which the teacher and I could not answer either). All groups finally put together a list, but not all the groups were satisfied. A lot were insecure whether their lists reflected the water quality criteria properly (and so were the teacher and I). This conclusion is partly based on their remarks when making the list and partly when presenting their judgments. The more general conclusion I draw was that, although students did experience the general relevance of this sub-question, I had underestimated the complexity of its details. As a result, it did not meet the need-to-know criterion.

4) What do you need to know in order to form a reliable judgment on the water quality?
This sub-question was to motivate students to get a deeper insight of the principles and accuracy of the test method. This was not achieved.
The teachers all addressed this question. They also addressed the example (see 'need-to–know: design and expectations', sub-question 4) of the situation in which the accuracy of a test result is questioned and at the same time decisive. The teachers used the example that I described before in their explanation why the next section (how accurate are the test methods?) is relevant and functional for answering the driving question. The teachers all very quickly moved on to the next question, going into the principle of 'colorimetric test methods', or what a calibration sequence is, without explicitly checking whether all students understood the functionality of this question as intended. I therefore strongly suspect, in light of how students addressed sub-question 5, that the students did not experience what they would learn by answering this particular question as a contribution to answering the driving question.

5) How and how accurate can we test the parameters?
The function of this question was to satisfy the *need-to-know* students were expected to have experienced in light of the previous sub-question: What do you need to know in order to form a reliable judgment on the water quality? This need obviously was not fulfilled, as there was no *need-to-know*.
In all classes, the students did the section on accuracy of colorimetric methods without complaining (much). That is, I observed that students generally enjoyed doing the practical and guessing the concentrations of the unknown solutions (like it was some sort of competition). They generally concluded fairly easily, as intended, that it 'was not a very accurate method', but none of the groups involved the accuracy of the test methods in their written and presented judgment of their water. At one point this topic really became meaningful in the sense that students saw the point of addressing the issue of accuracy and students and therefore *did* experience what they learned by answering this sub-question as a contribution to answering the driving question, only *afterwards*.
One group presented their findings and judgment to their classmates. They had judged their water as clean enough to swim in, although one of their test results was just below the norm. A classroom discussion emerged whether you could take such a risk: because the test methods are not so accurate, the *real* value of this parameter might very well exceed the norm. To collect evaluation data, I asked the students, in three somes, a few groups per class, about the section on accuracy in evaluative interviews.

During the interviews, they had their own workbook near at hand and could glance through it to recollect what they actually had done in the section. None of the students could adequately explain why the section was part of the module, what they thought it was about, what they had learned from it, let alone how, if at all, they used what they had learned from this section in judging their water. It can be concluded that the moment this sub-question was addressed, students did not experience what they learned by answering the sub-question as a contribution to answering the driving question (only *afterwards*).

6) Can we apply our knowledge to an example?
Two things were to be achieved with this sub-question: students were to gain experience with the steps involved in judging water quality, and with the actual tests.
All teachers, due to time pressure, skipped this sub-question. The first six questions had taken two lessons instead of one, and we foresaw that the testing of the water would take more than the one lesson we had planned.

7) Is our water clean enough for its function?
This was actually the driving question in which students were to apply what they had learned. The question was only partly answered.
Firstly, when doing the tests, I observed that students were generally very insecure when doing the 'black box' tests. They asked many questions and seemed to need a lot of reassurance. Students found it difficult to estimate concentrations by comparing the colour of their treated samples with the calibration series, using their eyes as instruments. In fact, these were doubts with respect to the accuracy of the tests. I found that also much insecurity was raised with respect to the tests.
In their written and oral presentations (see also sub-question 8A), all students involved the water function in their judgment of the water quality. As far as 'relevant parameters and norms' are concerned, it was sometimes difficult for students to explain their choices. As I mentioned before, the water functions were not always distinct, or easy to determine. Nevertheless, when the water function was established, they generally could easily find a list of standard parameters and norms for different water functions. In addition, the surroundings and circumstances of a water site were sometimes just too complex to determine what would be relevant parameters (for the students but also for the teacher and me). Not surprisingly this complexity somewhat killed off the class discussion. Based on this I concluded that, although the driving question was relevant to students, I had underestimated the complexity of its details.

8. Did we adequate judge the water quality of the water in our neighbourhood?
This sub-question was to make students reflect on 'What makes a judgment about water quality adequate?'.
Not surprisingly, in light of the previous findings, this sub-question was not often raised and answered as intended. Only a few of the 'critical classroom discussions' emerged around this sub-question. The different groups would present their findings and the teacher would ask questions. This pattern of communication actually followed the rules of traditional school setting: the teacher asks questions to find out if the students 'understood' and 'performed ok': students answer and the teacher evaluates

(see section 3.3). The moment when one group presented their findings with 'swimming water' they had judged as good enough (see 'accuracy of the test method'), this sub-question became central in a class discussion, which involved the accuracy of the method.
Sometimes students did express their doubts with respect to the list of reference parameters and norms they had established. This had two reasons. The first was, as I mentioned before, that students sometimes had difficulty with establishing the water function of a water site (it was not always very distinct), although most groups simply *decided* on a function. As a result, they were uncertain about what would be relevant parameters. Secondly, students tended to be uncertain about whether their list of standard parameters really covered what their water might contain in reality. To conclude: students had not experienced what they had learned by answering this sub-question as a contribution to an answer to the driving question.

9. What is the common procedure for testing and judging water quality?
This question was to make students explicitly realise that they had all actually applied a common procedure for testing and judging water quality. In other words: what have we learned?
This final question in the reflection phase 'What is the common procedure for testing and judging water quality', did not become relevant to the students, as they felt that they already completed the purpose of the lessons by answering the driving question 'Is the water of our neighbourhood (actually: some specific water site) clean enough?'. The common procedure never emerged in the reflection phase. One teacher finally explicitly asked students for the procedure, but students seemed to find the content of the procedure itself too obvious to bother.
As a result the criterion for this part of the *attention for student input* characteristic (b) was not met: students could not have felt that their input (presentations) contributed to to achieving the learning goal of making a common procedure explicit.

Summary and conclusions

The findings led to the following conclusions:

1. There were no indications at the time that the characteristic *context* needed rethinking. Students were generally motivated by the driving question, clearly considering it an important issue. They worked enthusiastically.
2. There were also no indications at the time that part of the characteristic *attention for student input* needed rethinking. Students appreciated the autonomy of choice as intended. However, concerning part b, the final class discussion did not unfold as intended and students could not have experienced that their input (their presentations) contributed to achieving the learning goal of making a common procedure explicit.
3. The operationalisation of the *need–to-know* characteristic proved to be inadequate: students generally had not experienced what they learn by answering the sub-questions as contributions to answering the driving question.
   The problem was that the sub-questions, which framed the sequence of learning activities, had not emerged from their own experiences, that is: the designed learning activities did not achieve among students that they appreciated the

questions and their sequence as logical and relevant in light of the driving question. In addition, some sub-questions turned out to be too complicated. The content of the design was far too ambitious for such a short module.

4.4.4 Implications for the second version

As described above, based on the findings with the first version three major problems came to the fore. These were the problems of the complexity of the design, of the inadequately designed *need-to-know* characteristic and of the fact that the procedure for judging water quality had not been an integrated part of the teaching-learning process. In section 4.4.4, I will discuss the implications of these major problems for the second version.

Need-to-know in light of a problem-posing approach

I considered the use of a problem-posing approach (Klaassen, 1995) for the second version of the design, because the design needed a higher didactical quality specifically for the *need-to-know* characteristic. A problem-posing approach basically means that teaching learning activities are designed in such a way that students are put in a position in which they feel the need and see the point to extend their knowledge in a certain direction, in light of their broad motive to answer the driving question. In the first version of the design, this was clearly not achieved (see previous section).
Klaassen (1995) refers to this 'knowledge need' as the development (among students) of *content-related motives*. The challenge is to design a learning activity in such a way that it builds on the prior one and induces a content-related motive in the direction of the next learning activity, and so on. In the designed learning activities proper use should be mad of the already existing (largely correct) intuitive notions students have of what is involved in, in this case, judging the quality of a water sample. A problem-posing approach is actually nothing more than a way of emphasising that in the design and in the evaluation of the teaching learning sequence special and detailed attention is paid to the creation of such content-related motives among students. Both the work of Kortland (2001) and Vollebregt (1998) involved designing courses from a problem-posing approach, which was introduced by Klaassen (1995).

I have built especially on the work of Kortland as the course in his research addressed a learning goal ('Making an adequate decision about the waste issue') which is closely related to that in my study. The design study of Kortland resulted, similar to that of Vollebregt, in a so-called didactical structure. Lijnse and Klaassen describe such a structure as follows:

> The main steps to be taken and stages to be gone through by teacher and students as derived from the final scenario are reflected upon and summarized in what we may provisionally call a possible didactical structure for the topic at hand.
> (Lijnse & Klaassen, 2004, p541)

Such a structure can be represented by a figure such as figure 4.4 (and 4.6A and B). Based on Kortland, there are three columns: an issue knowledge column, a content-related motive column and a procedural knowledge column (Kortland labels this

column as a ‘skill’ column, referring to the development of the decision-making skill). The structure solely reflects the content-related progression of the teaching learning sequence, thereby making the intended content-related motives explicitly visible in the middle column. These motives embody the problem-posing approach. They should logically follow from activities of which only the content-related progression is described. The structure should read as a story line. If not, it might indicate that something is wrong and the structure needs rethinking. This way it serves as a design tool, but it also helps to evaluate and reflect on what happened in the classroom, indicating the moments were students possibly did not follow the intended story line.
The didactical structure can also be characterised by a succession of phases, which ‘relate to particular didactical functions that have to be fulfilled in such a way that they assure the necessary coherence in the activities of the students’ (Lijnse & Klaassen, 2004, p244). Kortland (2001, p178) distinguishes six such phases in his teaching sequence:

- Phase 1: orientation on a topic and evoking a global interest in and motive for a study of the topic at hand.
- Phase 2: narrowing down this global motive into a content-specific need for more knowledge.
- Phase 3: extending the students' existing knowledge, in view of the global motive and the more specifically formulated knowledge need.
- Phase 4: applying this knowledge to situations the knowledge was extended for.
- Phase 5: creating, in view of the global motive, a need for reflection on the skill involved.
- Phase 6: developing a (still possibly contextualized) metacognitive tool for an improved performance of this skill.

The functional descriptions of Kortland’s phases reflect the problem-posing approach, and the successive phases can be seen as a rough version of the didactical structure (see figure 4.4):

> We remark that phases 2 and 5 represent one of the main points of a problem-posing approach. Such phases appear not to be present in the teaching cycles as published in the literature (Abraham, 1998). Those cycles almost exclusively deal with cognitive learning, even though it is also often written that one should not forget about the importance of motivation. In our [incl. Kortland’s] approach, however, both are taken together and integrated from the start.
>
> (Lijnse & Klaassen, 2004, p544)

Kortland’s distinction between successive functional didactical phases serves as a design tool because it helps to determine which functions might be relevant and how an activity might contribute to the respective function. It also serves as a framework for analysis of the realised teaching-learning process: are the functions fulfilled? Moreover, did that contribute to the criteria of meaningful chemistry education as intended and expected?
As can be seen in figure 4.4, for the second version of the teaching-learning process of the water quality course four of such phases inspired by Kortland were distinguished. Kortland’s phases 4 and 5 were made part of phase 3, as applying the knowledge and

creating a need for reflection can be considered part of ‘extending knowledge until a satisfactory procedure for judging water quality is achieved’.

The first phase is the ‘orientation and general motivation phase’. The function of this phase is to generally motivate students, which is a condition for their intended learning process. Students are to orientate themselves on the driving question of the lessons in such a way that they become generally motivated to get a deeper insight of what is involved (in order to answer the driving question).
In *the second phase*, a more specific knowledge need is to be induced among students, that is, a content-related motive in light of the broader motive. Students apply their intuitive notions of ‘what is involved’ to an example and experience ‘What they do not know, but need to know’ in order to answer the driving question.
The third phase is called ‘the extension phase’. Students are now to extend and reflect on their knowledge of the procedure for judging water quality, until the procedure is ‘satisfactory’. As a result, phase 3 is the most extensive phase in the teaching-learning process.
In *the fourth phase*, the reflection phase, students explicitly address the procedure, which steps are now ‘summed up’, and the question should emerge whether the procedure is applicable to a different exemplary problem. This way they are expected to reflect on what they have learned: a satisfactory procedure.

In the section 4.5 I will present an argued description of how this problem-posing approach was elaborated upon in each of the functional didactical phases of the teaching-learning process and what the expectations were.

The complexity of the design

In order to address the complexity of the first version, it was decided that students would all focus on judging the water quality of an exemplary, specific water *function*, instead of judging the water quality of a water *site*. This had several advantages: students did not need to establish the water function of some water site themselves (which several times proved to be not so straightforward). In addition, because students were explicitly asked to focus on an exemplary water function, the procedure for judging water quality was expected to become a more natural learning aim and actually the focus of attention throughout the teaching/learning process. I expected that it would be more natural to reflect in phase 4 on the question ‘Is this procedure, which we applied in this example, useful for judging the water quality of *other* water functions?’. Moreover, to further address the problem of complexity, the striking (I expected) example of drinking water quality was chosen. Students, I thought, did not need to worry about factors in the surroundings, like farming, that might influence the water quality and they do not have to worry about a *representative* water sample site either. Also, the question ‘What are the relevant parameters and norms?’ I thought to be less complex because of this choice for an exemplary water function. Not only did all students now address this question for the same water function, it was also easy to use the existing practice of testing drinking water quality as an information source: what is tested in reality and why?

Obviously, there is one big problem with the choice for the example drinking water quality: if tap water is to be judged, what is the challenge? To solve this problem, it was decided that students would *produce* drinking water from surface water they chose themselves first (I expected the students to have basic notions of such a production process, as they all had addressed the topic in a previous module). This way the question 'Is this water clean enough to drink?' would become, I thought, much more exiting, addressing students personally. It also seemed to have advantages for the *need-to-know* and the *attention for student input* characteristic. First, the question **'**Is this water clean enough to drink?' could be used in a *need-to-know* characteristic to refine the followed procedure and the usefulness of the issue knowledge (raising question like: What information do we need? How can we get this information? Can we answer the question now or do we need to take a next step? Etc. (see below). Secondly, it partly re-established part-a of the *attention for student input* characteristic, in the sense that students were still able to choose their own surface water to produce drinking water from, to ask them how they would do this and to let them present their own findings.

Integration of the procedure

First of all, in the second version of the design the students' intuitive knowledge of procedure steps was used to induce knowledge needs: they knew that the water should be tested, only not on what things. It was decided to include regular summarising activities in the course. In these activities, the students were to summarise what they have previously learned as a next step of the procedure. This was expected to be a logical set up for students since they have been asked to focus on the example of drinking water quality in order to gain insight in how the question 'Is the water clean enough' could be answered.

So, two ideas were expected to contribute to the aim that the procedure would be a for students functional part of the reflection phase. Firstly, the idea of focusing on an example. This would almost obviously lead, I thought, to the question 'And to what extent can we use this for other examples?'. Secondly the idea of adding specific activities in which the steps of the procedure were explicitly summarised.

The procedural steps of testing and judging water quality in the second version (see figure 4.3) consisted of the same steps as the first version (see figure 4.1). The only difference was in 1) the filling in of the steps, which is now simplified and 2) the accuracy of the test method step, which was now postponed until after doing the tests (see below 'need to know: design and expectations').

Figure 4.3 An overview of the second version of the procedural steps of testing and judging water quality

Judging water quality
Determine the water function *(drink)*, this determines the demands
Determine the demands *(safe)*, this determines the criteria
Determine the criteria: what parameters and norms should be tested?
Measure parameters
Compare test results with criteria
Reflect on the accuracy *(doubts)* of the test methods. Reconsider judgment with the accuracy of the test method in mind
Formulate your judgment *(including an account of the accuracy of the test method)*

4.5 The second research cycle: design, findings and implications

In this section, I will present in broad outlines the second research cycle. This section is structured as follows. In section 4.5.1, the second operationalisation of the three characteristics of meaningful is presented. Next, in section 4.5.2, the design is presented in broad outlines, including argued expectations as to how the criteria for meaningful chemistry education were expected to be met. The main findings of the evaluation are presented in 4.5.3 and finally, in 4.5.4 the most important implications are discussed (that is: those that concern the operationalisation of the three characteristics of meaningful).

4.5.1 The characteristics of meaningful in the second version

In this section, I will present how the three characteristics of meaningful were operationalised in the second version of the design.

Context

As I explained in section 4.4.3, the first research cycle did not give rise to a fundamental rethinking of the interpretation of the characteristics *context*. Therefore,

in the second version the context characteristic was (similar to the first version) operationalised as a relevant and recognisable driving question that was expected to appeal to students.
In this case the driving question was: Is the water clean enough?
The additional 'in our neighbourhood' was left out in this version, as all students were now to produce drinking water from surface water and were to test whether their product meets the criteria for drinking water. They were not testing and judging surface water in the neighbourhood to see whether the water meets the criteria for its intended use.

The *context* characteristic was evaluated based on the criterion (see also the first version): did the driving question appeal to students in the sense that they related the driving question to their daily life and were they broadly motivated to answer it?

Need-to-know

This characteristic was operationalised as follows. Within an overall broad motive (corresponding to the driving question) a series of connected and nested content-related motives should be triggered among students. This should be done by using the students' intuitive knowledge of what would be the next logical step of the procedure and by using the quality control question to induce the need for a next necessary procedure step. This way the students' content-related motives will frame the sequence of activities and the procedure can be expected to become a truly integrated part of the teaching-learning process.
In figure 4.4 the didactical structure of the second version of the design is presented: a problem-posing story line in which the intended content-related motives are made visible.

The *need-to-know* characteristic was evaluated based on the following criterion: were the intended content-related motives triggered at the intended moments in the teaching-learning process among students in light of their broad motive to answer the driving question?

Attention for student input

As I explained in section 4.4.2, the first research cycle did not give rise to a fundamental rethinking of the interpretation of the characteristics *attention for student input* either. However, because in a problem-posing approach much more attention is paid to students' content-related motives, it naturally follows that teachers should also pay explicit attention to these 'knowledge needs' at the level of students' remarks, as an additional way to let students feel that their input matters.

This characteristic was therefore operationalised as follows in the second version:

a. Students have certain autonomy of choice in particular the surface water from which they are to produce drinking water.
b. The input of all groups is necessary to achieve the learning goal of making the common procedure for judging water quality explicit. Students develop the insight that each group follows the same procedure and realise that all groups contribute

to this insight throughout the teaching-learning process. This insight is made explicit in the concluding reflective activity.

c. The teacher should pay attention to the student input, as far as it is an important driving force of the teaching-learning process, e.g. when students, at certain points in the teaching-learning process, express their content-related motives as questions.

The characteristic *attention for student* input was evaluated based on the following criteria. Do students feel that their input matters in the sense that:

a. they experience and appreciate that they can make their own choices?
b. they experience throughout the teaching-learning process that their input contributes to achieving the learning goal of making a common procedure explicit?
c. they experience that their input, in the form of their content-related motives, is now an important driving force in the teaching-learning process?

4.5.2 Design and expectations

The design was adjusted from the perspective of a problem-posing approach, thereby making use of those findings with the first version, which indicated which motives could potentially be triggered at which moments. In this section I will argue, based on the findings with the first version, why I expected each functional didactical phase of the designed teaching-learning process to trigger which content-related motives among students at what moments. Phase 3 covered the most extensive part of the teaching-learning process, as the students were to extend their knowledge until a 'satisfactory procedure' was reached.

Table 4.3 shows how the functional didactical phases each address (series of) questions, which were now expected to be posed by the students themselves.

Table 4.3 The successive functional didactical phases and the successive questions

Functional didactical phase		Question
1	General orientation on the theme raises a motive to find out about what is involved in testing and judging water quality….	What is the lesson series about?
		What is the driving question and, roughly, how are we going to answer this question?
		Which relevant experiences do we already have with judging water quality?
2	Applying the existing, intuitive notions of 'what is involved' in a specific exemplary situation induces a knowledge need in a certain direction.	Did we produce drinking water quality?
3	Extending knowledge in the direction of the induced knowledge need.	What stuff and how much of it is allowed in drinking water?
		What parameters should we test and how can we test them? (what is an appropriate procedure?)
		What are the test results? (is the water clean enough to drink?)
		Do we trust the test results? (is the water clean enough to drink?)
4	Reflection on the procedure for testing and judging water quality.	In what sense can we apply the followed procedure in another situation of water quality judgment (like: is the water quality of a ditch clean enough for its ecological function)?

Phase 1: general orientation on the theme raises a motive to find out about what is involved

In this phase students are to orientate themselves on the topic water quality, on the driving question 'Is the water clean enough?'. They are to get motivated to find out about a procedure for judging water quality and make their intuitive notions about the relation between water quality criteria and water function explicit as a first step of such a procedure.

After the teacher introduced the theme water quality, emphasising its importance by letting students think about how they use water, students were to think about their personal water quality criteria for two water different functions: swimming water in their neighbourhood and drinking water. This was expected to structure their intuitive notions of the relation between water function and water quality criteria, to be summarised as the first step of the procedure in a concluding activity. Students were to do and discuss these first few learning activities in groups before the results would be evaluated in a class discussion. The idea was that the question (and content-related motive) ‘What is involved in judging water quality?’ would naturally follow from the driving question ‘Is the water clean enough?’ (if you want to know if the water is clean enough, you need to know how you can judge the water quality).

I expected all this to lead to a more structured orientation phase than the introduction phase of the first version. In the first version students had, after the introduction of the driving question, immediately moved to choosing a water site and the question ‘What should we test?’, skipping questions like ‘What steps are involved?’ and ‘What things do you need to take into account?’. I expected the summarising activity, in which students were to formulate the first step themselves, to contribute to the students focusing on finding a procedure for judging water quality in order to answer the driving question.

Phase 2: applying the existing and intuitive notions of ‘What is involved’ in a specific exemplary situation induces a knowledge need in a certain direction.

In this phase, students were to focus on the example of drinking water. A knowledge need in the direction of ‘drinking water parameters and norms’ was to be raised among students: what and how much of what is allowed in drinking water?

Now that the goal was ‘to arrive at a general procedure for testing and judging water quality in order to be able to answer the driving question’, I expected that the teacher could easily make it plausible for students that it would be a good idea to focus on an example. So, to arrive at such a procedure, the teacher was to introduce to students the example of drinking water quality. This also created an opportunity to pay explicit attention at what had proved to be too difficult for students: to find out what parameters and norms are relevant to test their water on (see below for a more detailed description of how this issue was addressed).
For reasons I explained in section 4.4.4 ‘The complexity of the design’, groups of students were asked to produce drinking water from surface water they could choose themselves. The first knowledge need, ‘What and how much of this is allowed in drinking water?’ was expected to be raised when students were asked whether they trust their water to be of drinking water quality, and, if not, what specific information they felt they need.

Phase 3: extending knowledge in the direction of the raised knowledge need

In this phase, students were to extend their issue and procedural knowledge until they arrived at 1) the answer to the driving question and 2) a satisfactory procedure for judging drinking water. With respect to the procedure (2), the question had to be raised to what extent the procedure could be used for a different water function (of which drinking water was an example).

In this phase, the students extended both their issue- and their procedural knowledge until they reache a satisfactory judgment because of a satisfactory procedure. Two types of content-related motives were expected to drive the progression of the extension phase.

1. The need for more issue knowledge to be able to complete a step in the procedure, and:
2. The need for further procedural knowledge to be able to answer the quality control question ‘Would you now drink the water, do you think it meets the criteria?’.

In the first instance, the issue knowledge of the students was to be extended in the direction of that knowledge need that was raised in the previous phase: ‘What and how much of what is allowed in drinking water?’ This ‘knowledge need’ was expected to functionally introduce the next learning activities which addressed the legal list of drinking water quality parameters and norms. In these activities students were to work and discuss in their groups a selection of drinking water parameters and their norms. They had to look up in information maps the reasons why a parameter is on the list and why its norm is as it is. The parameters were selected in such a way that they covered different kinds of reasons. Some parameters concerned substances that are poisonous in very small concentrations, some in very large. Of some parameters, the experts are unsure how poisonous they are. Others, like calcium, are healthy and have a minimum norm. Some parameters, like iron, are on the list because they make the water look bad and taste bad. Next to chemical parameters there are microbiological parameters etc.

After students had learned about the legal list of drinking water quality parameters and norms, they were asked whether they thought they should test all these parameters. I expected students to realise that, although this obviously would be the best thing to do to be completely sure, it is impossible. As a result I expected students to pose the question ‘What parameters should we test and why?’. This issue knowledge-need is then simply satisfied by presenting the four standard parameters and norms by which drinking water quality is actually tested. Subsequently, I thought, this immediately would raise questions like: ‘Why these parameters? Why not others?’

Students were then expected to learn from the existing practice that these four give a *reasonable enough* indication of the chemical and the biological drinking water quality. And of course, this implies certain built in doubts. To emphasise this point, such doubts were explicitly discussed next, with the question ‘Why do they not test drinking water on the very poisonous mercury?’. This series of learning activities, concerning the parameters and norms of drinking water, resulted in the next summarising learning activity. In the summarising activity, the procedure is refined with ‘Establishing relevant parameters and norms’.

The next learning activities addressed the actual testing and judging of the water samples. Students were asked to write down a plan for judging water quality. They were asked to write it down in such a way that a third person would be able to understand what steps they planned and why. I expected them to write down and briefly explain the first two steps, also adding the steps ‘test the parameters’ and ‘compare the test results with drinking water norms’. I expected this ‘summing up’ of the steps to occur naturally, without leading to a refinement of the procedure.

When doing the tests, I expected that, like in the first version, several doubts concerning the accuracy and the reliability of the tests would be raised and that this would sometimes influence the different groups in the judgments they had to formulate based on their findings.
The final learning activities of this phase concerned a section which included the topics accuracy of the colorimetric test method and reliability of black box tests ('How can you be sure whether these black box tests really test what you want them to test?). The introduction of this section, both by the teacher and in the text, was supposed to link the raised doubts specifically to the topics accuracy and reliability. The learning activity on accuracy was unchanged: first students had to put the solutions of a calibration series of copper sulphate solutions in the right order with their groups. Next, after their calibration series was checked, they had to estimate the concentrations of two unknown solutions. I expected that this would lead to a variety of answers, illustrating the accuracy of the test method, similar to the first version. Students were finally asked whether they would like to reconsider their judgment and the teacher was to discuss this with the students.

In a next activity, students were asked whether they trusted the black box tests to test what they should. The students had to think about a way to 'test the tests'. After this, they were again to reconsider their judgment in a class discussion.
So now, in this second version, the topic of accuracy of the test method and reliability of the black box tests was addressed *after* doing the tests, combining it with the doubts raised when *doing* the tests, in the process of forming a judgment. I expected that in this way, the issue of accuracy would be truly rooted in relevant experiences of students (namely linked to the actual tests). They had to conclude each learning activity in which they thought about their doubts in relation to these issues with a *reconsideration* of their judgment of the water quality, instead of thinking about accuracy of the tests first and involving this topic in formulating a final judgment. In other words, the procedure included the step *reconsideration* of the water quality judgments in light of the issues 'accuracy of the test method' and 'reliability of the tests'.

At this point, after the students had reconsidered their judgments in light of what they had learned about accuracy of the test method and reliability of the tests, the students would have completed, I thought, the steps of a 'satisfactory procedure for judging water quality for the example of drinking water'.

Phase 4: reflection on the procedure

In the reflection phase, the students had to reflect on what they had learned by finding out if and to what extent the common procedure for judging water quality can be applied to other cases of water quality judgment.

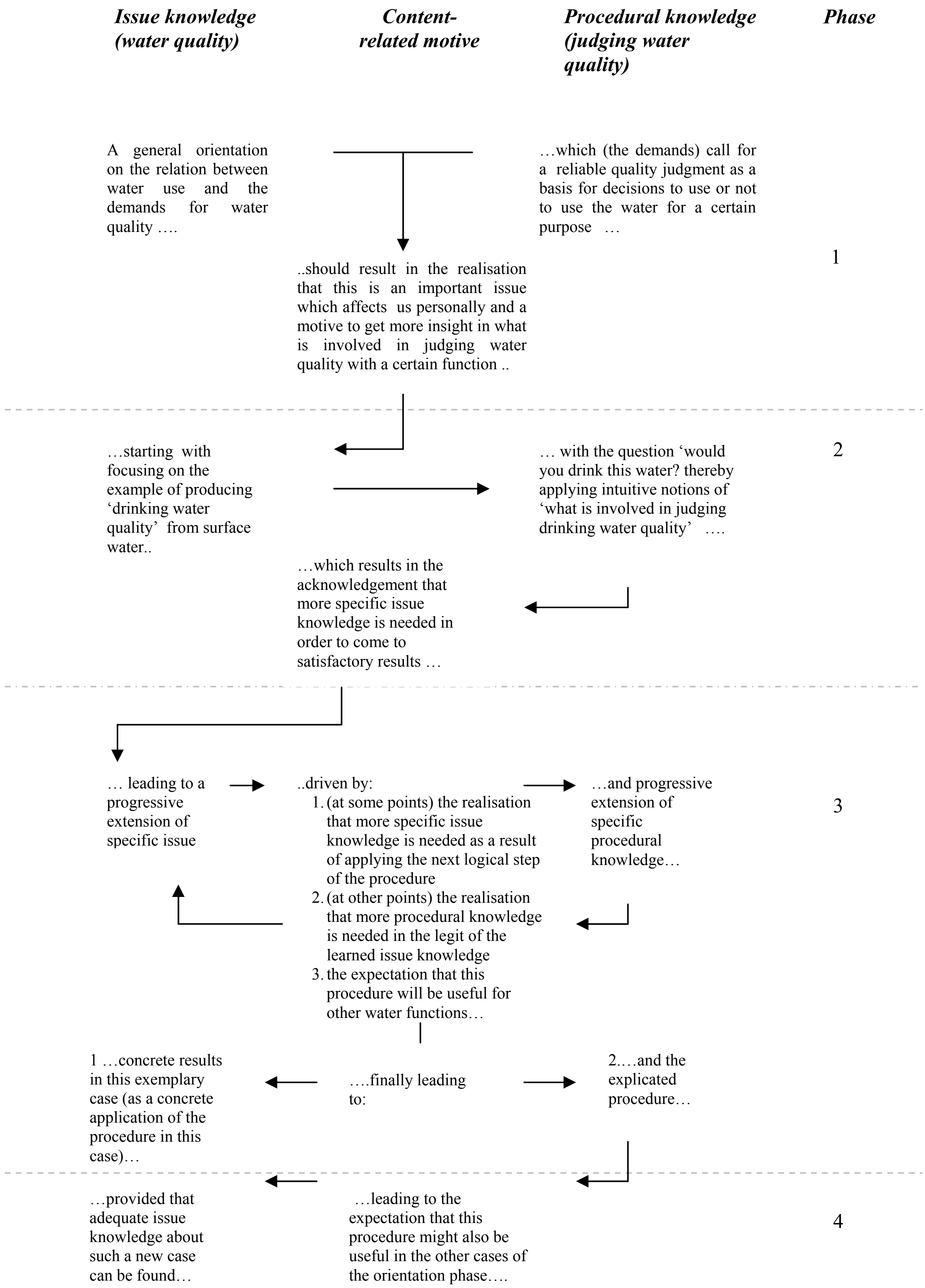
The teacher was to remind students explicitly that they now had answered the driving question for the *example* drinking water and thereby had explicitly developed a procedure. Finally, students were asked whether they can use the procedure to a different example: Is the water quality of this ditch clean enough in light of its ecological function? The students were to discuss the following questions in groups:

A. *How would a plan for judging ditch water with an ecological function look like?*
   Such a plan would have to consist of the following steps:….
B. *Does this plan differ from the plan you made for judging drinking water?*
   No, because
   Yes, it differs on the following points (explain your answer)…

This learning activity was to be followed by an evaluative class discussion, directed by the teacher in the direction of 'What is a common procedure for judging water quality?', which I expected to emerge naturally.

Because of such decisions, the general didactical outlining of the second version of the design differed considerably from the didactical outlining of the first version. This was mostly a consequence of designing a teaching-learning process according to a problem-posing approach. The thematic/basic knowledge and skills format was now replaced by a structure around content-related motives. The new structure did not discriminate between 'thematic knowledge' and 'basic knowledge'. As mentioned in section 4.4.4, figure 4.4 shows the general didactical structure of the content-related progression of the learning activities of the second design. It should now read as a story line, in which the content-related motives are a driving force of progression: there must always be a *reason* for students to take the next step.

Figure 4.4 The didactical structure of the second version of the design



4.5.3 Evaluation: the main findings

I will first present the overall results of the second trial, before describing the evaluation in detail.
In the second trial there generally was an improvement in the didactical quality with which the *need-to-know* characteristic was realised. Students were broadly motivated by the driving question (like in the first version) and the sequence of teaching-learning activities gave more rise to certain intended knowledge needs (content-related motives) and questions among students. Based on this I draw the conclusion that the *need-to-know* characteristic was improved and that the teaching-learning process was more (but certainly not always) adequate with respect to giving rise to the intended learning processes of the students. There still were some problems, however. Some of these problems concerned *the relation between* the characteristics *context* and *need-to-know.* Others concerned the *attention for student input.*
With respect to the relation between the *context* and the *need-to-know* characteristics, it turned out that the idea of a driving question as providing for a *context* did not sufficiently direct the *need-to-know* characteristic. It did give enough direction to what content-related motives might be triggered at which moments. As a result, the second version of the design still developed along too many lines, although it was less complex than the first version. This became, for example, apparent in the mixing up of the two practices ‘drinking water production’ and ‘water quality judgment’, which confused the students (see below).
Another problem concerned the procedure for judging water quality. Although the procedure was expected to have been designed as a leading theme to be explicitly developed in ‘summarising activities’ (this had been the aim) in the second version, it turned out that it still had not become an integrated part of the teaching-learning process for students. Students considered the summarising activities as something different and the concluding activity of ‘making the common procedure explicit’ a redundant and therefore irrelevant (and certainly not functional) activity.

Students still appreciated the autonomy of choice in the teaching-learning process (part-a of this characteristic, section 4.5.1). The fact that the procedure still was not an integrated part of the teaching-learning process obviously resulted in part-b of the characteristic *attention for student input*-criterion not being met: students did not experience that their input contributed to achieving the learning goal of making the common procedure explicit.
Also, the characteristic *attention for student input* clearly was not elaborated upon structurally and with enough detail with respect to part c: the teacher should pay attention to the student input, as far as it is an important driving force of the teaching-learning process, e.g. when students, at certain points in the teaching-learning process, express their content-related motives as questions. This became clearly visible in this second trial because now the *need-to-know* characteristic was designed on a detailed level. Teachers did not properly address the input of students when introducing and evaluating a unit of learning activities. The teachers were not used to this role and had not been prepared for this, because it just had not been an important issue before.

Context and need-to-know

In this section, I will give an overview of the main findings with the successive didactical phases, which have led to a rethinking of the relation between the *context* and *need-to-know* characteristics and to the most important design changes for the third version (see sections 4.5.4 and 4.6)

Phase 1: general orientation on the theme raises a motive to find out about what is involved

In this phase, students were to orientate on the topic water quality, on the driving question ‘Is the water clean enough?’ and on the structure and purpose of the course. They were to get motivated to find out about a procedure for judging water quality and make their intuitive notions about the relation between water quality criteria and water function explicit as a first step of such a procedure. The functions of this phase were partly fulfilled.

Firstly, in this phase, it proved to be difficult to determine whether the students were motivated by the introduction of the driving question and by the idea that they were to find out about ‘What is involved in judging water quality’. Similar to the first trial, the teachers introduced the lesson series in one long monologue, without involving the students or checking whether they appreciated the introduction as intended. This time, however, the students appeared much more expectant, unlike the immediate enthusiastic reaction of the students involved in the first research cycle, when they could hardly wait to get water samples and test them. All teachers, after briefly introducing the theme ‘water quality’, spent most time on explaining procedural issues, like: the instruction format and organisation of the learning activities (group work and class discussions). At a later stage, however, students showed enthusiasm for the driving question and the activities they were involved in.

Secondly, almost all students used their experiences in the group activities where they had to discuss the relation between water function and water quality criteria. In general, they easily made the first step of the procedure explicit: you have to determine the water function of the water because this determines the water quality criteria.

Phase 2: applying the existing and intuitive notions of ‘What is involved’ in a specific exemplary situation induces a knowledge need in a certain direction.

In this phase, students were to focus on the example of drinking water and a knowledge need in the direction of ‘drinking water parameters and norms’ was to be raised among students: What and how much of what is allowed in drinking water? This content-related motive was not induced and therefore the function of phase 2 was not fulfilled.

Students did raise the qualitative question ‘What *is* in the water?’, when confronted with the question whether they would drink their purified surface water instead of the intended question ‘What and how much of this is allowed in drinking water?’

Later, I concluded that the factual induced knowledge need referred to the procedural step ‘Test the water sample’. From the perspective of the students, this was a logical step to think about, more than ‘looking into the legal list of water quality parameters and norms' (phase 3). Moreover, this was actually not a step of a characteristic procedure by which drinking water quality is tested and judged. In one class, when the students were asked whether they would drink their purified water, some students called out: "Yes why not?". The teacher, however, ignored this and directed the line of thought towards "You should not drink it, it might not be safe as filtering the water does not take out the solved substances". By doing this, the teacher communicated to the students that this question was not about their *opinion*, but for the teacher to check whether students knew the right answer. In the school setting this is a very common pattern of communication (see section 3.3) but now it interfered with what was intended. After this, the teacher explained the difference between solutions and suspensions, thereby testing the students' knowledge through questions. Again, this is a very common pattern of communication between teacher and students, in which students are given the message 'Pay attention, this is important'. It was not, in fact, important in the flow of the water quality lessons and it only interfered with what was intended: paying attention to *student* input as a driving force of the process.

In addition, looking back, different *context*s seemed to become mixed up. Students had to *produce* drinking water from surface water. The following problem arose from this: they had chosen different strategies to purify their surface water. This made drawing conclusions for them about the drinking water quality more complex, because arguments related to the quality of the production procedure also began to interfere with arguments related to the quality of the testing procedure. In one class, some students even tried out different strategies to purify the surface water, of which they wanted to compare the results. Furthermore, it took students considerable time to purify the surface water, making it difficult for them to link what they were doing to the driving question 'How to determine whether the water is clean enough to drink (in this case)?'. Students tended to adjust the driving question to 'How to make water clean enough to drink?'. Moreover, because the students had used all different kinds of surface water, and obviously did not have the means to purify this water professionally, they did not think that the four standard parameters, by which real drinking water is almost constantly monitored and which are used to distinguish drinking water from household water, were adequate. And they probably were not. At the time, when analysing these findings, I did feel this shift of emphasis towards drinking water production to be a problem without being able to analyse it as a mixing up of *context*s or *practices*. This idea of analysing the design in terms of practices emerged afterwards, when the problems of complexity of the design and the procedure not being a main theme were addressed in designing the third version (see 4.6). Back then I just concluded that in a next version 'the evaluation of drinking water quality' needed to be put in to the fore and the production process to the background.

Phase 3: extending knowledge in the direction of the raised knowledge need

In this phase, students were to extend their issue and procedural knowledge until they arrived at 1) the answer to their driving question and 2) a 'satisfactory procedure for judging drinking water'. The question was to be raised as to what extent the procedure could be used for a different water function (of which drinking water was an example).

The function of this phase was only partly fulfilled, for different reasons.

Firstly, the learning activities in which students learned about the legal list of drinking water quality parameters and norms did not satisfy the induced knowledge need in the previous learning activity (see phase 2). The activities therefore formed a break in the teaching sequence. Students seemed to be more interested in what was to be the next question: 'What should we test the water on?'. After the students learned about the legal list of parameters and norms, the question 'What should we test?' and the realisation that this could not be the whole legal list of parameters and norms emerged again (only not for the first time, as intended), in a natural way. In general, students easily appreciated the information on what is tested in reality and why (four parameters, because they are good indicators of the chemical and bacterial quality of the drinking water). Some students found it dubious and inappropriate that only four parameters are tested. Generally, the students immediately recognised the functionality of this information. These four parameters and norms apparently sufficed and they used this information for their own drinking water quality judgment.

The learning activities in which students were to summarise the output of the prior learning activities explicitly in a next step of the procedure seemed to them somewhat distinct from the other activities. I concluded this from how they addressed these learning activities. Students would utter things like 'what do 'they' [*implying 'the authors'*] mean', indicating, I think, that they did not understand the purpose of the activity. Sometimes they would just fill in completely different steps from the ones they had taken themselves in the previous learning activities. Moreover, in the activity in which they finally had to summarise the steps before testing the water, they would sometimes do this in an incomplete manner. Sometimes they would add new steps. The procedure for testing and judging water quality, I concluded, still was clearly not a naturally integrated part of the teaching learning sequence and the students apparently had not adopted that arriving at a common procedure for testing water quality was a worthwhile learning aim.

Several doubts were raised when doing the tests, although generally less than in the first version of the design. Moreover, students did seem to connect their doubts that came up when doing the tests to the issue 'accuracy of the tests': they generally realised and could explain how it might have consequences for their water quality judgment. The teachers had collected and made an inventory of the test results and the doubts of all the groups on a poster. The teachers discussed these doubts as intended: paying real attention to the input of the students without quickly moving on to the topic 'accuracy of the test methods'. However, students seemed to consider the issue 'Do we trust the black box tests?' irrelevant ("Why would they make bad tests?"). They generally did not relate this issue to the doubts that came up when doing the tests. I concluded that I had misinterpreted some of these doubts. Students did not feel insecure about the tests, but about their own performances.

Phase 4: reflection on the procedure

In the reflection phase, the students had to reflect on what they had learned by finding out whether and to what extent the common procedure for judging water quality can be applied to other cases of water quality judgment.
The procedure was not an integrated part of the teaching-learning process, most students still experienced the concluding learning activity in which they had to apply the procedure to a different situation (reflection phase) as redundant and irrelevant. They did it (and were able to in general), but did not seem to experience a need for doing it.

Attention for student input

In this section, I will give an overview of the main findings with the successive functional phases, which have led to a rethinking of part-c the characteristic *attention for student input.* The findings with the second version did not cause for rethinking the characteristic *attention for student input* with respect to the aspects of autonomy of choice and contributing to a common learning goal.

Part-c was: the teacher should pay attention to the student input, as far as it is an important driving force of the teaching-learning process, e.g. when students, at certain points in the teaching-learning process, express their content-related motives as questions. With respect to part-c I found that two of the three teachers involved tended to ignore the input of the students, except for some special moments we had discussed before (see preparation of the teachers). In addition, frequently, the teachers used the content as a means to control the class. As a result the motivation to get a deeper insight of certain topics shifted from 'wanting to know in light of the driving question' (through content-related motives) to 'important for the test'.
In addition, one teacher of school A frequently silenced the class with, 'These are important concepts you need to know', implying 'Be quiet and listen', which happened immediately. This teacher also tended to draw attention to the 'important concepts', and how they should be defined, rather than their function in the procedure for testing and judging water quality. He would frequently question students to see if they had 'learned the definitions', implying 'this is important'. Especially this teacher, but also the other two, would also at some points dictate the 'right answers', rather than discuss student's answers, even when the questions referred to opinions, and there was not a 'right answer', only better arguments. As a result, I found that students more than once corrected their perfectly good, but differently formulated, answers and replaced them with the teacher's answer. The students of the teacher who did this the most, tended to be much more insecure than the students were of the teacher who paid more attention to the student input and valued it more. They asked more frequently what was expected of them, they found it more difficult to rely on their own judgment and they were overall less involved in discussions and in giving their opinion.

In general, the teachers relied on communication patterns which are common in the traditional school setting and raise certain expectations among students. Both teachers and students are used to these patterns and expectations (see section 3.3). However, in a problem-posing approach, both teacher and students have different roles that ask for

different communication patterns. The input of students drives the process much more and the teacher should pay attention to this, drawing back a little, trusting that students are able to (and will) come up with the intended input.
In general, the teachers did not introduce and evaluate episodes of learning activities in a *structural* manner, using the input of the students: their answers, their findings or their opinions. Based on the uttered and sometimes visible indifference of the students, their doubts and the fact that they tended to lose track (like I described above), I concluded that as a result students were not as involved in the process as intended, and that this contributed to a decreasing feeling that their input mattered, at least much less than desired. From observations it appeared that the students of the teacher who did pay more attention to students input (only not in a structural way) *were* more secure and, for example, more actively involved in class discussions.

I also found (not surprisingly) that the format of the interaction between teacher and students (and between students) raised expectations among students about the complexity of the learning activities (see also below: phase 1). For example, in the orientation phase the students were to discuss first in small groups their personal criteria for swimming water quality and drinking water quality. After this the teacher was to lead a class discussion on this topic, with the purpose of reaching a consensus about a basic list of criteria for each water function and making explicit the intuitive notion that water quality criteria depend on the water function. However, this instructional format was unnecessarily complicated. Different groups of students quickly came up with roughly the same lists of water quality criteria for each water function (e.g. list like: drinking water has to be clear, not smelly and should not contain bacteria or poisonous stuff). Students did not see the point of an additional class discussion: the fact that water quality criteria depend on water function was not surprising to them. Because of this mismatch between complexity of the content-related progression and complexity of the instructional format, students seemed to be confused (and bored). My interpretation of their confusion and boredom is that they expected that the class discussion was meant to lift their findings to some higher level, instead of recapitulating what they already knew.
Next, to complement these conclusions, I will briefly address the teacher's role in each phase.

Phase 1

As I explained above, in the attention for students input section, the instruction format of the first learning activities (first a group discussion and then a class discussion), proved to be too complicated for the aim 'activating relevant pre-knowledge'. This confused students, as they seemed to expect that the discussion was to lift their group-input to a higher level.
One teacher discussed the summarising activity (the first step of the procedure: 'Explain why it is important to determine the water function first?') as "Important because an important concept plays a role here and it starts with an F [function]". Especially this teacher would ask students' explicit attention throughout the whole teaching sequence for the 'important concepts'. This teacher evaluated this activity by asking one student his answer, and then dictating the 'right answer', which a lot of

students wrote down next to their own, perfectly good but less structurally formulated, answers.

Phase 2

One teacher tried to direct the class, after the knowledge need ‘what is in the water’ was raised, to ‘how much of what?’. The students did not show this part of the intended knowledge need and the teacher just added it in his formulation of the summarising question ‘What and how much of what is allowed in drinking water?’. This is what the students wrote down (next to their answers). The teacher, who frequently explicitly asked attention for “important concepts”, dictated the question and immediately connected the ‘what’ to ‘parameter’ and ‘how much’ to ‘norm’.

Phase 3

After the students had worked for a considerable time on the activities, in which they were to find out about the legal list of drinking water parameters and norms, one teacher dictated all the answers of the activities to the students, without asking for theirs.

I draw the conclusion that these findings reflected problems that are more general. It made me realise the following:

A. Besides letting students test their own water and present their own findings, the third characteristic of *attention for student input* needs *structural* attention at the detailed level of teacher-student interaction within some instructional format.
B. The instructional format, which structures the interaction, should match with the complexity of the learning activity.

As far as the characteristic *attention for student input* is concerned, it became clear that this characteristic should not only focus on the moments where an explicit knowledge need is induced, but more structurally at the introduction and evaluation of each unit of one or more learning activities. Obviously, because it was not an issue in the first trial, the teachers had not been sufficiently prepared for the fact that the input of students was now an important driving force in the designed teaching-learning process and *how* such a process should be guided. In addition, the teachers did not automatically listen to students and, for example, collect their answers at moments where it was appropriate (such as the summarising activity in which students formulate the first step). This asks from teachers that they trust students to be largely correct (section 3.3). This trust is necessary for students to feel that they own the final, neatly summarising formulated answer (and the teacher might then add: “I think this is what you actually said/meant?”). I concluded that this trust is a condition for proper attention for students input.

Without explicit preparation, the teachers relied on their own teaching style. And in line with Lemke’s findings, these were dominated by dialogues in the classroom with what Lemke described as a ‘triadic dialogue structure’ and ‘teacher lecture’ (Lemke, 1990). See also: Rop (2003), Watts *et al.* (1997) and Carr (1998) who show how difficult it is for a teacher to question differently.

The combination of these findings led to a rethinking of the characteristic *attention for student input*. I felt the need to redesign teaching activities in a more structural and detailed way using interaction structures (see section 4.6). These interaction structures were to make sure that teachers were prepared to pay proper attention to student input when introducing and evaluating a functional unit of one or more learning activities, and when guiding students in carrying out the learning activities, thereby using a format that matches the complexity of the learning activities.

Summary and conclusions

The following general conclusions were drawn based on the findings of this second trial:

1) The relation between the *context* characteristic (driving question) and the *need-to-know* characteristic (content-related motive) was not very strong.
2) The same might be said about the relation between the characteristic *attention for student input* and the *need-to-know* characteristic, which should have been more in tune with each other. The *attention for students input* characteristic needed rethinking on a much more structural and detailed scale: how, as a teacher, to create a *context* for the learning activities, to guide the respective learning activities, and to evaluate the respective learning activities.
3) The procedure for testing and judging water quality was still not an integrated part of the teaching-learning process and the students had not considered it a relevant learning aim.

Of course, the research cycle with the second version also showed some specific findings such as the finding that the 'how much of what' question did not emerge. Students had doubts about their own performances and not about the tests themselves. They also had doubts about the limited list of parameters on which the water was actually tested. When asked 'would you actually drink this water if the tests results were within the norms?', almost none of the students would.

4.5.4 Implications for the third version

In this concluding section, I will present how the findings with the second version have finally led to the third version of the design. Again, I will address the two most abstract levels of description: the characteristics of meaningful and the didactical structure with its functional phases (table 4.1). In chapter 5, the detailed scenario is presented.

The findings with the second version indicated that the three characteristics of meaningful and especially their interrelatedness needed rethinking.

Firstly, the rethinking of the relation between the *context* and the *need-to-know* characteristics, which proved not to be strong enough, resulted in the re-interpretation of the concept of 'context'. Instead of interpreting *context* as a driving question, it was now interpreted as an 'instructional version of an authentic practice' (Bulte *et al.* 2002). The idea is, basically, that an authentic practice, in this case the testing and judging of water quality in standard situations, is used for the design of a teaching-learning process *about* this authentic practice by imitating it to a certain degree.

Secondly, with respect to the teaching method and *attention for student input* characteristic, the second research cycle made me realise that explicit structural attention for the interactions between teacher and students and between students at the level of each activity was needed. The idea of interaction structures as a means to address structurally *attention for student input* at this level was inspired by Lemke's work (see also section 2.2.3.). He used the concepts 'episode' and 'dialogue and activity structures' to analyse communication between teachers and students in traditional science classes (Lemke, 1990). This rethinking eventually resulted in the formulation of a *desired* interaction structure for each episode of the teaching-learning process that matches the purpose of the episode. The description and underlying argumentation of desired interaction structures did not only form a basis for the preparation trajectory for the teacher involved in the third trial, but I also used it as a framework for analysis (see chapters 5, 6 and 7).

Context and need-to-know: an 'instructional version of an authentic practice'

The evaluation had the following implications with respect to the *context* and *need-to-know* characteristic. It was concluded that the idea of a driving question and the broad motive to answer such a question (*context)* did not sufficiently direct what content-related motives could be triggered among students at which moments (*need-to-know*). The relationship between *context* and *need-to-know* (broad motive and content-related motives) needs to be strengthened in this sense. Inspired by the research of Van Oers and Van Aalsvoort (Van Oers, 1998; Van Aalsvoort, 2000), this has led to the idea of establishing an instructional version of an authentic practice[6] (Bulte *et al*., 2002).
Central to Van Aalsvoort's work, which is inspired by activity theory, is the idea that students try out different roles of social [chemical] practices by simulating these roles in the school setting. By experimenting with different roles of different practices, students are expected to experience which roles appeal to them. It is important that students in their exploration of different roles get involved in a dialogue with the culture that is represented by the teacher and the texts. Van Aalsvoort emphasises that this will contribute to pedagogical goals such as 'contributing to the students' development as a person in such a way that students come to be more conscious of their decisions (in general)'. In her work, she primarily focuses on such pedagogical goals and not on *didactical* goals such as 'contributing to a teaching-learning process in which students know what they do and why they do it every step of the way to reach a certain width and depth of understanding' (Van Aalsvoort, 2005).

I will primarily use the term practice, however, as a *didactical* solution to a *didactical* problem (see section 4.5.3). It can therefore be seen as complementary to Van Aalsvoort's interpretation. I will explain below how, as a didactical tool, the idea of a

[6] It is inspired by Van Oers' and Van Aalsvoort's interpretations of the concept 'context' as a 'communal enterprise or practice'

practice could serve in the didactical outlining of the roles students are to adopt according to Van Aalsvoort. This way it contributes to then width and depth of their understanding of such a role.
A practice, as I use it, involves a central, characteristic procedure (or activity) with distinct purposes and aims (in this case, judging water quality in order to decide whether it meets the criteria for its use). The characteristic procedure which is followed in such a practice is functional for achieving these purposes and aims. In this case, that would be the water quality control procedure which is followed by chemistry analysts in a lab using standard methods. This is actually (in retrospect) a *different* practice than the water quality research practice described by Haselager (see section 4.2). Water quality research is in fact a practice that follows on standard water quality control. If something turns out to be wrong in a standard water quality check, then the *research* question emerges: what is wrong and how come?

The basic idea is that a characteristic procedure can be used to strengthen the relationship between *context* and *need-to-know* in an instructional version of an authentic practice, thereby making the procedure an integrated part of the teaching-learning process, if the following conditions are met:

a. Students should appreciate the purposes and aims of the authentic practice. This contributes to a broad motive to learn about the practice by simulating it, in the sense that it relates what is going to happen in chemistry class to society at large[7].
b. Students should have intuitive knowledge of the characteristic procedure which is followed in the authentic practice and appreciate its functionality in achieving its purposes and aims. This intuitive knowledge and appreciation of 'what would be a logical next step' can be used to induce content-related motives among students. That is, in order to take a next logical step students will at times find that they need more detailed information of a certain kind, which can be obtained by consulting the authentic practice. At other times, such information (e.g. the parameters to be tested) will induce the need for a refinement of such a procedure step (why these parameters?: we should find out if the parameters are to be trusted). Examples of this can be found in section 4.6 and, in more detail, in chapter 5.

Every new step of the procedure which the students develop is explicitly summarised on a procedure poster in front of the class. These summarising activities should form the leading theme of the module.

Framed as an heuristic for the design process, it can be said that the task is to determine the structure of functional activities of the authentic practice (such as the activities in water quality research according to Haselager, see section 4.2) and to transform this structure into a didactical structure of *good reasons* for students to perform actions in the instructional version of this authentic practice. When done

[7] Van Oers and Van Aalsvoort emphasise this aspect of enculturalization as an aim in itself.

properly, students are expected to see the point of every activity in the instructional version of the practice, beforehand, in light of achieving their aim. Although achieving this is not a straightforward process, this design heuristic is at any rate much more directive and supportive than the idea of a driving question, and can be expected to lead to a more cohesive didactical structure.

Attention for student input: interaction structures

In this section, I will explain how the idea of interaction structures specifically addresses the issue that the teacher should pay attention to the student input, as far as it is an important driving force of the teaching-learning process (part III). The second trial showed that the teachers tended to ignore student input, the instructional format of activities in which the interaction took place did not always match the complexity of the content-related progression, and teachers sometimes tended to use the content as a way to keep control. I will explain how, I think, interaction structures give content to the characteristic *attention for student input* at the detailed level of interaction in a learning activity.

Applying Lemke's (1990) observations to my case, helped me realise that the communication and specifically the interactions between teacher and students in traditional science classes differ in an important way from what the communication and interaction patterns should be in a problem-posing approach. I realised that the problems with the characteristic *attention for students input* in the second version of the design were for a large part the result of not considering this difference. As mentioned in chapter 2, one of the things Lemke observed is that in traditional science lessons, the teacher uses dialogue structures to communicate science, which promote the *teacher* to use scientific terminology rather than the students. As a result, students get little opportunity to explore the use of words that refer to scientific concepts and therefore their meaning. Lemke gives as examples structures such as 'triadic dialogue', the related structure 'external text question' and 'teacher lecture' (a teacher dominated monologue). A dialogue, which is dominated by the triadic dialogue structure typically shows the following triad of moves (the moves between brackets are optional): the teacher (or in the case of external text: the textbook) poses a question, the student or students answer and the teacher evaluates:

	(teacher preparation)
Triad of moves	**Teacher Question**
	(teacher call for bids)
	(student bid to answer)
	(teacher nomination)
	Student Answer
	Teacher Evaluation
	(teacher elaboration)

Lemke describes how this type of dialogue structures functions in a school situation in the sense that they raise certain expectations among students of what is happening, what the options are for what comes next, and who is supposed to do what. Students are used to and expect the teacher to ask a question s/he already knows the answer to, and they expect their answers to be evaluated positively or negatively. Students are used to listening to a 'teacher lecture' and being asked evaluative questions afterwards. They are not surprised when the teacher uses content ("write this down, it is important") as a means to keep control when a class is getting too noisy.

Such expectations might interfere with a new approach to the teaching-learning activities. As showed in section 4.5.2, this was exactly what happened in the second research cycle. The main problem, I think, was that it was not clear to the teachers (and to me at the time) what *their* content-related contribution should be and what content-related contribution they might and should expect from the *students*. In other words, at what points and how to hand over the responsibility for the content-related progression to the students?

As already described in the previous section, in a problem-posing approach this balance of who is responsible for the content-related progression differs from traditional chemistry or science lessons. In traditional science or chemistry lessons, students are not expected to have much responsibility in the content-related progression of the teaching-learning process. They are not expected to have a clear view of 'what comes next' (see also chapter 2), perhaps the subject matter is thought to be too difficult for that. Both teachers and students are very much aware of this and it directs the communication process. In a problem-posing approach, however, the students play an important role in driving the teaching-learning process. Their input forms the driving force in the sense that each activity builds on the previous one. Questions raised by students in the previous activity (knowledge needs) are addressed in the next and so on. For example, when students are to find out from the authentic practice what parameters they should test the water on, it is because *they* raised the question. When students are to find out about the accuracy of test methods through a series of activities, it is because *they* raised questions about the test results. And so on. If students are to feel that their input matters in the process, the teacher, as director of the process, should therefore pay attention to their input in such a way that the students experience that *their* questions, remarks, answers are actually an important driving force of the teaching-learning process.

I will now explain the idea of interaction structures and how I believe this idea to address the above-described problem. First, the teaching-learning process was divided in episodes, inspired by Lemke[8]. Each episode addressed an episode question that students were expected to feel they needed to answer in order to achieve their purpose of finding out about what is involved in water quality judgement. In other words, these episode questions (except the first one) reflect the students' content-related motives (see section 5.1, table 5.1).
Each episode has the following structure:

[8] Lemke uses the term episode for a part of a lesson. In his interpretation an episode is defined in different possible ways depending on what the researcher wants to analyse. According to Lemke, an episode might be defined by the domination of a certain topic or maybe by the domination of a certain type of dialogue structure. I used the idea of episodes differently.

Setting the stage for the episode question
Addressing the episode question
Evaluating the episode question

An episode consists of one or more activities, depending on the complexity of the episode. The simplest episode consists of one activity while episodes that are more complex consist of a series of activities. Interaction structures are to provide the teacher with guidelines as to how to direct the interaction in each episode as follows.

In the first and the second episode, the teacher is to direct the interaction according to the interaction structure 'collect input' as the content of these episodes is not expected to be complex for students (see sections 5.2.1 and 2). The teacher is therefore not supposed to contribute to the output of the activity with respect to the *content*. The collect input-interaction structure prepares the teacher for his/her role with the following guidelines:
-(for every interaction structure) In the stage setting of the episode students should be involved as much as possible in the bridging between this episode and the previous one. To achieve this the teacher should, when explicitly pointing to their answers to the previous episode question, ask their opinions or ideas on the relationship between the previous episode and the next. This should contribute to the students' feeling that their input matters (as it is actually asked for and used).
-When addressing the episode question students work individually or in groups on the tasks. The interaction between the teacher and students is limited and largely based on the initiative of the students (they might ask questions to clarify tasks etc.)
-When evaluating the episode question the teacher only needs to *collect* the answers of the students. Not all students need to call out. However, it is important that they are involved and express their ideas or opinions as much as possible. The teacher therefore poses questions to the whole class, and not one particular student, and s/he might trigger a discussion by asking things such as: 'anyone else?, who thinks the same/differently?' or posing a controversial opinion/idea. When evaluating their input the teacher can repeat their answers, verifying whether s/he understood them, thereby underlining the fact that s/he takes the students' opinions seriously. The teacher does not need to evaluate the content of their answers as 'positive or negative', to add to their answers, or to present an appropriate answer him/herself.

In the much more complex episodes 5 and 6 the teacher is to direct the interaction according for example to the interaction structure thinking-sharing-exchanging. When this interaction structure is used, it is expected that students cannot come up with a certain level of output solely by themselves, most of the time this involves a reflective, more abstract kind of output. The teacher then has the task both to help bring the student's output to a higher level of abstraction, and to let students experience that their output *really* contributed to this higher level. The thinking-sharing-exchanging-interaction structure prepares the teacher for his role with the following guidelines:
-(for every interaction structure) In the stage setting of the episode students should be as much as possible involved in the bridging the episode with the previous one. To

achieve this the teacher should, when explicitly pointing at their answers to the previous episode question, ask their opinions or ideas on the relationship between the previous episode and the next. This should contribute to the students' feeling that their input matters (as it is actually asked for and used).
-In the 'thinking and sharing' part students discuss their answers amongst themselves. The interaction between the teacher and students is limited and largely based on the initiative of the students (they might ask questions to clarify tasks etc.)
-during the exchange part the teacher collects the different group answers and let them clarify their answers when needed. The teacher is needed to formulate a satisfactory (complete) answer, raising it to a more abstract level, thereby making explicit use of all the different (uncertain and incomplete) answers of the different groups.

For the third version of the design, different interaction structures, matching the different complexities of the episode-purposes, were worked out. Appendix 1 contains a description of the interaction structures that were used.

4.6 The third research cycle: the didactical structure of the design.

In section 4.6 only a part of the third research cycle is presented. In section 4.6.1 the third operationalisation of the characteristics of meaningful is presented. In section 4.6.2 the didactical structure of the third version of the design is elaborated upon in broad outlines.
The remaining chapters cover the third research cycle: the scenario is presented in chapter 5, the methodology of evaluation in chapter 6, the evaluation and findings in chapter 7 and the implications (research questions, contribution to didactical theory and further research) in chapter 8.

4.6.1 The characteristics of meaningful in the third version

In this section, I will present how the three characteristics of meaningful were operationalised in the third version of the design.

Context and need-to-know

As described above, the reconceptualisation of (the connections between) the characteristics *context* and *need-to-know* led to the idea of establishing an instructional version of an authentic practice.
Students should become broadly motivated by the exemplary cases and therefore the purposes of the authentic practice to adopt their role in its instructional version and find out ('as interested students') how people in the authentic practice solve such cases (judge water quality) by simulating the authentic practice in a sense (that is: solving an exemplary case). The students' intuitive knowledge of such a procedure is used to design a problem-posing teaching-learning process, thus creating an *instructional* version of the procedure of the authentic practice.

The *context* and *need-to-know* characteristics are evaluated based on the following criteria:

a. are students broadly motivated by the purposes of the authentic practice to adopt their role and get involved in its instructional version as intended?, and:
b. do students experience at the intended moments that they need specific knowledge to take the next step in light of their purpose or at other moments that such specific knowledge asks for a refinement of a procedure step?

Attention for student input

In the third version, this characteristic was operationalised as follows:

a. Student input contributes throughout the whole teaching-learning process to the common purpose of developing a procedure for judging water quality.
b. The teacher directs interaction at the level of the stage setting and evaluation of *each* of the episodes of the problem-posing story line, according to the respective interaction structure (see chapter 5 and appendix 1).

The *attention for student input* characteristic was evaluated based on the following criteria:

a. Do students feel that their input contributes to a common purpose?
b. Are the designed interaction structures implemented by the teacher as intended, do they lead to the intended interaction patterns in the classroom and do students as a result feel that their input matters as a driving force in the teaching-learning process

4.6.2 The didactical structure of the third version

In this section, I will discuss the didactical structure of the third version of the design in broad outlines. Figures 4.6 A and B present the general didactical structures of the third version of the design. Figure 4.6A presents an overview of the didactical structure of the whole teaching-learning process. In figure 4.6B the didactical structure of phase 3 is presented in detail. This structure shows the interplay between the progression in issue knowledge and the progression in procedural knowledge of phase 3. Both figures show explicitly how one can use the intuitive appreciation by students of the logic of the successive steps of the characteristic procedure in inducing a specific knowledge need.

The teaching-learning process is divided in four phases (inspired by Kortland, 2001, see section 4.5.1). These phases describe in very broad outlines what should be achieved with students.

Phase 1 is a crucial phase in the sense that the stage of the instructional version of the practice is set in this phase. In this orientation phase, students should be invited by the teacher to play their part in the instructional version of the practice. The term ‘stage setting’ therefore seems appropriate. It is inspired by what Leach and Scott[9] have called the ‘staging of the scientific story’. In this case, the ‘scientific story’ should be replaced by ‘the instructional version of the practice’.

[9] Leach and Scott, in their turn refer to Tibergien (2000), who uses the phrase/term ‘mise-en scene’.

> The whole aim of staging the process (involving talk, other semiotic modes and various activities) is to make the scientific story appear intelligible and plausible (..) to the students. (…) A fundamental quality of the staging process is that it is persuasive (Sutton, 1996) in nature as the teacher seeks to convince the students of the reasonableness of the scientific story.
>
> (Leach & Scott, 2002, p122)

Students should become engrossed in some appealing exemplary cases of the authentic practice and should become generally motivated to find out how people solve such cases by simulating the authentic practice in an exemplary case in chemistry class (the instructional version of the practice). In all these cases, the same characteristic procedure is followed to solve them: water samples are taken (these are presented in the classroom), the water is tested and test results are compared to a reference to come to conclusions. Students proved to have (based on the previous versions) an intuitive knowledge of this procedure, but they lack the specific knowledge they need to actually carry out the procedure (see figure 4.5 for an overview).

Figure 4.5 The instructional version of the procedure for testing and judging water quality in the third version

Steps of the procedure	
1. Determine water function, this determines water quality criteria	What is allowed in drinking water? How much of this is allowed in drinking water? Selection parameters and norms
2. Test water sample: does the water quality meet the demands?	What does the water sample contain? How much of this does the water sample contain?
3. Compare	Do we trust the test results? i. Did we perform okay? ii. Accuracy: variability and position of the test results
4. Judge	Do we trust the fact that only four parameters are to be tested? How are these four parameters selected? How are the norms established?

In phase 2, students begin to apply their intuitive knowledge to a representative exemplary case with the aim to find out how people solve such cases. While doing so, they are expected to understand that they lack specific information to be able solve this exemplary case in order to find out about how people solve such cases. In this way, a specific content-related motive is to be induced. In particularly students are expected to become aware that they know intuitively that the water should be tested, and that they do not know what to test the water for and how to execute these tests. This provides them with a motive for inquiring.

In phase 3, students are to extend their knowledge until they solve the exemplary case and arrive at the end of the procedure. This process of knowledge extension is driven by content-related motives induced by the students' appreciation of the logical next step, in perspective of the main evaluative question 'Does the water quality meet the quality criteria?' (in other words, 'Can we solve the case yet?'), and their expectation that the procedure will be useful in the other cases too. Phase 3 starts with satisfying the knowledge need induced by phase 2: information is derived from the authentic practice about 'What do people, who are involved in this practice, test the water on in such cases?' Students are to test the water sample and their appreciation of the logical next step induces the next knowledge need, as students intuitively know that they need something to compare their test results to in order to be able to judge the water quality. This is used to induce a motive for finding out 'To what should we compare our test results?'. Relevant information is derived from the authentic practice: 'With what do people, who are involved in this practice, compare their test results in such cases?'. Next, the evaluative question 'Does the water quality meets the quality criteria?' is used in this phase to make students more prominently aware of (perhaps earlier felt) deficiencies in their issue knowledge: Are these the proper water quality criteria? Did we derive the test results in a proper manner? (Did we do the tests properly? Are the test results accurate enough?). These doubts are the next content-related motives that drive the learning processes. Students feel the need to extend (or at least see the point of extending) their knowledge in order to solve the respective doubts for solving the case. Phase 3 of the teaching-learning process ends with a solution for the exemplary case and the explicit description of the procedure. Figure 4.6B shows the detailed didactical structure of phase 3.

The general aim of the students is, in the first place, to find out how people monitor water quality, so students will still wonder whether the procedure is also useful in solving the other cases. This purpose is the content-related motive that is expected to lead students to the concluding activity of phase 4: find out to what extent the procedure is applicable to one of the other cases discussed in phase 1.

The implications for the design of teaching-learning activities are presented in chapter 5. In that chapter, I will describe the third scenario, which is based on this didactical structure.

Figure 4.6 A ***The didactical structure of the third version of the design***

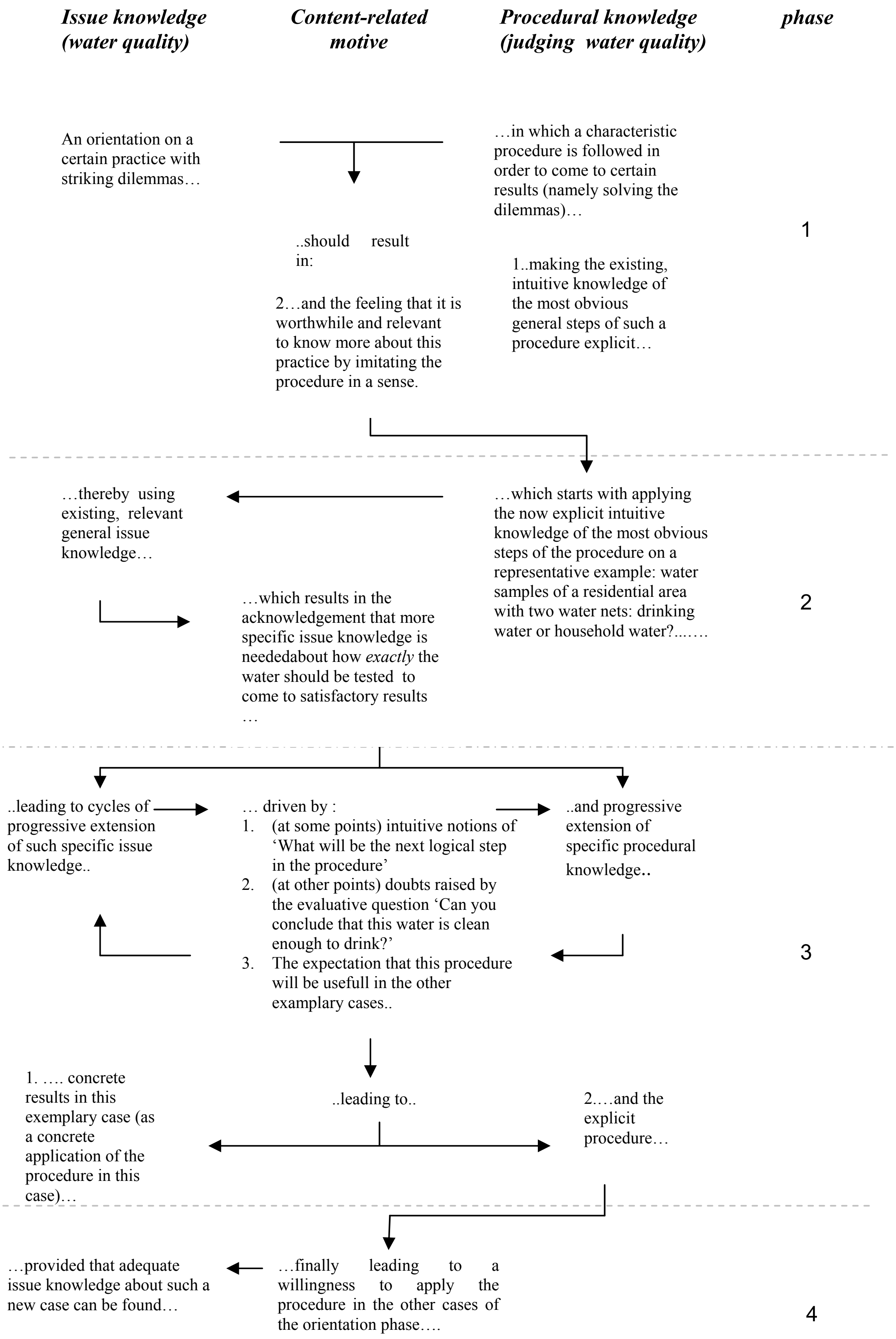


Figure 4.6 B The didactical structure of the third phase of the design

Issue knowledge	***Content-related motive***	***Procedural knowledge***

…which results in the acknowledgement that more specific knowledge is needed about how *exactly* the intuitive procedure steps should be executed to come to satisfactory results …

..which leads to using the authentic practice as an information source for answering the question what the water should be tested on as a first step in satisfying the induced knowledge need…

….and applying this information by testing the water samples in the instructional version of the practice on the same things, producing a list of test results and doubts about these test results

..which raises the need for a reference to compare the test results with ….

which leads to using the authentic practice as an information source: What reference is used in the authentic practice?…

….and applying this information in the instructional version of the practice by comparing the test results with the same reference as used in the authentic practice…

..which raises the question: ‘can we now judge the water quality and solve the exemplary case?…

…which leads to an attempt to solve the case based on the available knowledge…

..which leads to the realization that at first two types of doubts need to be solved in order to be able to solve the case:

1. concerning the limited list of tested parameters: Why these four, are there no others?
2. concerning the test results: Did we perform correct? Are the tests accurate enough?

..which leads to first solving the doubts about the reference ‘Why these four and why not others?’ by using the authentic practice as an information source.. .

…which partly resolves the doubts but still leave the doubts about the test results unsolved ..

..which leads to solving the doubts about the test results by expanding knowledge in the direction of ‘How do we know whether we did the tests correctly and how do we know if the test results are accurate enough?’….

Chapter 5

The design of the third scenario

5.1 Introduction

In chapter 4, I have described how the third version of the design is based on its previous versions. I showed how the interpretation of the characteristics of meaningful, *context*, *need-to-know* and *attention for student input*, and their inter relatedness evolved in the process of developmental research. It finally led to the idea of designing an instructional version of an authentic practice and interaction structures. The descriptions of the first two versions of the design and the findings with these versions in chapter 4 were on the rather abstract level of 'the characteristics of meaningful' and of 'the didactical structure & phases' (see chapter 4, table 4.1). The most detailed levels of description were only addressed in an illustrative manner, when functional for clarifying specific lines of thought.

In this chapter, I will present the third scenario on the two most detailed levels of description: the level of episodes and interaction structures and the level of learning activities. This way I want to give further insight in the scenario to be tested: a detailed description of the intended teaching-learning process with argued expectations of every step. The main evaluation question to be answered in chapter 7 is to what extent does the design of a teaching-learning process meets the criteria of the characteristics of meaningful and, as a result, can be seen as a case of meaningful chemistry education?

Before I present the actual scenario of the third version, I will first explain in section 5.2 how this scenario is presented.

5.2 The way the scenario is presented

As I already mentioned in section 4.6.2, the scenario can be divided into four different didactical phases, which fulfil different functions (based on Kortland, 2001). These functions refer to what should be reached with the students in the respective phase. Each phase consists of one or more episodes, which in their turn fulfil a function within the respective phase in the sense that each episode contributes to fulfilling the function of that phase. Each episode can be seen as addressing a question in one or more activities, depending of the complexity of the question. Except for the first episode in which students are to become broadly motivated, the episode-questions coincide with intended content-related motives. The formulation of episode-questions serves as an intermediate step in the design of learning activities. They help in checking the logic of the episodic story line.

Table 5.1 gives an overview of the phases, the connected episodes, their respective functions, the content of each episode and their interaction structures. For a more detailed description of the types of interaction structures: see appendix 1.

Table 5.2 shows how the activities are planned in each of the successive episodes.

Table 5.1 ***An overview of the functional didactical phasing of the scenario***.
The phases, episodes, their respective functions and the questions of each episode, the content involved and the respective interaction structures (see also appendix 1).

Didactical phase		Episode		Content	I.S.*
	Function of phase	**Question**	**Function of episode**		
1	Students are to feel broadly motivated to get involved in the instructional version of the practice about judging water quality. They are to get a clear view of the purpose of the instructional version of the practice, the way they are going to achieve this and their role in it.	1. What is involved in water quality judgment?	See function of phase	Exemplary case, descriptions of the practice 'judging water quality' and genuine water samples	CI
2	Students are to experience that their intuitive notions are not sufficient. They know that the water should be tested, but lack specific issue knowledge, which might be formulated, as the following knowledge need: 'What does the water sample contain?'	2. Does the water of the exemplary case meet the quality criteria for drinking water?	See function of phase	The exemplary case description of the residential area with two water nets and accompanying water samples	CI
3	Students are to extend their issue and procedural knowledge in progressive cycles in the direction of the raised knowledge need, until a satisfactory procedure is reached.	3. What does the water sample contain?	Students satisfy their induced knowledge need, which will raise the need for taking the next step in this process: 'how much of what is allowed in drinking water?'	The four parameters that are tested in the authentic practice, how this is done in broad outlines and how it might be done in the instructional version.	CI
		4. Does the water meet the quality criteria for drinking water?	Students are to link their doubts to the following categories: a. Trustworthiness of the limited list of parameters and norms. b. Trustworthiness of the test results.	The allowed values of these four parameters in drinking water to compare their test results with (information from the authentic practice)	CaI
		5. Do we trust the list of tested parameters and their norms?	Students address the episode question by learning how the list of four parameters and norms to be routinely tested in this case is established.	The argumentation behind a) the four standard parameters and b) a selection of parameters from the complete official list for drinking water.	TSE
		6. Do we trust our test results?	Students address the episode question by developing a notion about reliability and accuracy of the used test methods and how this might influence their judgment.	The test results of all the groups and the doubts the students have about these test results.	TSE
		7. Does the water meet the quality criteria for drinking water?	Students solve the exemplary case, summarise their procedural knowledge, and feel the need to test the usefulness of the procedure in the other exemplary cases (because they expect this to be the case).	A report, derived from the authentic practice	TTE
4	Students are to reflect on 'what they have learned': a procedure for how water quality is judged in standard situations.	8. To what extent can the procedure we used be applied in the other exemplary cases of phase 1?	See function of phase	The other case descriptions and for each case, a website with relevant parameters and norms to be tested	TTE

*Interaction structure

Table 5.1 shows that the main evaluative driving question for students, does the water meet the quality criteria for drinking water? (see for example episode 7), plays a role at different moments of the teaching-learning process. It is used at these moments to induce specific knowledge needs (depending on the stage of the learning process). As I described in chapter 4, there are three conditions for inducing content-related motives that drive the progression of the learning process in the instructional version:

A. Students should appreciate the purpose of the authentic practice and
B. Students should have intuitive notions of its characteristic procedure through which this purpose is achieved.
C. The authentic practice should provide students with information when needed.

When the evaluative question is posed to students, their appreciation for the purpose of the authentic practice is used to induce a knowledge need. Every time it is posed students are to realise that 'they are not there yet', that is (in this case): they are still not sure about the quality of the water, because they still lack specific information. The learning activities, which make use of the evaluative question, follow the same general pattern:

A. Does the water meet the quality criteria?
   Yes/no, because…
   I am not sure, because...
B. If you are not sure, what type of information do you need?
   I would need the following information: …

Question A induces a knowledge need (or at least a discussion on 'Are we there yet?'). Question B forces students to put their knowledge need into words.

Table 5.2 An overview of the six lessons

Lesson (min)	**Phase**	**Episode**	**Activities**
1 (50)	1 & 2	1 and 2	1,2,3
2 (50)	3	3	4
3 (75)	3	4 (evaluation), 4 and 5	5 –10
4 (50)	3	6	11-14
5 (50)	3 & start 4	7 and 8 (stage setting)	15, 16
At home	4	8 (addressing the episode question)	16
6 (75)	4	Evaluation 8	Test

In the next section, the actual scenario is presented. Sections 5.3.1 up to 5.3.8 each deal with an episode of the teaching-learning process. My aim is to make insightful the argumentation underlying the choice for the content, the interaction structures and finally the teaching-learning activities. This argumentation is for a large part based on findings with the previous research cycle. The main question to be answered in the scenario is to what extent does the design of a teaching-learning process meets the criteria of the characteristics of meaningful and, as a result, can be seen as a case of meaningful chemistry education? That is, the criteria:

- Are students broadly motivated by the purposes of the authentic practice to adopt their role and get involved in its instructional version as intended?, (*context)* and:
- Do students experience at the intended moments that they need specific knowledge to take the next step in light of their purpose or at other moments that such specific knowledge asks for a refinement of a procedure step? (*need-to-know*).
- Do students feel that their input contributes to a common purpose? (*attention for student input*-c).
- Are the designed interaction structures implemented by the teacher as intended, do they lead to the intended interaction patterns in the classroom and do students as a result feel that their input matters as a driving force in the teaching-learning process (*attention for student input*-d).

Each episode is presented as follows. First, the episode-question is mentioned in the title of each episode (see also table 5.1). Next, the function of the episode (what should be achieved in the respective episode with students) is described. Next, I will give an argumentation in the light of the episode function(s) of the choice for the content and, based on the complexity of the content and learning purposes, the desired interaction structure.
Finally, the expected unfolding of the episode is presented, using the basic structure of an episode: Setting the stage of the episode question, Addressing the episode question, Evaluation of the episode question. Most episodes end in addition with a 'wrapping up'-part. This description includes the planned teaching-learning activities and shows the detailed intended teaching-learning process. If more than one episode function is at stake, I will indicate between brackets what choice I expect to contribute to which episode function. In addition, an indication is given of how long the activities are expected to take.

The emerging procedure steps are collected in class on the 'procedure poster'. In the presentation of the planned teaching-learning process the moments at which the teacher collects the new steps is indicated as follows:

Figure 5.1 The procedure poster after episode 3.

Steps of the procedure	
1. Determine water function, this determines water quality criteria	**What is allowed in drinking water? How much of this is allowed in drinking water?**
2. Test water sample: does the water quality meet the criteria?	What does the water sample contain? **How much of this does the water sample contain?**
3. Compare	

The left column shows the successive steps of the procedure (up to 4: judge). The right column shows the refinements of these steps for the case of the two water nets. Every time a new step or additional refinement of a step is planned to be added, this is indicated in the text in bold. Figure 5.1 for example shows that at the end of episode 3 (when this figure is presented) a new step is planned to be added: '3. Compare'. In addition, steps 1 and 2 are to be extended with additional refinements (in bold).

To conclude, the third version of the scenario presented here, can be seen as underpinned by empirical evidence, but also as still hypothetical to a certain extent. In chapters 6 and 7, I will discuss the evaluation of the third design and answer the question to what extent the scenario was adequate in reaching its aims. What worked, what did not and (as much as possible) why?

5.3 The scenario in detail

5.3.1 Episode 1: What is involved in water quality judgment?

Functions of the episode

1A. Students are to feel broadly motivated to get involved in the instructional versionof the practice about judging water quality.

1B They are to get a clear view of
- 1Ba the purpose of the instructional version of the practice
- 1Bb in broad outlines how they are going to achieve this
- 1Bc their role in it

I will indicate between brackets what choice I expect to contribute to which function.

Justification of content in the light of the episode functions

Situation: the stage of the instructional version of the practice should be set in this phase and students should be asked to play their part in it.

To give the students a clear sense of purpose and broadly motivate them for the instructional version of the practice a selection of genuine water samples, which are taken for standard quality control are presented to them, together with case descriptions of the practice they are derived from (1A and 1Ba). In order to contribute to the functions of the episode, the examples are chosen based on the following criteria:

a. The examples must provide for students recognizable and appealing questions or issues (presented in the case descriptions).
b. All examples should follow the same general procedure for judging water quality: water use determines quality criteria, water quality criteria are expressed in a list of standard parameters and norms that give a 'good enough' impression of the water quality, these parameters are tested and test results are compared to the norms. A judgment is formulated based on this comparison, with the accuracy of the tests and how the tests were performed in mind.
c. Different ways of using water should be represented.
d. The examples must be genuine. With this, I mean that the water samples must be genuine water samples from genuine laboratories and the questions in the case description must be genuine questions.

I expect that if the examples meet the criteria they will contribute to fulfilling the functions of the episode as follows. Criterion 1 concerns such issues as 'Can we drink this water?', or 'Can we swim in this water?', described in the case descriptions. They should illustrate for students the importance of standard water quality control. It should motivate them to learn about how this is done in reality (1A and 1Ba). Criterion 2 should contribute to directing the students' attention towards the common procedure as a way to solve the interesting case descriptions. Because students have intuitive notions of such a common procedure, which I base on previous findings (see 4.5.2: *Context* and *need-to-know*), I expect them to recognize that this procedure underlies every case (1Ba). Criterion 3 serves, I think, three functions: a. it helps directing students to the first (intuitively known) step of the common procedure: quality criteria depends on water use as different water uses are represented. (1Ba); b. it contributes to a notion of the diversity of cases of the authentic practice of judging water quality (1Ba); c. the variety of cases provides for an increased chance of raising the interest of students (1A). Criterion 4, I expect to contribute to a feeling of authenticity, in the sense that 'What we are going to do in the lessons' is derived from genuine, existing exemplars of this practice (1A).
The following cases were selected based on these criteria:

1. Surface water quality in a nature reserve.
   This case concerns a Dutch nature reserve. The water quality in this reserve is being monitored closely. The question is, is this water clean enough for the flora and fauna it is supposed to accommodate?
2. Swimming pool water quality
   The case concerns a swimming pool nearby the school. The water quality is constantly checked. The question is, is this water suitable for swimming?

3. Drinking water quality
   The case concerns a special residential area, which has two water nets: one for high quality drinking water, one for so–called less quality household water. The water quality is monitored to avoid mistakes like a mixing up of the two water types. This has actually occurred a few times in the past, which have been reported in the news. The water quality is regularly checked. The question is, is it drinking water or household water?
4. Purified sewer water quality.
   The case concerns a sewer water purifying installation in the area of the school. After the sewer water is purified, its quality is checked before it is discharged of in the river (which obviously is also in the neighbourhood of the school). The question is, is the water clean enough to be discharged?
5. Brew water quality.
   This case concerns a brewery. It uses special water with a specific water quality to brew beer from. Checking the water quality is part of the production process. The question is, is this water good enough to brew beer from?
6. Aquarium water quality
   The case concerns the aquaria of a big Zoo in Amsterdam. The fresh water for the very sensitive rainbow fish is kept in huge basins in the cellars of the buildings. The water quality of this storage water is continuously monitored before it is pumped to the aquaria. The question is, is this water suitable for the rainbow fish?
7. Natural swimming water quality (surface water).
   This case concerns a lake, again in the neighbourhood of the school, which is used for recreation. The question is, is this water safe to swim in?
8. Sea water quality
   This case is about the European standards for ‘safe beaches’. If the water quality of a beach is considered safe to swim in, it gets a ‘blue flag’. The question is, does this beach deserve a blue flag or not?

All these cases are genuine and together demonstrate a diversity of water uses. In all cases, the same basic procedure is followed to monitor the water quality. All cases address questions, which I expect are appealing to students, because they refer to acute issues. Some of them refer to basic needs: if the water you swam in or drank was not safe, you might get ill (or at least it might taste bad). Some of them are, I expect, appealing because other things are clearly at stake: if the water is not clean enough, the rainbow fish might die or it will affect the flora and fauna in the river the water was discharged in.
I expect that functions 1Bb and 1Bc will be fulfilled by the way the cases are addressed in the activities.

Justification of the interaction structure in the light of the content

The teacher prepares the context of the episode question. The two main activities in this episode concern questions that 1) make explicit already existing knowledge and notions of students, and 2) their opinions. I therefore expect that students can easily answer the questions.

The teacher should not interfere with the answers of the students, as students are expected to be able to come up with adequate answers. I expect that the teacher can almost suffice with collecting students' answers when evaluating the activities. This should add among students to their feeling that their input matters.
I think that the teacher only needs to help students with generalizing their answers in the formulation of the first two steps of the procedure (see below). However, I do not expect this to be a complex step at all, as it concerns the already existing intuitive notions of students (see section 4.5.2). The teacher should always build explicitly on students input, when summarising answers as a procedural step by for example adding remarks like: 'is this what you mean?'.
The interaction structure, which fits the purposes and complexity of the activities, is 'collecting input' (CI, table 5.1, appendix 1).

Next, I will present more specifically how I expect the episode to unfold.

Expected unfolding of the episode

Setting the stage for the episode question (15-20 min)
The teacher introduces the first activity in the module, with respect to the content 'What is it about?' and procedural 'How is the activity organised?'.
First, the authentic practice is introduced. The teacher can show the students the water samples, emphasising that they are genuine, from genuine laboratories. The teacher might tell the students that the module is about what these labs actually do with such water samples and how this might result in an advice. S/he can elaborate on one or two of the cases so that students get the picture of the purpose of the authentic practice. To involve students more and to check whether they understand the examples s/he might ask them if they can come up with some good examples themselves (1A and 1Ba).
The teacher mainly collects the student input, as I expect it to be an easy question for them. I also expect that students will get involved in the topic judging water quality this way and feel that their input matters as they are made part of a discussion.

The teacher also has to explain how the activities of this first episode are organised:
Students will work in groups (three to four students). Every group will address a different case. At this point, they all should have a clear view of the type of cases. Each group is to discuss their case and answer a number of questions. They are to write down their answers on a poster, which will be displayed on the classroom wall for everyone to see.

Addressing the episode question: what is involved in water quality judgment? (15 min)
Students work on a case in groups of four. They are to discuss their case and answer the following questions on a poster:

a. What is the function of the water in this case?
b. What would water quality criteria in this case be, do you think?
c. How, do you think, do they check whether the water satisfies the criteria (what steps are involved)? (1Ba)

d. Do you consider it interesting to find out in chemistry class about how people do this? Yes/no, because… (1Bc)

All posters are displayed on the classroom wall. The students get a few minutes to view the posters of their classmates and to think about differences and similarities in the answers.

I expect that these activities will direct the students' attention to the role of water quality control in genuine cases and to the purpose of the instructional practice: to find out about *a procedure* for doing this (1Ba). Their intuitive notions about 'What steps are involved' are to be made explicit here and even more so in the evaluation part as the teacher explicitly uses this input of the students to come to a formulation of the first two steps of a procedure (see below) (1Ba). I expect, based on the findings with the previous version in which they had to answer the same type of question (see 4.5.2, *context* and *need-to-know*), that students will be able to answer the questions easily and will come up with water quality criteria like 'clean', 'clear', 'taste good/smell good', 'no poisonous stuff, no bacteria' and so on.
On the question, 'How do you think they check whether the water satisfies the criteria (what steps are involved)?' I expect answers such as 'they should test the water', 'Find out what the water contains', 'Investigate the water'. Maybe answers include steps like 'Take a water sample' and, although this answer was not triggered in previous versions 'Compare the water with good water' (1Ba). This expectation is based on what happened in the second version where the students wanted to test their water sample and a need for knowledge about *what* should be tested was raised (section 4.5.2).
Next, students should look at the posters of the other groups, from the perspective of the question 'What are similarities, what are differences?'. I expect that the students' attention will be drawn to the fact that different types of water use require different water quality criteria. I also expect them to appreciate the similarity of the procedural steps (question c). The last question on the poster is meant to invite students to take the role of students who are interested in (or at least see the point of) learning about judging water quality (1Bc).

Evaluation of the episode question (10 min)
The teacher can refer to the context setting: different cases of the practice of water quality control with different questions. The similarities and differences between the answers of the first three questions on the posters will be discussed. I expect that the students will bring up the following (in their own words), which the teacher can collect:

1. Water quality criteria depend on water use
   I expect this to emerge as a sort of conclusion when the first two questions of the poster are being discussed in relation to the question 'What are similarities and what are differences between the posters?'
2. For every example there must be a standard procedure, namely to answer the question, the labs will have to:
   a. Establish the water use and based on that the water quality criteria

   b. Test if the water meets the criteria (what it contains, that is) (1Ba)

   I expect this to emerge as a sort of conclusion when the third question of the poster is discussed in relation to the question 'What are similarities and what are differences between the posters?', as I expect that the different groups generally will come up with the same steps. As mentioned above, this expectation is based on the evaluation of the second version, see section 4.2.5.

3. Answers to the fourth question whether they think it is worthwhile to find out about this practice in chemistry class might, I expect, vary from something like 'because I want to know the answers in these cases' or 'it will be interesting to know more about how they do these kind of things' to 'I do not care/ am not interested' to. The teacher should explain *his/her* motivation for getting involved in the instructional practice: 'It is important to find out about how this is done in chemistry class, because it will give you an idea how chemistry actually functions in this practice' (1Bc).

I expect that students are focused by this activity on the purpose of the instructional practice: to arrive at a procedure for judging water quality. In addition, I expect that, since the teacher explicitly used their input (the intuitive notions), students feel involved in the teaching-learning process and that this contributes to the purpose that they will experience that their input matters.

Wrapping up (5 min)
The teacher can use the last question to direct the discussion to what will be one of the products of the instructional practice: a procedure for judging water quality. The teacher might emphasise that the steps of the procedure will emerge throughout the teaching sequence at certain specific moments.

At this point, students should l have made explicit their first intuitive notions of the procedure and the procedure poster should now look something like this:

Steps of the procedure

1. **Determine water function, this determines water quality criteria**
2. **Test water sample: Does the water quality meet the criteria?**

All posters stay on the wall of the classroom.

5.3.2 Episode 2: Does the water of the exemplary case meet the quality criteria for drinking water?

Function of the episode

Students are to experience that their intuitive notions are not sufficient. They know that the water should be tested, but lack specific issue knowledge, which might be formulated as the following knowledge need: 'What does the water sample contain?' (function 2)

Justification of content in the light of the episode function

Situation: the importance of finding out about what is involved in judging water quality in chemistry class should be clear now. The poster on which the procedural steps are to be written during the teaching-learning process stays on the wall for students and the teacher to refer to.

To induce the intended knowledge need, students are to apply their intuitive notions of how to judge water quality on the example of the residential area with the two water nets. This example will affect students personally, as it leads at several moments in the teaching-learning process to the question of quality evaluation: 'Do you now think the water meets the quality criteria for drinking water?'. I expect this question will stimulate students automatically to wonder about the more personal question: 'Would *I* drink it?'. The second version showed that students are personally affected when they are to judge drinking water and ask themselves 'Would I (dare to) drink this?'. For this reason, but also for the practical reason that all the necessary information was already prepared for the second version, the case of the two water nets was chosen as 'exemplary case', although the other cases were equally possible.
Based on previous findings (section 4.5.2), I expect that students easily come up with intuitive notions such as 'we do not know what is in the water, 'it might contain pollutions', 'we do not know what is in there'. Justification of the interaction structure in the light of the content

Justification of the interaction structure in the light of the content

The teacher, when setting the context for the episode question, should connect the outcome of the previous episode 'Do we think it is important to find out about how they solve such cases?' with this next one. S/he should make clear to students that also in this episode their opinions are asked for, and that there is no right or wrong answer. Because the questions are not complex and I expect students to come up easily with the intended knowledge need, the teacher has only to collect the input of the students and maybe summarise it more neatly in a single question: 'What does the water sample contain?'. The interaction structure, which fits the purposes and complexity of the activities, is 'collecting input' (CI, table 5.1, appendix1)

Expected unfolding of the episode

Setting the stage for the episode question (5 min.)

The teacher bridges the notion that it is interesting to find out about what is involved in judging water quality in chemistry class and the exemplary case: let us all find out for this example how it works. The groups all get the case description of the residential area with the two water nets: one with drinking water and one with so called household water. The question is: 'Is the water sample indeed drinking water as it is supposed to be?' The teacher might point at the poster of the group, which addressed this case: what quality criteria have this group written down? All the groups get a genuine water sample of the respective practice.

Addressing the episode question: does the water of the exemplary case meet the quality criteria for drinking water? (10 min)

At this moment (actually the first time), the purpose of the authentic practice is used to induce a content relative motive. I explained in the introduction the main evaluative question is used at specific moments in the teaching-learning process to make the students realise that they cannot answer it yet and that they lack certain specific information.

In the learning activity of this episode, students are to discuss the following:

You can take the caps from the water samples and examine the water. How does it look? How does it smell? Maybe you can look at the water through a microscope. What do you see?

A. 'Do you think this water sample meets the drinking water quality criteria and are you sure about this?'
I think this water meets/does not meet the criteria for drinking water quality.
I am sure about this/not so sure, because…

B. If you are not sure, what kind of information do you need to be sure?
To be sure that this water is drinkable, I need the following type of information…

When discussing question A, students are expected to apply their intuitive notions of judging water quality on this example. They can open the cover of the water sample bottle and smell the water. They can look at the water through a microscope. Nevertheless, I expect that the majority of the students will conclude that they cannot be sure about the water quality. Because, I expect, although they intuitively know that the water should be tested, they do not know what to test. Maybe students will remark that one needs information on 'What is allowed in drinking water?'. However, considering that this did not happen with the previous version (see 4.5.2) I do not really expect this. Question B will direct them specifically to express their knowledge need: the type of information they lack and need.

Evaluation of the episode question (5 min)
The teacher collects the opinions of the students and the information they say they need in order to judge the water quality. This way I think the teacher might secure that the students feel that their input matters. S/he summarises the student input: 'We do not know what the water might contain', and explicitly check with the students if this is what they meant.

Wrapping up (5 min)
The question is also the next step in the procedure to be written down on the poster, of which I expect students to feel that they came up with it themselves. The teacher wraps up the activity with the next step:

Steps of the procedure	
1. Determine water function, this determines water quality criteria	
2. Test water sample: does the water quality meet the criteria?	**What does the water sample contain?**

5.3.3 Episode 3: What does the water sample contain?

At this point, with episode 3, phase 3 starts (see table 5.1). In phase 3 (which covers episodes 3-7) students are to extend their issue and procedural knowledge in progressive cycles, in the first instance in the direction of the raised knowledge need, until a satisfactory procedure is reached (see also figure 4.6). Episode 3 is the first of the five episodes of this phase.

Functions of the episode

3A. Students satisfy their induced knowledge need.
3B. This will raise the need for taking the next step in this process: 'How much of what is allowed in drinking water?'
I will indicate between brackets what choice I expect to contribute to which function.

Justification of content in the light of the episode functions

Situation: every group has its own water sample. All the water samples are part of one series of samples of the same sample site, so the test results should be pretty much the same. Students should now want to know 'What does the water contain?'. This *episode starts in the student materials with chapter 2: 'What is in the water samples, and what is allowed in the water samples?'*

The students find out from the authentic practice what is tested in reality in this case. This is simply presented to them. I expect that the students will not question the information, as it is derived from the authentic practice, and value the information as useful in the light of their knowledge need (3A).
The students are also explained how they can actually test the same parameters in the instructional version of the practice (3A).

Justification of the interaction structure in the light of the content

The content-related progression of this episode is as follows:

1. Students produce a list of test results
2. Doubts are raised and written down by students about the course of the events when doing the tests. These doubts will later on frame the topics 'Do we trust the test results?'(see next episode).
3. The next knowledge need: 'How much of what is allowed in drinking water?' emerges naturally because the students need to compare their test results with some reliable reference (3B).

The teacher only needs to summarise this second knowledge need raised by the students as the next steps of the procedure. I expect that students find this next step obvious. The interaction structure, which fits the purposes and complexity of the activities, is 'collecting input' (CI, table 5.1, appendix 1). The teacher has only to point out to the students that their knowledge need 'to compare the test results with how much of the parameters is allowed in drinking water' can be seen as a (obvious) next step in the procedure.

Expected unfolding of the episode

Setting the stage for the episode question (10 min)
The teacher bridges the knowledge need 'What does the water sample contain?' to the introduction of chapter 2 in the students' materials. In this introduction, it is explained that the water in this case (household water or drinking water) is routinely tested on four parameters: chloride, nitrite, e-coli bacteria and pH. For every parameter it is briefly explained how it is tested in the authentic practice, in broad outlines. Based on findings with the second version, I expect students to appreciate the usefulness of the information on what parameters are tested in reality in this case (3A). At the same time I expect them to appreciate the difference between *how* they are tested at school compared to how they are tested in the authentic practice (much more sophisticated) (3A). As such, this is exemplary for other moments in the teaching-learning process, where on the one hand the authentic practice functions as information source for the instructional version. On the other hand, it is possible to make explicit for students why and how they deviate from the authentic practice, being in a school situation

Addressing the episode question: what does the water sample contain? (50 min)
The groups test their water sample on the four parameters, which are tested in reality.

To the students all the tests are 'black box tests', in the sense that they do not need to understand, and therefore are not explained how and why the tests work exactly. Nitrite and acidity are tested with standard Merck-tests. Both are indirect colorimetric tests: chemicals are added to a distinct volume of the water sample by following a prescription. This creates a specific colour. The coloured solution is than compared to a standard calibration series on a colour card. In the case of nitrite: the more nitrite the more intense the colour of the solution (if no or very little nitrite is present, then the solution remains colourless). In the case of acidity: the type and intensity of the colour determines the acidity.
The E-coli test is very simple. Students are to put a 10 ml of the water sample in a test tube with specific bacteria medium-pill (Coli bacteria food). The test tubes are kept at room temperature in the dark. After at least thirty and at most thirty-six hours, the E-coli bacteria can be determined as follows: a gel appears, but when it stays at the bottom of the tube, and when the fluid above the gel is clear yellow or reddish, and there are no gas bubbles, than the result is negative. If many gas bubbles have appeared and the gel has moved upwards in the tube, the liquid has gone turbid and the colour yellow, than the result is positive. The test had to be adjusted however, because it needed to be within the range of the norm. Students were told that when negative, there are less than 10 colonies per litre in the water sample (which is exactly the drinking water norm) and when positive there are more than 10 colonies per litre in the water sample. The actual values of this test are: positive, when more than 200 colonies in 100 ml water, and negative when less than 200 colonies in 100 ml water.
The chloride test was done as follows: a solution of silver nitrate (an excess) was added to 100 ml of the water sample. The precipitation of silver chloride was filtered and the dried filter plus precipitation was weighed. Of course, the dry filter had to be weighed beforehand also. The students were given a simple formula in which they had to fill in the number of grams precipitation and with which they could simply calculate the chloride concentration (in milligrams per litre). This, of course, was a very inaccurate method.

Students are explicitly asked to write down in a special appendix everything concerning the tests they are uncertain about. Finally, each group is to write down their test results on a poster in front of the class. This way the test results of all the groups are collected and lined up. I expect students to get an overview this way and they might see that the test results show certain variability (see section 4.3).
Again, the authentic practice is used in the form of a simple evaluative question, to induce a specific knowledge need. The students are asked the following questions:

A. *What are the test results?*
   E-coli bacteria:
   Nitrite:
   Chloride:
   Acidity (pH):
B. *Can you judge based on these test results whether your water meets the criteria for drinking water?*
C. *If not, what type of information do you need?*

Question B should trigger students to realise that although they intuitively know that they need to compare their test results with some sort of reference, they do not know what would be an appropriate reference in this case. Question C forces the students to put their knowledge need into words (3B).

Evaluation of the episode question (incl. wrapping up: 5 min)
When the teacher asks the students if they can judge the water quality now, I expect that the students realise (if not already) that they need a reference to *compare* their test results with. This naturally brings up among students the quantitative question '*How much* of these four things are allowed in drinking water?'. So instead of expecting this quantitative question and the question 'What is allowed in drinking water' to emerge in episode 2 as part of the knowledge need (which did not work in the previous version of the design, see section 4.5.2, context and need-to-know, phase 2), it is now coupled to the test results and the induced need to compare.

Wrapping up
The teacher can now summarise the input of the students into the next steps (a new one and refinements of the first two) of the procedure:

Steps of the procedure		
1.	Determine water function, this determines water quality criteria	**What is allowed in drinking water? How much of this is allowed in drinking water?**
2.	Test water sample: does the water quality meet the criteria?	What does the water sample contain? **How much of this does the water sample contain?**
2.	**Compare**	

5.3.4 Episode 4: Does the water meet the quality criteria for drinking water?

Functions of this episode

Students are to link their doubts to the following categories:

4A Trustworthiness of the limited list of parameters and norms.
4B Trustworthiness of the test results.

Justification of content in the light of the episode functions

Situation: all the test results are collected on a poster in front of the classroom. Students (should) have written down their doubts in the special appendix.

The doubts of the students form the basis of this episode and are to be categorized as follows: A. Doubts about the test results, and: B. Doubts about the limited list of parameters and their norms. A. is used to let students link their experiences to the topic of accuracy and own performance (4A), B. is used to let students link their experiences to the topic of 'legal list of drinking water quality parameters and norms' (4B). I expect that students will appreciate these topics as functional in the process of judging water quality adequately, because they are now both, in contrast with the previous version (see 4.5.2, '*Context* and *need-to-know'*) truly rooted in their experiences, namely their own doubts that need to be solved.

Justification of the interaction structure in the light of the content

The teacher prepares the context by collecting the doubts of the students raised when doing the tests. The teacher, I expect, needs to use the remarks of students to lift them to the more abstract level of 'Trustworthiness of the test results: accuracy and own performance', or 'Trustworthiness of the list of parameters to be tested: are there no others? Why these four? Why these norms?' (4A and B). I expect that students will have no difficulty with seeing the logic of such a categorization. The interaction structure, which fits the purposes and complexity of the activities, is 'categorize input' as the teacher needs to collect the student input and categorize the doubts properly (CaI, table 5.1, appendix 1).

Expected unfolding of the episode

Setting the stage for the episode question (5 min)

The teacher points out in the students' materials the list of official water quality criteria for this case (the list of norms belonging to the four parameters) and bridges the previous episode with this one by saying something like: 'We have compared the test results with the official norms, can we now answer our question?'

Addressing the episode question: Does the water meet the quality criteria for drinking water? (10 min)

The main evaluative question is used yet again to induce the next specific knowledge need. In this case, the students are expected to be uncertain about their judgment because they are uncertain about the test results and about the list of four reference parameters. These doubts, which form the knowledge need because they need to be solved in order to be able to achieve a judgment (the purpose), are made explicit in learning activities. These activities follow the same general pattern as the previous learning activities in which the evaluative question was used to induce a knowledge need (see the introduction).

Students answer the following questions first with their group:

A. *Does your water sample meet the official list of water quality criteria? Yes/no*
   I am/am not sure about this, because…

B. *Suppose your water sample does meet the official list of water quality criteria, would you trust it and drink the water?*
   I would /would not trust it, because…

I expect that these questions will trigger students to express their doubts in a peer-discussion:

A. With respect to the test results: 'We could not see very well what the concentration was, we are not sure', 'We are not sure if we did the test correct' 'We did not follow the prescription all the time' (based on findings with the second version, see 4.5.2) (Ib).
B. With respect to the limited list of parameters and norms, of which they do not know the background. I expect students to accept that this is tested in reality, but still to be uncertain when asked whether they would actually dare to drink it. I base this on previous findings, which indicate that students do not trust the limited list (see the discussions that emerged when students found out that mercury was not tested routinely, see section 4.5.2). However, I also expect that some of the students will dare to drink it (Ia).

Evaluation of the episode question (15 min)
The teacher collects the answers of the students. Based on findings with the second version, I expect question A to bring out the doubts concerning the test results (accuracy and own performance) and question B to bring about doubts concerning the *limited* list of tested parameters. The input of students will be on the level of: 'We could not see clearly if the concentration was more closely to X or more closely to Y', or 'We stirred three minutes instead of two, it that ok?'.
The teacher needs to categorize these doubts as follows, as I do not expect students to be able to do this by themselves:

A. Doubts about the test results
A-1 Did we do the tests correctly?
A-2 Are the tests accurate enough?

B. In securities about the limited list of parameters and their norms
B-1 Is this list sufficient? Why are there no other parameters on it? Why *these* four?
B-2 Are the norms trustworthy?

Wrapping up (5 min)
The teacher summarises the student input as the next step of the procedure, in such a way that the students recognize their input:

Steps of the procedure

1. Water function determines water quality criteria	What is allowed in drinking water? How much of this is allowed in drinking water?
2. Test water sample: does the water quality meet the criteria?	What does the water sample contain? How much of this does the water sample contain?
3. Compare	**Do we trust the test results?** **i. Did we perform okay?** **ii. Accuracy**
4. Judge	**Do we trust the fact that only four parameters are to be tested?** **How are these four parameters selected?** **How are the norms established?**

5.3.5 Episode 5: Do we trust the list of tested parameter and their norms?

Function of this episode

Students address the episode question by learning how the list of four parameters and norms to be routinely tested in this case is established (function 5).

Justification of content in the light of the episode function

Situation: the poster with doubts about the official list of parameters and norms to be tested and the list of test results is in front of the classroom.

To address their doubts about the four parameters, students have to develop a notion of the argumentation behind the choice of the four parameters:

c. Together they give a good enough impression of the chemical and bacteriological quality of the water.
d. There *is* a longer list of parameters and norms, but all those other parameters on the long list of drinking water quality criteria are *improbable*.

In other words, if one of the four standard parameters exceeds the norm, it is a sign that the water *might* be polluted (with something else). The implication of this idea is also that there *are* doubts, although limited. There are doubts about the choice for the four parameters (you just cannot test *every* possible parameter). Moreover, there are doubts captured in the norms. I think it is difficult for students to appreciate at once what 'good enough', as described above, means.

To fulfil the function of this episode, students should learn that there actually *is* a much longer list with all kinds of possible parameters and norms for drinking water quality.
The argumentation behind the short list of four parameters comes down to: together they give an accurate enough view of the chemical and bacteriological water quality, each of the four parameters adds to this. To be able to understand why the other parameters on the list are *not* tested they should learn about two things.
First, they should learn about the different types of argumentations behind the parameters on the long list, which should show them what the other parameters are about and that there are in fact parameters on the list that are very poisonous. Secondly, students should learn why the other parameters, and especially the very poisonous ones, are not tested routinely. Students are to conclude is that the four are good enough, because they are indicators for pollution and at the same time because the other parameters are *improbable*. In addition, students probably realise (they did in the previous version, see 4.5.2) that testing all parameters would be practically impossible: too time consuming and very expensive.
To achieve all this among students, they are to find out about the argumentation behind a selection of parameters from the long list which:

2. Reflects the different types of argumentation.
3. Includes a very poisonous one

These criteria have led to the following selection of parameters of which students have to find out why they are on the list (next to the list of four), but are not tested routinely:

1. Nitrate: not more than 50 milligrams per litre
2. Mercury: not more than 1 microgram per litre
3. Iron, not more than 0.2 milligrams per litre
4. Calcium, at least 60 milligrams per litre

These four parameters reflect different types of argumentation:

Nitrate is on the list for health reasons, but research does not agree on the precise effects of nitrate and how poisonous nitrate actually is. The norm is set partly by the fact that ground water and surface water is polluted with nitrate anyway due to extensive use of fertilizer in agriculture, and it is impossible to get it out of the water.
Mercury is very poisonous, which is reflected in the very low norm. The norm is based on the minimal risk value, which means that when 1 microgram per litre mercury is present in drinking water, there is a minimal risk with average consuming, of one person in a million to die. This parameter is not tested.
The norm for iron is not for health reasons but for esthetical purposes, the water will colour brown if the norm is exceeded.
Calcium concerns a minimal norm. Drinking water must contain at least 60 milligram per litre calcium, for health reasons.

Justification of the interaction structure in the light of the content

Because the content-related progression is expected to be complex for students to fully grasp at once, the content is divided over a series of activities, which gradually lead to the full scope of the argumentation behind the choice for the four parameters. The evaluation part consists of evaluative activities (see below).
I expect, based on the previous version where this did not happen either (see 4.5.2), that not all groups cover all different arguments in the evaluation part. Although students easily proved to appreciate the logic of the arguments (see 4.5.2), an exchange part is needed to line them up (see below). The teacher is needed to guide the class discussion in which the students exchange ideas finally in the direction of the appropriate conclusion about the selection of the four parameters and about the final question: Do you now trust the list of four?
The appropriate interaction structure, matching the complexity and purposes of the activities, is: thinking-sharing-exchanging (TSE, table 5.1, appendix 1)

Expected unfolding of the episode

Setting the stage for the episode question (40 min)

The teacher introduces the series of learning activities of this episode by bridging the central question of this episode: 'Can we trust this official list to be adequate?' (which is also the title of the chapter) to the doubts about the official list of parameters and norms to be tested: 'Why only these four?'. The teacher introduces the legal list of drinking water quality parameter and norms, to show students that there *are* actually much more parameters and norms for drinking water. In the first instance, I expect this will raise the questions: 'Why are the four parameters on this list?' and 'Why are all those other parameters on the list?' which students will have and want to find out first. The teacher might introduce the activities using these specific questions, thereby indicating that 'If we want to find out why the four parameters are selected from this list and, if this is to be trusted, we first need to answer these questions'

The teacher also explains the format of the activities: the different groups are going to work on the tasks. All groups get information maps. In these maps, all kinds of documents are collected, from newspaper articles to official reports of toxic-effects of certain parameters to consumer tests. The documents contain information on a selection of parameters from the legal list of water quality criteria and specifically why the four parameters are on the list to be tested (and not other parameters, or more parameters), and why the norm is what it is.

Addressing the episode question: Do we trust the list of tested parameter and their norms? (40 min)
Students are to work in their groups on the tasks.

Activity 6 of the student materials
In the first task, they have to find out from the information maps, of all four parameters, that are routinely tested why they are on the list and why their norm is what it is. In the introduction of this task, E-coli is worked out as an example. The questions are as follows:

Why is the acidity of the water on the list?
Because…
Why must the acidity be neutral (between 7-9,5)?
Because…

Students answer the same questions for chloride and nitrite. I expect that students are able to find in the information maps the following, because the sources almost literally present this information:

Chloride is an indication for the quality of the process of drinking water purification. When exceeding the norm, something might have gone wrong and the water might not be properly purified.
Nitrite, when exceeding the norm, is an indication for possible organic waste
Acidity is a more general indication of the chemical quality of the water. When the acidity exceeds from the norm, it might be a sign that the water is polluted.
E-coli (the example) is an indication of the bacteriological quality of the water. If the norm is exceeded this might be a sign that the water is infected with faecal pollutions. Besides this, if the norm is exceeded people can acutely get seriously ill.

Based on this kind of answers, I expect that the students have developed a notion of: if these four parameters exceed the norm, than this is a warning that something is wrong.

Activity 7 of the student materials
In the next task students have to answer the same questions for a selection of parameters and their norms of the official, legal list of drinking water quality parameters and norms, which is described in the student materials: nitrate, mercury, iron, calcium. They will see that 1) the list is a long one, and 2) different parameters are on the list for different reasons, and that the norms are established for different reasons 3) that some of the parameters are actually very poisonous, but not tested.

Activity 8 of the student materials
The teacher uses this next task, called 'The establishing of parameters and norms', to let students summarise their answers, using the question: 'What type of argumentations have you found? . The reason for this is as follows. Students have worked on the tasks for considerable time, it will take too much time to go through all the answers, but students must be given the feeling that their efforts actually serve the unfolding instructional practice. They have seen that there *is* a longer list, what the types of argumentations are behind this longer list and that there are in fact very poisonous parameters on the list, which are not tested routinely. In other words, they are worthwhile to discuss in a sense. The task consists of the following questions:

A. *Did you encounter in your opinion surprising reasons for why a certain parameter is on the legal list? Yes/no, namely*
B. *What sort of reasons did you find for settling a norm?*

Evaluation of the episode question (10 min)

The evaluation, for which the main activities form the basis, comes down to formulating a founded answer to the question ‘Do we trust the list of tested parameter and their norms?’. Students now know about the type of argumentation behind the long list and specifically the four selected ones. They know that the other parameters are sometimes very poisonous, but not routinely tested. They are explicitly asked to think about this in their groups.

I expect that most students will conclude that, besides that the four give an accurate impression of the water quality, it *must* be that those very poisonous ones are improbable; otherwise, the risk would be too high. I also expect that at least some students will additionally mention that it would be too much work to test all the parameters anyway. Maybe they will also mention that it would be too expensive (for such a low risk).

To direct the students to these considerations, the phase consists of two activities (9 and 10, see below) which further develop the idea among students (what it means) that the four parameters are thought to be sufficient for routine testing, and other parameters are not probable, students have to think 1) about why the very toxic parameter mercury is not tested routinely and 2) of a situation in which the very toxic parameter mercury *is* tested.

Activity 9 of the student materials

The first activity of this phase consists of the following questions:

1. *Why does the lab routinely test the drinking water samples only on these four parameters, and not, for example, on the very poisonous mercury?*
2. *Can you think of a situation where the lab will test the water on extra parameters?*
3. *You tested the water on these four parameters yourself; do you now consider this to be sufficient? I consider this to be / not to be sufficient, because…*

Activity 10 of the student materials

In the second activity, called ‘Why these norms?’, students are to answer the following question:

Do you consider the norms on the list of drinking water quality criteria of the lab adequate?
I consider / do not consider the norms the lab uses sufficient, because…

The students think about, discuss and answer the questions with their group. After this, the answers are exchanged and elaborated upon by the teacher in an evaluative class discussion. The exchange is, I expect, necessary to direct the students to the full scope of the appropriate conclusion: “In the authentic practice the four parameters are considered sufficient for routine testing, because they provide for an accurate enough indication of the chemical and bacteriological drinking water quality. The legal list includes all possible parameters, the probability that the purified water contains too much of one of those parameters is very low. Besides that, it is not possible to test the water standard on all the parameters on the list. That would be too time consuming and too expensive".

Wrapping up (5 min)
The teacher wraps up the episode in the class discussion on students opinions on the question: do *you* trust the list of standard tested parameters and their norms to be adequate (now you know about the type of argumentation behind the four parameters and norms)?
Finally, the procedure is not extended, only the terms 'parameter' and 'norm' are added. The teacher wraps up the episode by adding these terms to the procedure:

Steps of the procedure

1.	Determine water function, this determines water quality criteria drinking	What is allowed in drinking water? How much of this is allowed in water? **Selection parameters and norms**
2.	Test water sample: does the water quality meet the criteria?	What does the water sample contain? How much of this does the water sample contain?
3.	Compare	Do we trust the testresults? i. Did we perform okay? ii. Accuracy
4.	Judge	Do we trust the fact that only four **parameters** are to be tested? How are these four **parameters** selected? How are the **norms** established?

5.3.6 Episode 6: Do we trust our test results?

Function of the episode

Students address the episode question by developing a notion about reliability and accuracy of the used test methods and how this might influence their judgment.

Justification of content in the light of the episode functions

Situation: the posters with the categorized doubts and the poster with the test results are still on the classroom wall. The first question: 'Do we trust the limited list of *water quality criteria?' has been addressed in the previous episode. The question 'Do we trust the list of test results?' is still open.*

The question 'Do we trust the test results?' is addressed in this episode. The doubts of the students show that students have intuitive notions that their performances on the tests (did we do it right?) and the accuracy of their measurements (we could not see clearly what the right concentration was), might form a problem if they want to interpret their test results correctly in formulating a judgment.
I expect that students feel that their doubts need to be addressed. In order to address them properly, students need to develop the following notions.

1. Based on their intuitive notion that their performances on the tests might form a problem in the sense that it might influence test results and therefore judgments, students should develop a notion of the type of strategies that address this problem, e.g.:
    e. See if performing the test in the exact same way with a known solution gives divergent test results, or
    f. Perform the test precisely according to the prescript with the same sample and see if the results differ

Students are to come up with these strategies, and appreciate their more general usefulness in 'checking tests'.
To achieve this, students should reflect on their own doubts about their performances. They discuss the strategies with their classmates, who reflected on *their* own, sometimes different, doubts, in a class discussion. Because all students did the same tests, they should easily empathize with each other's doubts. This will contribute to their appreciation of the usefulness of the activity: it actually solves a genuine problem they all have. I expect that the general character of the strategies, applicable in all the different exemplary doubts, will contribute to their appreciation of their more general usefulness.

2. Based on their intuitive notions that the accuracy of their measurements should be considered when formulating their judgments, the students should learn about how to do this. They are to consider:

    a. The number of tests: the more tests, the accurate the average of the test results
    b. The degree of variability in test results: less variability more accurate and (see section 4.3):
    c. The position of the test results with respect to the norm: the closer the test results, including their variability are to the norm, the more risk the true value exceeds the norm.

With respect to the accuracy of the test results, in order for students to consider the accuracy of the test methods in their judgments as described above, the choice of content must meet the following criteria:

A) It should motivate them to find out more about the principle and accuracy of the colorimetric test method, in which eyes are used as measuring instrument: how accurate are these type of tests?
B) It should contribute to the creation of a common situation specific experience among students that the accuracy is reflected in the extent of variability in test results (see section 4.3 for an explanation of the concept of variability). Moreover, it should contribute to the realisation that the concept of 'variability' would help them to reflect on the accuracy of their own test results in the light of formulating an adequate judgment. Also: the more test results, the accurate picture
C) It should strengthen more general notions of combining variability in test results and the position of the test results with respect to the norms as strategies for formulating a judgment in which the accuracy of the test results is accounted for (see section 4.3).
D) It should contribute to creating a common experience among students that the roughness of the scale of a test method reflects in fact also the accuracy of the test method and might influence a judgment.

These criteria led to the following choices:

1. The doubts of the students, which concern the accuracy of the test methods, are used to meet criterion A. These doubts, similar to the doubts about the performance of the tests, are raised when the students were asked if they could formulate a judgment yet. They form a knowledge need in the sense that they need to be solved to be able to formulate such a judgment. This should motivate students to get involved in the activities (because they will solve this problem)
2. To create a common experience as described above under B, a colorimetric copper-sulphate test, in which unknown copper sulphate solutions are to be estimated, is used as an exemplary colorimetric test method (similar to the pH and nitrite tests). This will create a common experience with these type of tests, as all students do the same estimations, and a range of test results with a certain variability (like in the previous version (see 4.5.2). Similar to the test results of the genuine water sample, the 'true values' of the copper sulphate solutions remain unknown, which strengthens the experience that variability stands for accuracy, because the range of estimations 'is al we can go by' and the more estimations, the more accurate idea is possible of what the true value might be
3. The exemplary set of test results, which meets criterion C, is the set of chloride test-results on the test-results poster, which is on the classroom wall. In addition, the variability in pH and nitrite test results might be used here. I expect that using the variability in chloride test results will contribute to making explicit the more general (already existing) notion that variability reflects the accuracy of the test method: the greater the variability, the less accurate. The norm for chloride and the position of the chloride test results with respect to the norms should contribute to the notion that this is a second criterion for a judgment.

4. The E-coli test meets criterion D. It forms an example of a test students have done and which has a rough scale (positive means more than the drinking water norm of 10 colonies per litre (bad), negative, means less than the drinking water norm of 10 colonies per litre (good). Also, the consequences of a wrong estimation might be severe in this case.

Justification of the interaction structure in the light of the content

I expect the scope and depth of the question 'Do we trust our test results?' and the impact of this question on the judgment of the water quality 'Does this water meets the drinking water quality criteria or not?' to be complex for students.
The first version of the design showed that it is certainly not a topic, which students easily grab and integrate meaningfully in their judgments of the water quality. The question 'Do we trust our test results?' is therefore addressed gradually (see above). In different successive activities, students have thought about all aspects of this question: performance, variability and position of the test results, scale of the test method. They have discussed their thoughts with their group-mates and in a class discussion. In the evaluation part, I expect that the teacher is needed to bring all the output of the previous activities together when students are finally to formulate their judgment of the drinking water quality. The teacher should make sure that students pay attention to every issue when they evaluate their water quality test results. He is needed to direct students in their final reflection on their test results as this has proven to be a difficult issue (section 4.5.2). To achieve this, the input of students (per group), which I expect in general not to cover all the different aspects 'at once', should be exchanged in a class discussion. The interaction structure, which matches the purposes and complexity of this episode, is thinking-sharing-exchanging (TSE, table 5.1, appendix 1)

Expected unfolding of the episode

Setting the stage for the episode question (5 min)

The teacher refers to the poster on which the doubts of the students were listed. S/he might emphasise that the question 'Do we trust the limited list of water quality criteria?' has been addressed and that the question 'Do we trust the list of test results?' is still open and need to be solved to be able to judge the water quality properly.
The teacher explains that the episode question is addressed in two parts, each referring to a different type of doubts. Part A is about doubts concerning the question 'Do you trust the tests and the way you performed them?. Part B is about doubts concerning the question ''Do you trust the accuracy of the test results?'

Addressing the episode question: Do we trust our test results? (40 min)
Students start with Part A: 'Do you trust the tests and the way you performed them?'

Setting the stage
The teacher, when setting the context for the respective activity, should refer to the doubts of the students, which concern the tests and the way they performed the tests.

Students should reflect on the following questions with their doubts in mind, and with the purpose to solve them (or at least address them).

Activity-11A/question: 'What is a good strategy to test the tests and the way you performed them?'
Students are to think over with their groups the following question:

Can you think of a way to find out if the test and the way you performed the test really tested what you wanted to test?

Based on previous findings, I expect students to come up with answers like: 'You might check the results by doing a different test', 'You might sent the water to a sophisticated lab', 'You might test a known solution'.

Evaluation/ setting the stage for activity 11B/question
The teacher is needed to collect and discuss the input of students: what would be a good way to test the tests and the way they are performed? I expect that students are able to think of 'testing a know solution'. Some of them might suggest sending the sample to a lab (see previous version, 4.5.2). I expect that the students see the need to apply the strategy in the instructional version of the practice in order to genuinely solve the doubts. I also expect students to appreciate the general usefulness of the strategy, as it might be used to solve different doubts. Suggestions like 'sending the water sample to a lab' might now be judged not practical (in the instructional version of the practice).

Activity11B/ question: 'Can we trust the tests and the way we performed them?'
Next, when the students have decided, guided by the teacher, on adequate strategy to address the doubts, the teacher actually needs to apply it. Students have to think about the results in terms of: are the tests and the way we performed them trustworthy?

Evaluation
The teacher evaluates the activity by letting students discuss the following question:

When you think back, do you now trust the tests you have done and the way you have done them?
Yes/no, because…….

The teacher should make sure that students answer this question using arguments like: we are sure, because when we tested the way we performed our test on a known solution, the results were accurate, so it did not matter that we 'stirred two instead of three minutes' and so on.

Wrapping up
The teacher wraps up the activity by pointing out that part of the doubts are addressed now, but there are still doubts, which need to be dealt with.

(Next, students address Part B: 'Do you trust the accuracy of the test results?')

Setting the stage
The teacher draws the attention of the students to the doubts, which reflect the accuracy of the colorimetric pH and nitrite tests. These doubts are addressed next. The teacher explains that in the next activity the principle and accuracy of such tests will be explored.
In the student materials, the activity is introduced as follows:

> *As you have noticed, the Merck-kit tests are based on the following principle: the higher the concentration of the solution, the more intense the colour of the solution. That is why these type if tests are called colorimetric tests. To give you an impression of the accuracy of such tests, you will get a series test tubes containing different solutions of copper sulphate. They are numbered 1 to 10.*

Activity 12/question: 'How accurate are the colorimetric nitrite and pH test results?'
The students are to answer the following question with their group:

Place the numbered test tubes (1-10) in order of increasing concentration.

The teacher collects the written answers of each group. I expect, because the differences in intensity are sometimes difficult to see, that the students will not always agree with each other, thus experiencing the doubts, and that the different groups will come up with different answers (at least in some cases).
Every group is now to put the test tubes in the right order and two unknown solutions of copper sulphate, nr 11 and 12, are handed out. The groups are to discuss the following question:

Estimate the concentration of the solutions 11 and 12.

The groups are to write down their estimations on the black board. I expect differences of opinion within the groups and variability in the results of the groups.

Evaluation
The teacher is to bring forward the variability of the results on the black board and discuss in a class discussion with the students why they think this is so. I expect that this will lead to remarks like 'it is hard to see', and 'we could not agree', which now can be easily linked by the teacher and therefore by the students to accuracy of the colorimetric method using eyes as an instrument. The fact that the teacher will not mention the true concentrations, because he does not know it (and I expect that students will really want to know), will further contribute to a notion among students that the results represent the true value in a way and that this true value therefore, must lie somewhere between the extreme values of the cloud of test results.
The teacher can summarise the above as follows, asking students if they agree:

‘the extent in variability represents the accuracy of the test method’

Wrapping up
The teacher wraps up this part of the episode by pointing to the variability in test results of pH and nitrite with the question:
‘Think back to the colorimetric tests you did with the Merck kit. Do you now consider them accurate enough?’

This question is meant to include the position of the test results with respect to the norm in the argumentation. The students are to develop a notion that both variability in the test results and position of the ‘cloud of test results’ with respect to the norm determines if the test method was accurate enough. I do not expect this to be obvious to students

Setting the stage
The teacher draws the students’ attention to the chloride test results on the test results poster and the extent in variability in these test results.

Activity 13/question: ‘How accurate are the chloride test results?’
The students are to think about the following question:

Compare the chloride test results of the different groups.
What do you think of the accuracy of the chloride test?

I expect that students are generally able now to give an argued opinion about the accuracy of the chloride test, in terms of variability in the test results.

Evaluation
Students exchange their views on the question ‘What do you think of the accuracy of the chloride test?’ in a class discussion. The teacher has to make sure that the students, again, explicitly link the variability in test results to the accuracy of the test.

To include also here the position of the test results with respect to the norm in the argumentation, the teacher posses the question to the class:

Do you consider the chloride test in this case accurate enough?
I think that the chloride test was/ was not accurate enough, because…

At this point, I expect students to be able to argue if the chloride test was accurate enough in this case, for example: ‘Although the variability in test results was considerable and the test method therefore not very accurate, all the test results were far below the norm. So in this case, the method was accurate *enough* (and visa versa).

Setting the stage

The teacher should now draw the students' attention to the E-coli test results and the fact that variability is not an issue here, because the test results can only be negative or positive. This roughness of scale **(?)** is the problem.

Activity 14/question: 'How accurate are the E-coli test results?'
First students are to answer the following questions with t heir group-mates:

What does the prescription of the E-coli test say about the accuracy of the test method?
The prescript says:…..

Do you consider this test to be accurate enough?
I think that in this case the E-coli bacteria test was /was not accurate enough, because…

I expect that these questions will direct at least some students to think about the accuracy of this test in terms of 'roughness of scale'. When this comes forward in the class discussion, I expect students to see the point of the argument 'E-coli test is not so accurate, because the scale is very rough. The question is how great the risk is that the test gives a negative result when the 'true value' is actually above the norm.

Evaluation
Students exchange their views on the question 'What do you think of the accuracy of the E-coli test?' in a class discussion. The teacher needs to guide the discussion to the concept of 'roughness of the test' and the argument described above.

Wrapping up
The teacher wraps up the episode by asking students if they think their opinion on the accuracy of the E-coli test should have consequences for their judgment:

Do you want to reconsider your judgment? Do you consider this test accurate enough?
I think that in this case the E-coli bacteria test was /was not accurate enough, because…

I expect that students will say it gives an indication but you cannot be sure unless you know the risk of the test being wrong.

Evaluation of the episode question (15 min)
The teacher points out that the students have now explored all the aspects of the question: 'Do we trust the list of test results?'. S/he points out that having answered properly 'how much of what' does the water contain, means including an account for the trustworthiness of the test results. The class must now, like in the authentic practice, reach to a conclusion about the question: 'Is this water clean enough to drink'. The groups think about this question, sharing their views.

I expect that in the class discussion, students exchange their views, which will lead to an argued conclusion about the water quality. The exchange is guided by the teacher and is needed, to come to a conclusion, which includes a full account of the accuracy of the used test methods. This includes an account of the accuracy of the test results in terms of variability and position (pH, nitrite, chloride) and roughness of the scale (E-coli)

Wrapping up (5 min)
The teacher, based on the conclusions of this episode extends the procedure with the next steps:

Steps of the procedure		
5.	Determine water function, this determines water quality criteria	What is allowed in drinking water? How much of this is allowed in drinking water? Selection parameters and norms
6.	Test water sample: does the water quality meet the criteria?	What does the water sample contain? How much of this does the water sample contain?
7.	Compare	Do we trust the test results? i. Did we perform okay? ii. Accuracy: **variability and position of the test results**
8.	Judge	Do we trust the fact that only four parameters are to be tested? How are these four parameters selected? How are the norms established?

5.3.7 Episode 7: Does the water quality meet the quality criteria for drinking water?

Functions of the episode

Students solve the exemplary case (I) and summarise their procedural knowledge (II). They feel the need to test the usefulness of the procedure in the other exemplary cases because they expect this to be the case (III).

Justification of content in the light of the episode functions

Situation: the students have gone through all the steps of the procedure. The poster with the procedure is completed. Students know what advise they should bring out, based on the test results and considering the accuracy of the test methods and reliability of their performances on the tests.

To summarise the procedure in a way that reflects the authentic practice, students are to fill in a lab-report, which is based on the type of reports used in the authentic practice. The different steps of the procedure are all represented in the report as necessary information to be able to interpret the report. The report merely reflects what students already know, namely: the procedure (II). The choice to let students fill in such a report is based on the following arguments:

1. It strengthens the relationship with the authentic practice in which such reports are a common way of communicating results.
2. It explicitly emphasises for the students' that the exemplary case of judging the water quality from water of the residential area with two water nets has been brought to a close. The procedure has been determined and all activities have resulted in a final judgment in this exemplary case (I).

This explicit emphasis should form the basis for the next episode, in which students address the question 'To what extent does the procedure we used apply in the other exemplary cases of the orientation phase in which testing and judging water quality is a central activity?

Justification of the interaction structure in the light of the content

I do not expect this to be a very complex activity. However, it is important that every student is involved in thinking about the report within his or her group, because of the purposes described above under point 1 and 2. The teacher is needed to emphasise explicitly that the exemplary case is now solved, and to point back to the other examples of the orientation phase, with the question: 'What about them?'. The interaction structure, which matches the complexity and purposes of this episode is thinking together-exchanging (TTE, table 5.1, appendix 1)

Expected unfolding of the episode

Setting the stage for the episode question (10 min)

The teacher points out to the students that in the genuine practice the lab now would bring out an advice. The teacher introduces the report, based on the kind of reports, which are used in reality in such cases, in the students' materials. Each group is to fill in the report.

Addressing the episode question: Does the water quality meet the quality criteria for drinking water? (15 min)

The groups are to fill in the standard report. All steps of the procedure, which form the argumentation of the final advice, can be found one way or another in the report. By filling in the form, I expect that the students will go over all the steps again and experience that *their* procedure is in fact based on the procedure of the practice: this is how they find out if the water quality satisfies the criteria (I and II).

Evaluation of the episode question (10 min)

The teacher discusses the reports with the students from the perspective: 'we now worked out and applied a procedure for the example 'household water or drinking water?' (I). By referring to the other exemplary cases of the orientation phase, which (should have) motivated students for to learn what is involved in the first place, I expect that the question 'to what extent does the procedure apply in the other examples?' naturally emerges among students.

5.3.8 Episode 8: To what extent can the procedure we used be applied in the other exemplary cases of phase 1?

Function of the episode

Students are to reflect on what they have learned: a procedure for how water quality is judged in standard situations.

Justification of content in the light of the episode function

Situation: the students brought the exemplary case of the two water nets to a closure. I expect that they now wonder to what extent their procedure applies in the other examples.

Each of the student groups is to apply the procedure for a different exemplary case of the orientation phase. They need to realise by now that the only thing they need to look up is the specific water quality criteria for the respective water function and relevant parameters and norms to be tested (as test methods are given, like in the worked out example).
This, I expect, contributes to their appreciation of the usefulness of the procedure. It also contributes to their insight in how the procedure helps in establishing the type of information they need in order to be able to apply the procedure for any given example (namely the specific water quality criteria and parameters to be tested of a specific water function). This should form the basis of a reflection on what they have learned and on the worthiness of what they have learned

Justification of the interaction structure in the light of the content

The different groups each work out an example, which I do not expect to be very difficult for them, maybe only time consuming. However, all students should be involved in this episode; they all should to reflect on what they have learned. The teacher is needed to guide the reflection in the evaluation part as described above, implicitly (or better still explicitly) addressing the question, which was raised in the introduction part: what have we learned and was it worthwhile?'. The interaction structure, which matches the complexity and purposes of this episode is thinking together-exchanging (TTE, table 5.1, appendix 1).

Expected unfolding of the episode

Setting the stage for the episode question (10 min)

The teacher might bring out the case descriptions and water samples of the examples of the orientation phase. S/he might also bring out the posters the different groups made of the different exemplary cases of the practice in this phase. The different groups are each asked to apply the procedure for one of the exemplary cases.

Addressing the episode question: To what extent can the procedure we used be applied in the other exemplary cases of phase 1? (40 min)
The groups are asked to produce a standard form for their case similar to the form they filled in for the drinking water case. They are asked to write an instruction for the student group who is to investigate the water quality. The groups are given existing websites on which for each case information can be found about parameters and norms, which are routinely tested or monitored in reality, often including a motivation. I expect this to be a time consuming but not a difficult task for students. I think they know by now that the differences are laid in water function and water quality criteria, and therefore in which are relevant parameters and norms to be tested.

Evaluation of the episode question (15 min)
The teacher is expected to evaluate this activity by addressing the questions of the orientation phase:

How, do you think, do they check if the water satisfies the criteria (what steps are involved)?

Wrapping up
The module has come to an end now. The teacher wraps it up by initiating a discussion on 'what have we learned and was it worthwhile?'
The teacher might bring back the posters of the orientation phase, on which students were to 'look forward' and let students discuss it, maybe compare their answers with the procedure poster.

5.4 Concluding remarks

I have presented the concrete levels of description of the idea of creating an instructional version of the authentic practice of judging water quality.

Next, in chapters 6 and 7, I will present the findings with the third version of the design. In chapter 6 I will present the data processing and presentation of the third trial. In chapter 7, the actual evaluation is presented. Based on this detailed empirical evidence, I will draw conclusions on the extent in which the expectations were realised and the functions of each phase fulfilled.

Chapter 6

The final test: methods of evaluation

6.1 Introduction

In chapter 5, I presented the scenario of the third and final version of the design based on what the three characteristics of meaningful, *context*, *need-to-know* and *attention for student input*, finally evolved in: an instructional version of the authentic practice of judging water quality and interaction structures. I showed the adaptation of this authentic practice for education at the level of functional episodes and the teaching-learning activities, which includes the role of the teacher in the form of interaction structures.
In chapters 6 and 7, respectively the evaluation methods and the results of the trial with this version of the design are presented and discussed.

In chapter 6, I will present how this trial with the third version of the design was set up. First, I will briefly describe in section 6.2 the school involved, its background, and the two classes that participated. Next, in section 6.3 I will discuss the teacher's preparation trajectory for his role, which is now given shape by interaction structures. In section 6.4 I will give an account of the way the data were collected, processed and will be presented in light of the evaluative research questions: which data were collected, how the data were analysed and are how the findings are presented in chapter 7.

The actual evaluation of the realised teaching-learning process is presented in chapter 7. I will answer the question to what extent the design lived up to the aims, in the intended way. In other words, to what extent is this third version of the design was adequate in meaningfully guiding the students to the intended learning goals: insight in the procedure for judging water quality and the related authentic practice. The scenario serves as a reference in this process: for each step of the teaching-learning process, specific argued expectations were formulated as to why the specific step can be expected to contribute to the function of the respective episode and finally the criteria for meaningful chemistry education.
I will describe how the respective data were interpreted in order to answer the evaluative questions. This results in section 7.6 in a final reflection on what worked and what did not work and why. This evaluation forms the basis for chapter 8, in which the research question is answered and reflected upon:

What is an adequate teaching-learning process in a module about judging water quality for initial chemistry education (students: 14-15) that properly embodies the three characteristics of meaningful: a motivating context, a proper need to know and a proper attention for student input?

6.2 School situation of the final test

6.2.1 Background school

The school involved in the final test is situated in an average village in the centre of the Netherlands (26.000 inhabitants) and has a Christian identity. On the web site, the school describes itself as ‘Christian with a deep respect for other religions’. In 2003 the school, which has one location, served 1674 students divided over all different possible levels. The average performance of the students, except for the first two years, which score just below average, is in line with the national average performance (Trouw).

The educational culture of this school (school C) resembles more the culture of school A than the culture of school B involved in the previous research cycle. Schools A and C both emphasise in their teaching practice the cognitive development of the students. At both schools A and C, the teacher mostly directs the unfolding of the teaching-learning process at the level of the whole class. Individual students do not get much saying in this. The general teaching style is characterized by whole class instruction (Teacher C, private communication).

6.2.2 Classes

The module was put in to practice in two different classes in the spring of 2003. One class was used as a ‘try-out class’; the other was the ‘focus class’. The factual evaluation is based on the trial in this focus class. The activities in the try-out class were part of the teacher’s preparation (see below).

The focus class consists of thirty students (19 girls/11 boys); the try-out class consists of twenty-six students (13 girls/13 boys). The teacher is mentor of the focus class. The teacher knows the students of this class therefore well and he describes the atmosphere as pleasant. In the try-out class the teacher-class relation is somewhat different in the sense that the teacher knows this class much less and considers this class ‘smarter, but more difficult to handle and less motivated’. He says he needs to put in much more effort to keep control in this class. I observed the differences in atmosphere in these two classes, some weeks before the test took place. In the focus class, the students seemed much more at ease with the teacher than the students of the try-out class. The teacher also seemed more relaxed with the focus class. Based on this it was decided that the mentor class would function as the focus class.

6.3 Preparation by the teacher

The tests with the second version had revealed that I had underestimated and rather neglected the teacher role (see section 4.5.2). In chapter 5, I have described how I addressed the teacher role in the design of the third version, using interaction structures. I will now elaborate on how the teacher in this test did prepare for his role and the use of the interaction structures.

First, I will give a description of the background and start situation of teacher C: how he sees his functioning in the school, how he approaches his participation in this project and how his attitude, ideas and practice in broad outlines did or did not match with the key principles in the design of the module about water quality. The description is based on an interview with teacher C and lesson observations of two lessons prior to the test. Next, I will discuss what I expect to be challenging for the teacher about the intended teacher role, based on this first description, my view of what interaction structures demands of a teacher and what appeared to be problematic for the teachers of school A and B (see section 4.5.2).

In section 6.3.1, I will discuss how I, based on an open interview and lesson observations, determined the start position of teacher C. Next, in section 6.3.2 I will discuss what I expected to be challenging for teacher C about the intended teacher role, based on his start position and what in my opinion the implementation of interaction structures demands of a teacher. In section 6.3.3, I will discuss the preparation trajectory: the main choices, the outlining and the unfolding.

6.3.1 Background and start situation of teacher C

Teacher C has almost 4 years of teaching experience and has two colleagues in the chemistry department. He indicates that he wants very much to develop his teaching practice. He feels that it depends mostly on his individual initiative if and how he can achieve this. In this sense, teacher C experiences a lot of freedom of choice. At the time of this test, teacher C is for example also involved in a cross-curricular project of the Utrecht University which aims at developing a learning line for students for doing research, because it ‘seemed interesting’. In this project, teacher C cooperates with researchers and teachers. He co-develops materials, puts them into practice and reflects on his own practice and on the quality of the design in evaluative interviews with the researchers. These researchers observed his lessons and guided the process and peer discussions with other teachers. His lessons are video- and audio taped in that project, so he is used to being observed.

When asked about his motivation to participate in the water quality project, teacher C explains that he is curious about the Dutch initiative to reform the chemistry curriculum. In his opinion, school chemistry is much too abstract. His concerns and ideas about what school chemistry should be about are in line with the ideas behind the design of the module on Water Quality:

> I am very curious about the new developments in school chemistry. Actually, I think the subject is rather abstract. Many chapters seem to start very quickly with all kinds of abstract ideas and symbols. I wonder what students are supposed to do with that.
> It is often not clear to them what it is used for. For instance, I recently looked at Thieme's [publisher] new method. What did they do? They started every chapter with an article about a social topic and immediately afterwards they shifted to using abstract symbol-language. Therefore, that was far from innovative. I believe it is important that students learn how chemistry is applied, what its use is, especially in the third grade. That entire abstract content, I do not approve of

> it. They had better start with getting familiar with chemistry in practice. On the other hand, it should connect to what they can expect in the fourth grade.

From this, I conclude that teacher C agrees with the basic idea that students should learn about how chemistry functions in society and that he is sensitive for the importance of carefully thinking through the choice of content, because it should match this aim. He indicates that he recognizes the use of those appealing articles in the schoolbook as a trick to attract student attention. I also interpreted the above citation that teacher C expresses that a choice for *relevant* content, that is content that really matches the aim, is more important than preserving the current content. However, he immediately tones this down to “this must not interfere with a proper preparation of students who want to continue their chemistry education”.
Besides interviewing him, I also observed two successive lessons in different classes to get a first impression of how teacher C pays attention to student input. I wanted to get an impression whether teacher C does or does not perform in line with the basic features of interaction structures: does he listen to students, does he give them opportunities to express themselves, and is he able to empathize with their way of reasoning? Does he introduce and evaluate activities in such a way that it is clear to students why they need to get involved in the activities and how does he involve students in these introductions and evaluations? The following general impression of how teacher C handles students input emerged from my observations.

When evaluating teacher C invests a lot of effort to involve students and trying to understand what a student means. He is very patient and gives students frequent opportunities to ask questions and start discussions. He also frequently asks students evaluative questions to see whether they understood or to try to activate specific knowledge. He takes time to listen to student questions, and he answers and encourages students who have difficulty to express themselves clearly. He also really tries to evaluate what the student means, rather than evaluating whether the student uses the correct formulation. He frequently comments on the descriptions in the textbook by putting himself in the position of the students and trying to rephrase texts from their perspective, in their words. The teacher is very accessible to students. The atmosphere is relaxed and friendly, although more in the focus class than in the try-out class.
In both observed lessons, students were supposed to work a period on individual activities, although they could discuss the activities with each other. Teacher C explained that this was part of the normal course of a lesson, after going over homework and introducing new content. Often, because of the long duration of the lessons, such a period of individual work would be interrupted by experiments, either demonstrated by the teacher or done by the students. In the lessons I observed how teacher C, in his instructional role, merely introduced the individual activities and student experiments, with respect to their procedure, by elaborating on the question “What are we going to do and how?”, and hardly on the question, “What are we going to do and why?”. In addition, he scarcely evaluated activities with respect to their purpose: what did we do, what was the reason we did this and what have we learned? Sometimes he only referred to an isolated issue in his evaluation, but many times he

did not evaluate at all or just asked the class if there were any questions. Subsequently he obviously did not use student input in the introduction or evaluation of activities.
Teacher C confirmed these impressions when I discussed them with him. Looking back with him at the video recordings of his lessons (immediately afterwards) he commented his role and said that he sees the point of introducing and evaluating activities with respect to their purpose and involving students in this process, if only to verify if they have understood the content-related progression as intended. Teacher C explained that watching the video from the perspective of the students, made him realise that the purpose of the activities and their interrelatedness were clear to *him*, but that that does not mean that this was clear to *students*.

In summary, I have the following impression of teacher C's start situation: he has experience with being involved in a research project, in which his lessons are videotaped and observed. He also has experience with reflecting on his own practice. He seems very motivated to get involved in similar activities and expects it will improve his practice. He is specifically motivated to get involved in the water quality project, because it matches his personal aims: learn students about how chemistry functions in society.
Teacher C has attention for student input at specific moments in the teaching process. He takes time to let students explain what they mean and makes an effort to understand them. He is sensitive to the student's position, their way of reasoning.
He is not, however, used to pay attention to introducing and evaluating activities in such a way that he prepares a context for activities, addressing for students the questions: "Why are we doing this? How do the different activities related to each other?" Obviously, he is not used either to think about what input the students might give and how to involve students properly in introductions or evaluations. After reflecting on the videos of his lessons, he at least sees the point of paying attention to these issues.

6.3.2 Expected challenges for teacher C

My expectations with respect to what might be challenging for teacher C about his intended role in the trialled lessons were based on my impression of his starting situation as compared to the intended teaching role. I also base my expectations on what happened in the classroom with the three teachers involved in the second round of trials (see 4.5.2). Although the teaching styles of these teachers differed considerably, ranging from mostly teacher centred to more student centred[10], they all seemed to experience the same difficulties (only in different degrees). Kortland described the same type of challenges for the teacher with respect to the teacher role

[10] Teacher centered: the teacher takes the responsibility for the content-related progression most of the time. He/she dictates what are important issues, the right answers, the proper discussions, and so on. Students have little influence.
Student centered: the teacher hands over part of the responsibility for the content-related progression to the students. There is room and attention for the issues they raise, the discussions they start, their opinions, their answers and questions.

in this experiment (Kortland, 2001, p99-100)[11]. This led to the formulation of the following three related challenges:

1. The teacher is less in charge of the content-related progression and has to hand over some of that responsibility to the students. This means that the teacher has to trust that students will come up with intended input, although sometimes maybe formulated not as neatly as the teacher might wish. In addition, the teacher needs to listen to student input and interpret their input from their point of view (prepared by the expectations in the scenario), which means that the teacher must withdraw a bit, to avoid that he himself formulates the answers. This also means for example that the teacher should not use the content as a means to control the class, as this would undermine the basic idea underlying the problem-posing approach (see 3.3).

I do not think that listening to students and taking their input seriously is a challenge for teacher C, as he does this already in his own practice. However, teacher C does this when he tries to activate what he expects to be prior knowledge or when he evaluates homework. I do expect that handing over part of the *responsibility* for the content-related progression to the students, which is a different aspect, to be a challenge for teacher C.
Teachers A1, A2 and B all had difficulty with structurally handing over responsibility to the students to come up with specific input (section 4.5.2). They all (one more than the other) tended to take over and explain or fill in for the students. Kortland describes very similar experiences with the teacher involved in his design experiment. Sometimes the teacher, in his instructional role, is eager to give the right answer; sometimes the teacher is eager to interpret students' remarks in the direction of the expectations of the scenario (Kortland, 2001, p99-106).
I therefore expect the same pitfalls for teacher C, especially when the activities address more complex content (episodes 5 and 6) or have a more reflective character (episodes 7 and 8).

2. In contrast to traditional chemistry lessons, which teacher C is used to, the learning activities in water quality each build on each other in the sense that the outcome of the previous learning activity forms the basis for the next and so on. On a second level, the learning activities are grouped in episodes around episode questions. The episodes have the basic structure of 'setting the stage' –'addressing the episode question'- 'evaluating the episode question'. The students should experience the logic of this episodic story line and the functionality of each of the activities within the story line in the sense that they serve to answer an episode question. This means that the teacher has to pay explicit attention to the story line by using the outcome of the previous episode in setting the stage for the next one and by evaluating each episode using student input properly (challenge 1).

[11] Kortland studied the design and development of a teaching-learning process about decision making from a problem-posing approach (see also section 4.5).

Introducing and evaluating activities is something teacher C is obviously not used to in his own lessons although he seems to see the point of introducing and evaluating activities in such a way that the story line and the way the activities build on each other becomes explicit (see above). To change practice is however a different issue than to see the point of change (Roth, 2003a).
Of course this second challenge is related to the first one in the sense that ‘setting the stage for the episode question’ and ‘evaluating the episode question’ (the basic structure of an interaction structure) are *the* moments when the teacher needs to make proper use of the student input by handing over part of the responsibility for the content-related progression to the students.

3. The third challenge concerns the content: the fact that the teacher is not familiar with the type of content involved and especially the fact that students are to answer an open question: water quality is tested and the results and doubts involved are unknown beforehand. In the preparation trajectory of the teacher this third challenge should be addressed, so that he feels he has an overview, and feels comfortable and at ease about it.

In summary, the preparation trajectory should address three challenges and lead teacher C to:

Aim 1. A general understanding of the design and the main arguments for the activities in the scenario.

Aim 2. Being able to make the story line of water quality explicit by focussing on the functionality of each of the activities in light of the episode question when setting the stage (What are we going to do in the light of the episode question?) and evaluating (Can we answer the episode question now and what might be a next step?), thereby properly using student input.

In section 6.3.3, I will discuss how I expected that these aims could be achieved. I will indicate between brackets which aim is addressed.

6.3.3 Outline and unfolding of the teacher’s preparation trajectory

As I already mentioned in chapter 3, the key problem I had to consider when making choices about the teacher’s preparation trajectory was that he had not been involved in the design process at all. He had no insight in its history and in why the scenario is as it is. Based on this argument the teaching sequence and the (general) reasons that led to it should not only be discussed with the teacher, but he should also practice with it, directed by feedback, as much as possible to gain experience with teaching the teaching sequence and specifically with the interaction structures. This is in line with findings of Showers & Joyce (1980) and Stolk (in preparation) who report that practice and directed feedback are necessary for the effectiveness of teacher preparation trajectory.

The preparation trajectory (about two months) is described below. The outline of the planning is mentioned in the sections A-F, immediately followed by the description of the realisation. I include here a summarised part of the interview (A, see below) and the reflection on the two recorded lesson observations (B) as steps of the trajectory because they not only served to give me a first impression of the starting position of the teacher (see section 6.3.1), but also as an orientation for teacher C on the water quality lessons and the intended teacher role.

A. To give the teacher a first impression of the ideas and intentions of the teaching sequence I start a discussion of those issues with the teacher: e.g. the idea of a story line in the teaching sequence which resembles the practice of judging water quality, the importance of preparing a context for activities and evaluating them, and the importance of involving students by paying proper attention to their input to make such a story line explicit (aim 1).
B. In order to get a deeper understanding of what these ideas mean for his practice the teacher first reflects on his own practice directed by specific feedback: to what extent does the teacher already practice the characteristic features of interaction structures (although the term ‘interaction structure’ is not mentioned as such): setting the stage for activities and evaluating them and involving students by paying proper attention to their input to make such a story line explicit. We discuss videotapes of two of his lessons from this perspective. Teacher C should get a deeper understanding of what the ideas mean, and see the point of these ideas (aim 1).

Based on his reactions on the interview (A) and the feedback sessions (B), I concluded that teacher C understands and sees the point of introducing and evaluating activities with respect to their meaning and involving students in this process, if only to verify if they have understood as intended (see section 6.3.1).

C. In order to get a deeper understanding of these characteristic features of interaction structures and their function (to involve students by paying proper attention to their input) the teacher prepares a part of a lesson from this perspective. We discuss the planning of the lesson, e.g. its ‘story line’, and teacher C’s intentions beforehand. Afterwards we discuss the videotape of the lesson. Teacher C reflects on his practice directed by the questions ‘How did you introduce, evaluate and involve students?’ and ‘Were your intentions met and how do you know?’ (Aim 1).

Teacher C prepares in this way two lessons for two different classes in which he practices with making a storyline explicit for students and involving them properly. We reflect on each lesson afterwards whether it did go as intended, using the video recordings of the respective lesson and my observations.
The first lesson teacher C is very much in his instructional role. He does tend to explain the meaning of each experiment and activity (about the conservation of mass), but at the same time seems to forget to involve students. It strikes teacher C when watching the video that he is speaking so much. He also remarks that as a result he actually does not know whether the students have understood and followed the storyline as he intended. For the next lesson, teacher C wants to pay explicit attention to student input and to involving them in the storyline.

The second lesson teacher C prepares is about the molecular model. Although he has tried to create a storyline, the topic of the lesson turns out to be so very abstract and difficult for students that he needs to abandon the story line and to explain a lot. The intended idea of involving students is lost. Afterwards we discuss that the content involved should be accessible for students if you want to involve them properly and how this is the case in the Water Quality module.

D. To prepare the teacher for step E, practicing with the teaching sequence in the try-out class, the story line of water quality is talked through in general and each episode in more detail. Regarding each episode it is discussed how the context should be prepared for its main question, how the activities are expected to address the episode question and how the episode question should be evaluated. The expectations about student input are discussed and if necessary, for example when teacher C has doubts whether students will come up with certain input, underpinned by experiences with the second version (aim 2).

Although he had received the materials some time before our meetings, teacher C was by now very anxious to discuss finally the lessons about Water Quality in detail, to see what he could expect. We discussed the general planning of the lessons: how the episodes might be divided over the lessons, how to organise the experiments etc. We also discussed the successive episodes in detail. Teacher C shows to be very enthusiastic about the consistency of the teaching sequence.

E. The teacher practices with the teaching sequence in the try-out class that is two lessons ahead of the focus class. We discuss the planning of each lesson beforehand. After each lesson we discuss the questions ‘Did it go as intended and if not, why not?’ and ‘Should anything be adjusted for the focus class?’, using the teacher’s experiences and my observations (aim 2).

The choice for a try-out class appeared to be the main preparation for teacher C for what to expect. Teacher C tried very hard, and was sensitive for his intended teacher role. He explained in an evaluative discussion after the second lesson in the try-out class that he still found the following very difficult:

-to wait for students responses at the proper moments
-to ignore the urge to direct students to the ‘proper’ steps of the procedure
-to structurally prepare a context for the episode question and evaluate the episode question of an episode as intended.

Teacher C tended at moments to loose track of the storyline himself. He blamed this for a great part on the inflexibility of his role, which made him ill at ease, afraid to give answers away or say the wrong thing at the wrong moment (see also section 7.5.2 ‘Retrospect of the teacher’). Teacher C felt that the development of the procedure for Water Quality control made his role inflexible: he felt he had to make sure that the proper steps emerged at the proper time, which was sometimes unnatural in his experience.

The use of a try-out class also led, at some points, to small adjustments for the focus class. Teacher C had for example a difficult time in letting students come up with the first steps of the procedure in the first episode. The students did not seem to know what teacher C was asking for. They had already formulated these first two steps of the procedure on their group poster in the first activity, although without explicitly labelled as such. We discussed this and concluded that the first too steps were probably too obvious for students and because of that, the question ‘what are the first two steps’ was confusing for them. It seems much simpler and less artificial for teacher C to point at the different group posters and label their answers as the first two steps himself.

F. In order to be fully prepared, including the possible adjustments as a result of practicing with the try-out class, the teacher and I discuss his planning in detail before and after every lesson in the focus class: did it go as intended, and if not: are there implications for the next lessons? (aim 2).

How teacher C finally fulfilled his role in the focus class is evaluated in chapter 7, in section 7.2 and in particular in section 7.3. I want to emphasise that each time I conclude that the teacher did not fulfil his role as intended in whatever way, I do not mean this as a criticism of teacher C. I want to stress here that I take it first of all as a sign that either teacher C was not properly prepared or that the intended teacher role needs rethinking.

6.4 Methods of evaluation

In this section, I will discuss how the data were collected and analysed in the evaluation process and how the results are presented in chapter 7.
In chapter 7, the realisation of the design is evaluated from different perspectives. The methods of evaluation and the presentation of the results differ for each of these perspectives and are therefore discussed in separate sections. Section 6.4.1 discusses the methods of evaluation of the realised teaching-learning process from the perspective of the content-related progression. In section 6.4.2, the method of evaluation with respect to the realised interaction is discussed. Section 6.4.3 discusses the evaluation of the final test. Finally, in section 6.4.4 the methods of evaluation with respect to the students’ and teacher's appreciation of the Water Quality lessons in retrospect is discussed.

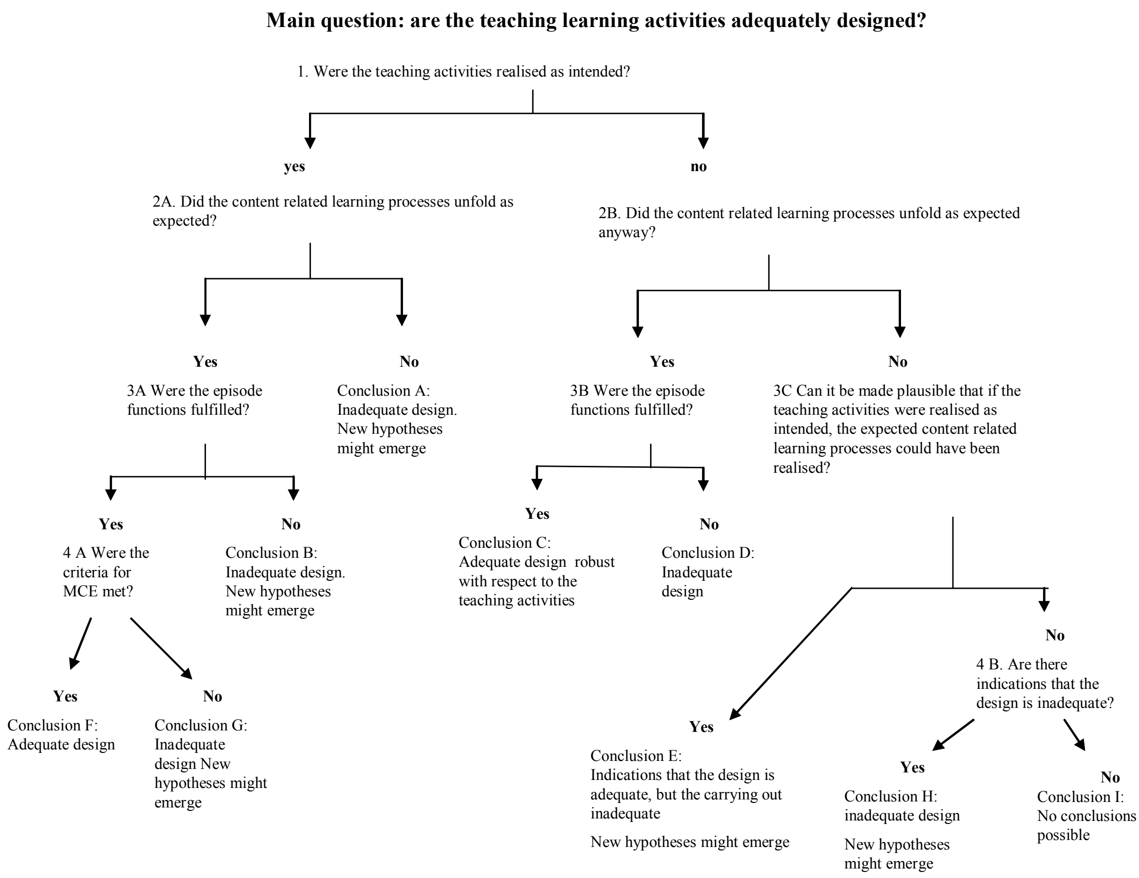
6.4.1 Evaluation of the instructional version of the authentic practice

In section 7.2, the design is mainly evaluated from the perspective of the content-related progression. The main question from this perspective is to what extent does the design of the teaching-learning process meet the following criteria of the characteristics of meaningful and, as a result, can it be seen as a case of meaningful chemistry education?

- Are students broadly motivated by the purposes of the authentic practice to adopt their role and get involved in its instructional version as intended?, (*context)* and:
- Do students experience at the intended moments that they need specific knowledge to take the next step in light of their purpose or at other moments that such specific knowledge asks for a refinement of a procedure step? (*need-to-know*).
- Do students feel that their input contributes to a common purpose? (*attention for student input*-c).

Figure 6.1 shows in broad outlines the decision scheme, which will be followed in the evaluation of the design in section 7.2. Although a bit rough and oversimplified, it shows the main questions that need to be answered with the available data, and the main types of conclusions that might follow from this.

Figure 6.1: The evaluation decision scheme with respect to the adequacy of the design



For each part of each episode, (the setting the stage-, addressing the episode question– and evaluating the episode question-wrapping up -part), a list of evaluative questions was formulated to start the evaluation of the realised teaching-learning process. These questions reflected the intended teacher activities and the expected student activities and responses as described in the scenario. The questions addressed both the content-related progression (section 7.2) and in some ways the interaction (section 7.3). They were used as a rough checklist for observation and for analysis of the data. To give an impression, a few of the evaluative questions formulated for the first episode are listed below.

Some evaluative questions of the first part of the first episode:

Setting the stage for the episode question

1. How and how enthusiastic does the teacher introduce the module. Does he use similar dilemmas as the exemplary ones? Does he involve the students by asking them for examples?
2. Do the students seem interested? Are they listening? Are they (and how) reacting to the teacher's introduction?
3. Do they come up with adequate exemplary cases themselves?

Answering the evaluative questions led to a first evaluation of what had happened, what seemed to have worked, what not and possible causes.
In the whole process of data interpretation triangulation took place on two levels. Firstly, at the level of information sources. Different information sources were used in this process: video recordings, observations, protocols of audiotapes and worksheets.
Secondly, triangulation took place at the level of reaching agreement about the analysis and interpretation of the data. Different science education researchers (see below) checked my argued interpretation of the teaching-learning process based on various information sources. The scenario, of course, served as a reference throughout the process.
Answering the evaluation questions gave a first overview of the realisation of the teacher role and the learning process of the students.
Next, I will present in detail how the data were processed in the evaluation process and how the evaluation results are presented in section 7.2.

The data were processed and presented as follows:

1. Every lesson was videotaped. The video recorder was placed at the back of the classroom facing the teacher. Expressions of students were therefore less visible. However, when students worked in their groups, it was possible to tape their expressions more.
   The videotapes of the first four lessons (the fifth lesson was not taped for technical reasons) were studied twice. First they were run through as it were, sequentially, mainly to bring back to mind the lessons as the analysis took place some weeks after the trial. This first view of the tapes provided for a quick overview and a general impression of the unfolding of the content-related progression.

The second time the videotapes were studied detailed notes were taken of how the teaching-learning process unfolded. These notes served as an additional source of information to the field notes that were taken when observing the lessons. The videos provided for more specific information on things one does not always fully grasp when observing the lesson as it unfolds. Such as the teacher's body language and his facial expression when for example evaluating student input, which students he addresses, how he handles the posters, walks around, the atmosphere in the class, the involvement of the students, their attitudes, tones of voices and so on.

2. The teacher had worn an audio recorder throughout every lesson. In a next step, these teacher-audiotapes of all five lessons were transcribed. The protocols of these tapes were not transcribed completely verbatim. Things such as stammering or 'eh' pauses were left out as it irritates the reader and were considered not to provide for relevant additional information for this evaluation. Also, it is impossible to translate spoken Dutch properly into spoken English.
Additional information that was considered relevant for the interpretation was indicated between brackets. The following example illustrates the implications of this process. When the teacher says: ' eheh let's see, what are, what are we going to do? In eh the next chapters we are trying to solve these', it is transcribed as:

T: 1. Let's see, what are we going to do?
2. In the next chapters we are trying to solve these [*points at the listed doubts]*

(The first two sentences of protocol 5A)

It is relevant information that the teacher points at the poster on which the doubts of the students are listed: he refers to the student input (their doubts) by doing this.
All the protocols of the teacher's audiotapes and of the audio-taped group discussions are numbered per episode in the sequence in which they are presented in section 7.2. Protocol 5A is for example the first protocol (A) that is presented in the evaluation of episode 5. The teacher is referred to as T, students are called S-1, S-2 etc.
As student's individual learning processes are not object of study, S-1 in one protocol fragment might very well be a different student than the student S-1 from another protocol fragment. And student S-1 whose written answer is cited, might very well be yet another student. Of course, *within* each fragment, S-1 is the same student and S-1 and S-2 are different students.

The protocols of the teacher's audiotape provided for a detailed overview of the content-related progression of the lessons. Apart from the evaluation of question 5B, episode 4, and some brief moments of class activity of little relevance (when students worked on the activities), they were completely and integrally used in the evaluation and the presentation of the evaluation.

In the part of 'addressing the episode question- part' of each episode students worked mostly on tasks in groups. The main information source on the content-related learning processes in these parts were the worksheets of 16 students of four randomly chosen groups. In addition, all posters of all groups of the first activity were analysed. It was thought that the worksheets of these 16 students would give an adequate overview of the type of answers students come up with because they were interpreted in combination with observations and the video recordings of these parts. These data provided for additional information on the type of answers students come up with and the type of questions they ask the teacher. Three of the four groups of which the worksheets were analysed were also audio taped. However, it turned out that the recordings generally did not provide for much additional information, and were therefore used only at some points where it added some relevant information Evaluative interviews were merely used to clarify and verify interpretations.

3. In a next step, he first version of the interpretation and analysis was discussed in detail with one other researcher. This researcher judged the evaluations of every episode, the type of data used, the interpretations of the data and conclusions on plausibility and trackability (Lijnse, 2003). Unclarities, differences of opinion ranging from interpretations to the type of data used were extensively discussed. This procedure continued until agreement was reached. The next version of the evaluation was discussed in the same way with two other researchers.

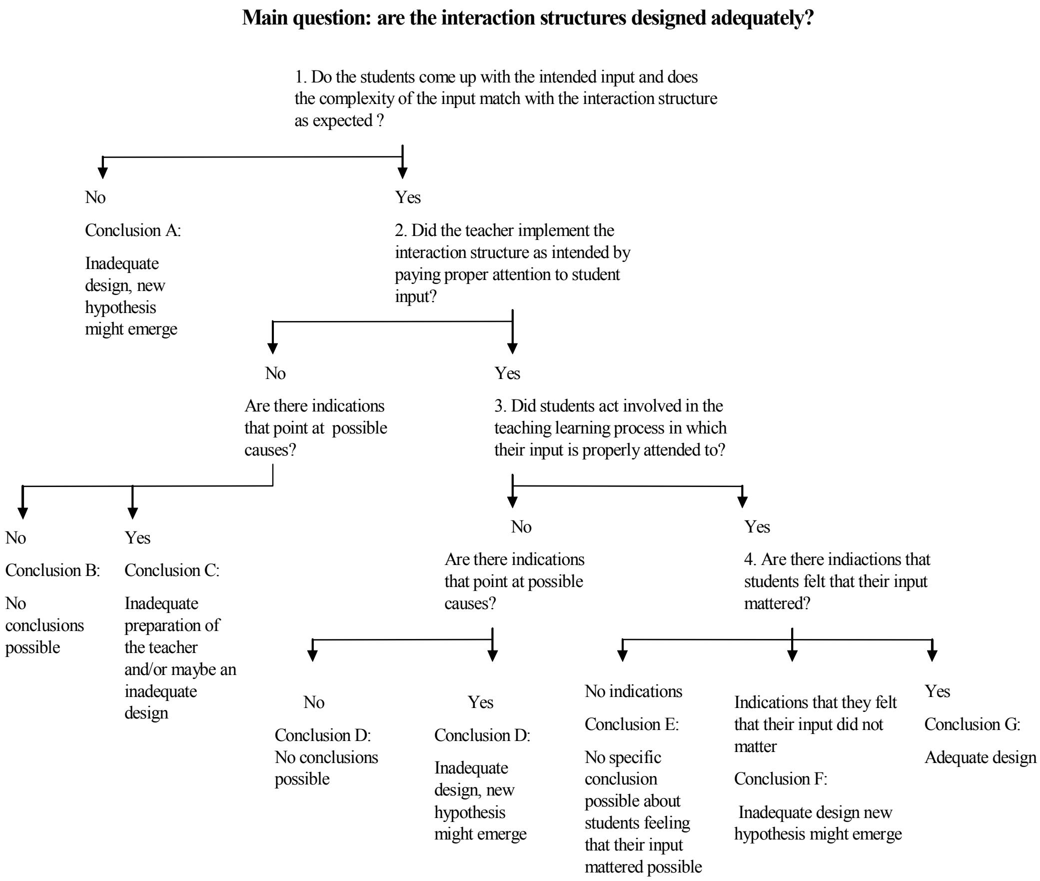
In the next section, the data processing and presentation of the evaluation of the realisation of the interaction between teacher and students is discussed.

6.4.2 Evaluation of the interaction in the teaching-learning process

In section 7.3, the design is mainly evaluated from the perspective of the interaction in the classroom. The main question that needs to be answered is:
Are the designed interaction structures implemented by the teacher as intended, do they lead to the intended interaction patterns in the classroom and do students as a result feel that their input matters as a driving force in the teaching-learning process (*attention for student input*-d).

Figure 6.2 shows in broad outlines the, somewhat rough and oversimplified decision scheme that was followed.

Figure 6.2 The evaluation decision scheme with respect to adequacy of the interaction structures



The data were processed as follows. Protocols of the teacher's audiotapes (see section 7.2) provided information on how the teacher involved students in setting the stage and evaluating the episode question. The transcription of the audiotapes provided information on whether the teacher was adequate in directing the intended interaction. The audiotapes also provided information on the tone of voices with which students and teacher spoke out: enthusiastic, bored, inviting, insecure and so on. In section 7.3.1 the interaction structures are evaluated for each episode. When protocols of section 7.2 are used in the evaluation, it was thought to be sufficient to refer to section 7.2. When videotapes and the field notes of the classroom observations served as additional information sources this will be indicated between brackets.

The first version of this evaluation was checked by and discussed with in first instance one other researcher and in second instance with two other researchers similarly as in the evaluation from the perspective of the content-related progression (see 6.4.2).

6.4.3 Evaluation of the final test

In section 7.4, the final test is evaluated. The final test served as an additional source of information to evaluate whether the individual students had achieved the intended learning goals. The evaluation of the content-related progression of the teaching-learning progress including the outcomes of the final, reflective, activity provided for primary information on the question to what extent the learning goals were achieved by students.

The final test questions were determined in the following procedure:

First, the learning goals were listed (see section 4.3) based on which a first version of the test was designed, including a model for judging the answers.

Next, face validity was achieved on two aspects. Firstly, some science education researchers evaluated the adequacy of the test as a research instrument. Secondly, two teachers, teacher C and one other teacher whose class also did the water quality lessons but was not involved in this research, evaluated the adequacy of the test from their perspective. All of them were given the test, the answer-model, the list of learning goals and the student materials.

They were asked to evaluate the test based on the following question: Do you think the test represents the learning goals sufficiently? And if not, why not? They were also asked to give other comments for example on the formulation of the questions, the layout and so on. This discussion led to a few adjustments in the type of questions and in the formulation of some questions and introductory texts.

The test was first carried out in the try-out class. Students had no questions when doing the test, at the time, and did not seem to experience problems in understanding the meaning of the questions. Therefore, no further adjustments were made.

In evaluative interviews 16 students were asked whether they had found anything that was not logical in the test. Not one student answered this positively.

The test results of the students were evaluated as follows. First all the students' answers were marked by the first researcher using the criteria for the correct answers. The answers were categorized as 'adequate', 'incomplete but sufficient' or 'wrong/ insufficient'. In section 7.4, the test questions and the criteria for an adequate answer are presented. These criteria were based on the assumption that the teaching-learning process was realised as intended and therefore were solely for research purposes (the marks the student got for their tests were of course adjusted to what actually had happened in the classroom). This way the answers could as much as possible be traced back to what actually happened in the classroom. For example, if all students' answers on a question lacked a specific argumentation, this could be traced back to the fact that the specific argumentation had not come to the fore in the respective episode (this happened e.g. with question 1A).

Next, a second science education researcher independently marked the students' answers a second time, using the written criteria for an adequate answer. The two evaluations were compared and discussed.

Some of the written criteria turned out to be unclear. After clarifying the respective criteria (this version is presented in appendix 2) some of the marks of the second researcher were adjusted. This was obvious for both researchers.
Table 6.1 shows the extent of agreement in percentages between the first and the second researcher as follows. A. presents the percentage of agreement after the criteria were clarified. B. presents the percentage of agreement after comparing and discussing the argumentation of the first researcher with the argumentation of the second researcher, only when agreement was quickly reached and beyond doubt. A few differences (C) remained.

Table 6.1 An overview of the extent of agreement between the two researchers about the test results

%agreement/question	1A	1B	2A	2B	2C	3A	3B
A.	78	85	85	85	85	74	89
B.	93	89	100	96	100	100	100
C.	7	11	-	4	-	-	-

A: % agreement after clarifying criteria
B: % agreement after discussion, without doubts
C: % of questions about which agreement is reached, but with more doubts

Finally, both researchers reached 100% agreement to validate the scores of the test results. Next, the second researcher evaluated the conclusions the first researcher had drawn based on these results. Both interpretations of the test results were to be discussed until agreement was reached.

In chapter 7, an overview of the relative scores of the students (n=27) is given. Finally, the interpretation of the overall scores is presented: to what extent they can be seen as indications for a learning effect and how they might be explained in view of the realised teaching-learning process. At some points, exemplary answers are presented as an illustration of how certain answers were judged.
In some cases the student answers will be interpreted in terms of the realised teaching process, e.g. if they reflect what the teacher had evaluated as ‘correct answers’ to the same kind of questions.

6.4.4 Retrospect from the students and teacher

In section 7.5, a retrospect by students and the teacher is presented. The main research question was: ‘How did the students appreciate the unit in general and what was their perception of the logic and functionality of the unfolding activities of the teaching-learning process afterwards?’. In order to get an idea of the students’ appreciation of the teaching-learning process students were given a post-trial questionnaire and 11 of the 16 students whose worksheets were analysed and who were audio taped, were interviewed in four groups two students and one group of three students. Not all of the 16 students were interviewed, simply because they were not available. However, the 11 students were considered representative for the class.

Twenty-two of the 27 students responded on the questionnaire. It consisted of two parts. The questions of the first part were about how the students had generally appreciated the teaching-learning process, the working methods, the topic, how they compared these lessons with their regular chemistry lessons and so on.
The second part was about how students had generally experienced the logic and functionality of the activities of the teaching-learning process, referring to how they in retrospect experienced and appreciated the didactical structure of the realised teaching-learning process.
The questionnaire consisted mostly of open questions. It was developed in first instance for the second round of trails and inspired by the questionnaire Kortland used (Kortland 2001, p.154).
This version of the questionnaire was evaluated by a number of researchers (face validity) and had been used before in the three classes involved in the trials with the second version of the design.
The questionnaire was adjusted to the third version of the design of the water quality unit and again evaluated by a number of researchers. This mainly meant adjustments in the second part of the questionnaire were students were asked about specific tasks. All the answers on the open questions were transcribed and categorized per question. Students were for example asked 'How do you feel about the water quality lessons if you compare them to your regular chemistry lessons?' Answers varied from 'I liked that we were more independent' (for this question, answers of this type were for example marked as 'category a'), to 'the activities were all very clear' (category b), to 'it was a bit unclear sometimes' (category c). An answer such as 'we could decide a lot ourselves' was for example marked as a 'category A' answer. The scores per category were counted and this provided information on the students' appreciation of the teaching-learning process in retrospect.

The evaluative interviews provided for more specific richer information on the students appreciation in retrospect. The evaluative interviews were structured around four main questions that were thought to provide for relevant information on the students perspectives on the realised teaching-learning process:

1. Looking back, what do you in general think of the water quality lessons
2. Do you think the water quality lessons differ from the regular chemistry lessons? In what respect?
3. Did you ever think during the lessons 'Why are we doing this?' When? When not?
4. Did you feel that your input/answers mattered, was used maybe? When? When not?

Every interview would proceed as follows: the main question was asked first and it depended on the answers how the sequence of the conversation went afterwards. Sometimes students were asked to explain themselves further; sometimes they felt eager to share their opinions themselves (which not always had something to do with the theme).

Information on the teacher's appreciation of the teaching-learning process and especially his role, looking back, was derived from an evaluative post-trial interview. The evaluation interview with the teacher was structured around three main questions:

1. What is your general impression, looking back, of the lessons?
2. How do you look back on your role as a teacher, especially compared to the role you usually fulfil in regular chemistry lessons?
3. Your aim was to improve your practice by participating in his project, how do you look back on that?

Next, in Chapter 7 the evaluation of the third version of the design is discussed in detail. The main question is 'What worked, what did not work and why?'

Chapter 7

Evaluation of the scenario

7.1 Introduction

The detailed argued expectations of every step of the designed teaching-learning process described in the scenario (see chapter 5) were empirically tested in a third trial. In chapter 6, the set-up of this trial was discussed together with the data collection, processing and analysis (in broad outline). In chapter 7, the third version of the scenario is evaluated. The main question is: what worked, what did not work and why?

The results of the evaluation are described from different perspectives, each contributing to the main question. These perspectives are addressed in the successive sections as follows. In section 7.2, the emphasis of the evaluation is on the content-related progression. The question is: to what extent does the design of the teaching-learning process meet the criteria of the characteristics of meaningful from the perspective of the content-related progression and, as a result, can the teaching-learning process be seen as a case of meaningful chemistry education?

In section 7.3, the evaluation is focused on the teacher role and the realised interaction structures. The main question that needs to be answered is: are the designed interaction structures implemented by the teacher as intended, do they lead to the intended interaction patterns in the classroom and do students as a result feel that their input matters as a driving force in the teaching-learning process?

Section 7.4 addresses how students scored on the final test. This evaluation serves as an additional indication of the learning effects of the realised teaching-learning process. The test results are traced back to the realised teaching-learning process as much as possible. Section 7.5 addresses students' (7.5.1) and the teacher's (7.5.2) appreciation of the water quality lessons in retrospect. The question is how, in retrospect, students appreciated the lessons with respect to such themes as general appreciation, logic and usefulness of the successive activities (referring to the didactical structure), and the feelings of the students in being involved and having their input matter. The teacher was asked how, in retrospect, he appreciated the lessons with respect to their general design, his role and his original aims and expectations for participating.

In section 7.6 the main conclusions on the third version of the design are summarised.

7.2 Evaluation of the instructional version of the authentic practice

In this section, I will present the evaluation of the realised teaching-learning process with emphasis on the content-related progression. In section 7.2, the design is mainly evaluated from the perspective of the content-related progression. The main question from this perspective is: to what extent does the design of the teaching-learning process meet the following criteria of the characteristics of meaningful and, as a result, can it be seen as a case of meaningful chemistry education?

- Were students broadly motivated by the purposes of the authentic practice to adopt their role and get involved in its instructional version as intended?, (*context)* and:
- Did students experience at the intended moments that they needed specific knowledge to take the next step in light of their purpose or at other moments that such specific knowledge asked for a refinement of a procedure step? (*need-to-know*).
- Did students feel that their input contributed to a common purpose? (*attention for student input*-c).

Section 7.2 is structured as follows. Every episode is evaluated in a separate section: 7.2.1 (episode 1) up to 7.2.8 (episode 8). The headings of each of these sections refer to the episode question. Which phase of the teaching process the episode belongs to is indicated between brackets. Section 7.2.1 has, for example, the following title: “Episode 1: What does water quality judgment involve? (Phase 1)”.
In a subsection, the episode functions are listed. For some episodes, it turned out to be necessary to adjust the episodes functions in view of the unfolding of the previous episode(s) at the time the lessons were given. In these cases, the adjustments are explained in the subsection “Reviewing the episode functions’. Next, the main conclusions on this episode are briefly presented in the subsection ‘main conclusions’ after which the evaluation of the realised teaching-learning process is presented in the subsection ‘Evaluation of the episodic teaching-learning process’. Every section ends with the subsection ‘Conclusions’. Table 7.1 shows how the eight episodes were divided over the six lessons.
All the episodes were evaluated in detail. In this chapter, I will present the evaluation of a selection of the episodes in such detail: episodes 1, 2, 5 and 8. The selection of these episodes is based on the type of conclusions drawn from their evaluation. They turned out to be the most crucial in the sense that they had the most impact on the research question of how to design an instructional version of the authentic practice of judging water quality. The detailed evaluation results of all episodes is available in an internal report (Westbroek *et al*., 2005)
The subsections that present the detailed evaluations (7.2.1, 7.2.2, 7.2.5 and 7.2.8) are outlined as follows. First, the function of the respective episode is described. Next, the evaluative questions are answered for each part of the episode: ‘setting the stage for the episode question’, ‘addressing the episode question’ and ‘evaluating the episode question’. The key question is: To what extent are the expectations about the unfolding of each episode met? I will present in detail how the data were interpreted and used to evaluate the realised teaching-learning process. Third, based on this, conclusions that are more general are drawn as to what extent the expected episode functions were fulfilled.
Of the subsections 7.2.3, 7.2.4, 7.2.6 and 7.2.7, I will present the full conclusion but a summary of the evaluation.

Table 7.1 An overview of the six lessons (as they unfolded)

Lesson (min)	Phase	Episode	Activities
1 (50)	1 & 2	1 and 2	1,2,3
2 (50)	3	3	4
3 (75)	3	4 (evaluation), 4 and 5	5 –10
4 (50)	3	6	11-14
5 (50)	3 & start 4	7 and 8 (stage setting)	15, 16
At home	4	8 (addressing the episode question)	16
6 (75)	4	Evaluation 8	Test

7.2.1 Episode 1: What does water quality judgement involve? (phase 1)

Episode functions

In this episode, the following functions were to be fulfilled:

1A Students are to feel broadly motivated to get involved in the instructional versionof the practice about judging water quality.

1B They are to get a clear view of

1Ba the purpose of the instructional version of the practice

1Bb in broad outlines how they are going to achieve this

1Bc their role in it

Main conclusions

Although the episode unfolded in several ways as expected, not all its functions were fulfilled. This was partly due to the design of the activities. They need rethinking at some points, as it might be questioned whether the activities were putting students (and the teacher) on the right track at specific points (especially activity 2). In addition, the teacher did not address the content-related purpose of the episode (why are we doing this?) when setting the stage for the episodes. He also did not do this for the activities. As a result, the students were not provided with a clear sense of purpose. However, students gradually saw the purpose and became generally motivated by the instructional version of the practice. Also, the estimation of the students' intuitive knowledge about the procedure proved to be correct and they generally learned about the first two steps of the procedure.

Evaluation of the episodic teaching-learning process

Setting the stage for the episode question

Although the teacher made a lot of effort to involve students by paying attention to their input, he did not address the purpose of the episode as intended. Because of this, students did not grasp the purpose of the activities in this episode beforehand. However, it became clearer to them with the unfolding of the activities (afterwards).

Unfolding of the stage setting
[Observation, video, protocol analysis]
The lesson started with some procedural remarks and a brief introduction and explanation of the background of the water quality module. The teacher pointed out the video camera and audio tape recorders, which he had already announced in a previous lesson. After this, the teacher began with the actual introduction.
The teacher's introduction was short (see below) and he acted a little nervous, searching for words. Although he practiced it in the try-out class, the teacher did not set the stage as was intended. In his introduction, he emphasised the authenticity of the descriptions and the accompanying water samples (protocol 1A, lines 6, 7, 8, 11), but he did not direct the attention of the students sufficiently to the cases and in what sense they were exemplary for the practice of judging water quality: situation descriptions in which the water quality is determined in order to decide whether it is 'clean enough' for its function. In lines 2, 12 he referred to the driving question and to the procedure respectively, thereby showing the different 'genuine water samples' to the class, and in lines 10,13 and 15 he already referred to the first step of the procedure: water function determines quality criteria. However, he did this in an indirect manner, without directing attention to the cases: *this* is what the lessons are about. Furthermore, by asking students for more exemplary *water types* (line 16) he directed their attention further towards water types instead of the water quality cases as a result of which students did not come up with 'similar exemplary cases' but with 'different water types' like ditch water and tap water (lines 17 to 21). The teacher (line 22), in turn, reinforced this.

Protocol 1A

T [..]

1. Ok, what will you be doing.
2. The main question is: 'is the water clean enough?' and you will make eh a poster about that later on.
3. What do we mean by this?
4. See, you have all types of different water.
5. This water is for example brewing water [*holds up the brewing water sample*].
6. We got it from an official brewery.
7. All the water here is real water.
8. So, from a brewery.
9. Brewing water.
10. You can imagine that this will meet different criteria that sea-water [*Picks up another water sample*].
11. This water is from Wijk aan Zee.
12. How can you determine whether it is clean enough?
13. This example [*holds up the brewing water sample*] is probably cleaner than this one [*holds up the sea water sample*].
14. Right?
15. This one [*sea water*] probably does not need to meet such strict criteria.
16. Eh eh can somebody think of other examples of water types?

S-1:
17. Tap water!
S-2:
18. Mineral water.
T:
19. Tap water, very good, mineral water.
S-3:
20. Dune water.
Different students:
21. Ditch water; sewage water.
T:
22. Ditch water, very good. Look, I have got tap water right here [*holds up tap water*].
23. There are a lot more examples, now what you are going to do today is first you will form groups of four eh you will work in groups from now on.
24. You will make a poster.
25. Let's see if I got it here. You will get a poster like this and answer the following questions [*goes over all the questions*].

From this point the teacher continued with explaining the organization of the main activities.
The teacher could not have been sure whether the students got the picture of the authentic practice they were going to learn about, because the introduction was not about this practice. Moreover, there were several indications that the students did not have a clear sense of this purpose (see below). However, they did feel involved by the teacher's introduction; they were motivated to start [observation]. Students started discussions on which case they wanted to address. All but one group immediately started to read and discuss the case description they were handed out.

Addressing the episode question

Although students did come up with the intended input in activity 1, the unfolding of this part of the episode indicates that the students started without a clear sense of purpose, probably because of the way the teacher set the stage. Activity 2, which introduced the procedure poster, unfolded a bit chaotic. This indicates that it was not a logical next step for students, and also not for the teacher, partly caused by what happened previously, but mainly by the design of this activity.

Unfolding activity 1
[Observation, video, protocol analysis]
Students received the necessary materials (posters, crayons etc). In general, the students were involved in discussing the questions on the poster. They read the case descriptions and started to discuss the questions. Some groups opened the water sample bottles. Students frequently asked the teacher questions, which could be interpreted as a sign that they were involved.
There were a lot of 'What are we supposed to do?' questions, which indicated that students lacked a sense of purpose when they started the activity. Also, when the groups worked on the case descriptions and the questions, there were still indications that they did not fully grasp the exemplary case descriptions and the purpose of the activity [observations, video]. All the group answers are presented below, in tables 7.2 A and B.

The group which addressed the case of ‘the beach of Wijk aan Zee’, for example, asked what the function of the seawater was (all groups have a water sample) [observations]. The case was about the question whether this beach deserves a blue flag, which means that the water is found to be clean enough to swim in by an independent lab. The water has to meet European quality criteria. All this information was available to the students in the case description. Obviously, this group did not grasp what the water sample stood for. Moreover, the group not only asked for the function of the seawater, but also wrote down a number of possible sea water functions on their poster which had nothing to do with ‘blue flags’: shipping, salt winning, accommodating sea animals. A similar thing happened with the other groups who addressed cases in which the water also had different functions than the function central to the case: the case of the water from a protected area that has an ecological function and surface water that is used as swimming water.

Also, all groups answered question 4 positively: ‘Do you consider it interesting to find out in chemistry class how people do this? Yes/no, because…’. Four of the groups (1,3,4,6, see tables 7.2.A and B), formulated answers which could be interpreted as referring to a more fundamental motive for getting involved in the learning process, when they discussed and wrote down their opinions on the question whether they think it is important to learn about how this is done in reality. These groups all filled in typical general answers like: it is very important [*to learn what is involved*], because water is essential to life, and you need clean water. The students of these groups did not formulate more specific motives, referring directly to the cases and a willingness to learn *how* they are solved, in other words: the procedure to be followed. Groups 2, 5, 7 and 8 formulated answers that did refer to the specific cases [analysis of posters].
Also, the students’ opinions sometimes differed more than the answers on the poster suggested (see e.g. protocol 1B)

Protocol 1B

S-1 [reads aloud]:
1. ‘Do you consider it important to learn about this [*refers to the previous question; how they check whether the water is clean enough*] in the coming lessons?

S-2: 2. No
S-3: 3. I do
S-2 4. Well, I just do not think it important
S-3: 5. Yes it is! Water is the source of life!
S-2 6. Ok just fill in something
S-4: 7. I don’t think it important at all
S-1: 8. Excuse me, do you know how important water is? [*indignant]* If there was no water we would be dead now!

In at least three groups, students had differing opinions and similar discussions [observations].
Apart from this, the answers on the posters were mostly as expected (see tables 7.2 A and B for an overview). The students did not have difficulty with filling in the water

function once the case was clear (apart from the ‘blue flag’ group) [observations]. The answers to the criteria varied, as was expected, from: ‘it must be clear, taste good’ (drink water) to ‘it should not contain substances which make you ill’ (swimming water). Six groups referred to substances that the water should or should not contain; four groups referred to bacteria or algae the water should not contain. To the question: “How do you think they determine ..?”, all the groups gave answers which referred to testing. Sometimes groups gave answers, for example like: pH tests, salt-concentration, chlorine test, skin-test (group 5), temperature tests. Three groups only mentioned testing as a general step. One group added that this should be done regularly. Five groups mentioned taking samples as a step. One group (the ‘blue flag’ group) mentioned that more than one sample should be taken from different locations. The group that worked on the case of the two water net system suggested that nitrite, chloride, pH and E-coli bacteria should be tested. They obviously got their answer from the student materials [analysis of the posters].

Unfolding activity 2
[Observations, video, protocol analysis]
By the time the different groups finished filling in the posters, which were placed on the wall, the teacher told the students to look at each other’s posters. He did not give students specific directions such as see if they could find differences and similarities between the posters:

Protocol 1C

> T: I can see that everybody is sort of finished by now. I will give you five minutes to look at each other’s posters and I will ask you some questions about it afterwards.

While the students walked around and looked at each other’s posters, the teacher remained behind his desk. The students showed that they were not sure what they were supposed to look at. The atmosphere was somewhat restless and most students started to discuss other things than what was on the posters. This was as an indication that activity 2 was not meaningful to them in the sense that it did not serve a clear purpose [observations]. In retrospect, even if the teacher had asked students explicitly to look for similarities and differences, it could be doubted whether they would have linked this activity functionally to the purpose of learning about the procedure, which was the intended purpose of activity 2. The design needed rethinking at this point. The way the actual introduction unfolded, confirmed this. The teacher got too little support from the design.

Table 7.2 A The poster of activity 1: answers of groups 1-4

Group	**Question**			
	1. What is the function of the water in this case?	2. What do you think would be water quality criteria in this case?	3. How, do you think, do they check whether the water satisfies the criteria (what steps are involved)	4 .Do you consider it interesting to find out in chemistry class about how people do this? Yes/no, because…
1. Surface water in a nature reserve	Life for plants and animals	No chemicals; clear natural water; neutral pH	pH test; make a chromatogram	Yes: life needs lots of water. Water is the source of life.
2. Swimming pool water	For swimming	Correct concentration chlorine; no bacteria; must not be turbid; the right temperature; not too much calcium	Test in a lab & by the swimming pool itself (chlorine). Take a sample; measure the temperature	We think it is important, because you want to know in what kind of water you are swimming.
3. Aquarium water	This water is used in the fresh aquaria	It must be fresh; right temperature; right salt concentration; quality of the water must not vary/change	Measure the salt concentration; measure the temperature; determine whether the temp. and the salt conc. is constant	Yes, we want to know what is in the water and for that you need chemistry. We think it is very important.
4. Drinking water	For drinking For in the kitchen For cleaning	No harmful bacteria; neutral pH; E-coli bacteria, chloride, nitrite	Take water samples; measure pH; spectrophotometer; potentiometer	Yes, because water is of vital importance.

Table 7.2 B The poster of activity 1: answers of groups 5-8

Group	**Question**			
	1. What is the function of the water in this case?	2. What do you think would be water quality criteria in this case?	3. How, do you think, do they check whether the water satisfies the criteria (what steps are involved)	4 .Do you consider it interesting to find out in chemistry class about how people do this? Yes/no, because…
5. Swimming water (surface water)	For swimming, water sports	No substances that cause complaints like the blue algae; no toxic substances; safe for health	Take samples; test the samples in a lab (e.g. microscope, indicator); see if skin reacts to the water	Yes, it is also important for us. A lot of people get sick.
6. Sewer water	For dumping	It must not be polluted, because it is dumped	Take samples and test them	Yes, because we are for a better environment.
7. Brewing water	It is an ingredient for beer	As pure as possible; drinkable	Let a lab test the water regularly to see if the quality is still ok	Yes, because people drink it.
8. Swimming water (sea water)	For shipping, recreation, for animals to live in, salt winning	No toxic substances; no harmful bacteria	Take water samples at different spots in the sea and let a lab test the samples	Yes, people can get sick if they swim in polluted water

Evaluation of the episode question and wrapping up

The unfolding of the evaluation part and specifically the introduction of the procedure poster indicated that the students did not experience this step as useful. There were strong indications that students did not consider learning about the procedure (to solve such cases) to be the purpose of the module. Besides that the teacher did not direct students to this purpose, the design needs rethinking. Maybe it was not recognised well enough how unusual it is for students and teacher to consider a procedure as an object to learn about. When students have to learn a procedure, they usually are supposed to read it in their books and be able to reproduce it, instead of developing it themselves as intended in this episode (see also the evaluation of test question 3 in section 7.4). This might have interfered with the intentions of the Water Quality design.

Unfolding evaluation-part
[Observations, video, protocol analysis]
The teacher started the evaluation with briefly 'going over' all the posters and the types of answers. He read aloud some answers, sometimes in an appreciative and

sometimes in a doubtful manner. The question of activity 2 "look for similarities and differences" was discussed now, but floated in the air and did not link the students' posters of the cases (activity 1) functionally to an introduction of the procedure poster (protocol 1D). As a result, the evaluation unfolded a bit disorderly.

Protocol 1D

T: 1. Ok, eh the posters, what can we make of the posters?
2. You can see that water can be used for various purposes.
3. So, if you look at the question, eh, 'we think that the water should meet the following criteria',
4. What stands out? What is the difference between all those things, between all these posters?
5. Is there a difference? Is there a similarity or a difference?

Several students:
6. A difference
S-1: 7. They're all different water
S-2: 8. Yeah
T: 9. And what is the reason for that you think?
S-1: 10. Because they have used all different types of water
T: 11. All different types of water
S-3: 12. With different purposes!
T: 13. Different purposes, yes
Different students:
14. Yes yes
T: 15. Very good, different purposes.

The teacher turned to the next question on the posters: how do you think they find out whether the water meets its quality criteria?:

Protocol 1E

T: 1. And next, if you look at…eh…how do you think they will check whether the water meets the criteria?
S-4: 2. Testing
S-5: 3. They test the water
T: 4. They test the water.
5. I can see that many groups have written down real extensive answers, like 'chromatography'.
6. That's very good.

The students' answers to the last question, "Do you think it is important to find out how they do this?", and the teacher's appreciation of their answers (Protocol 1F, lines 11, 12, 15) showed that all of them lacked a sense of purpose: the answers did not reflect the purpose of learning about the procedure (see "unfolding activity 1"). The teacher also did not use the students' answers to invite them to take up their role, which would have involved them more (probably) and would have focused everyone's attention on the purpose (connected to their role) of finding out about an adequate procedure. In retrospect, this is not surprising, as the design (see chapter 5) did not provide the teacher with guidelines for how to achieve this. Also, the teacher was not explicitly prepared for inviting students properly to play their part at the time, as this was not considered a very important issue at the time. Besides this, establishing the students' purpose in the instructional version of the practice proved to be more complex than was expected. I will elaborate more on this issue in chapter 8.

Protocol 1F

T: 1. Ok, then the last question: do you think it is important to learn how this is done? Why?
S-5: 2. Sure
T: 3. Why is it important? …[hurriedly:] or not?
4. Who does not think it is important at all?
5. Ah you know what I want to hear!
6. I think there must be some of you who do not think this important!
7. Why is it important?
8. What have you written down S-6?
S-6: 9. Well, water is just necessary for life.
10. Without it you just cannot survive.
T: 11. Very good.
12. Water is of great importance.
13. What else? Who?
S-7: 14. Your body exists for a great deal of water.
T: 15. Very good.

The actual introduction of the procedure-poster finally did unfold a bit disorderly and in a by the way manner. First, the teacher just briefly mentioned 'what we are going to do in the next few lessons' (developing a procedure) and referred thereby already to the drinking water example of the case of the two water nets. To come to the first steps of the procedure, he basically asked the class again how they thought people find out whether the water meets its quality criteria (line 2).

Protocol 1G

T: 1. Ok, what are we going to do in the coming lessons?
2. We are going to develop a kind of procedure to, let's say, decide if we can drink this [*holds up the water sample of the two water-nets case*] water.
S-7 3. You should test it.
S-8 4. Send it to a lab
S-9 5. Test persons!
S-10 6. Let's have S-11 drink it!
T: 7. Ok, ok, lets focus on what we are going to do in the next few lessons.
8. We are going to develop a sort of procedure by which we can establish if <u>this</u> water is drinkable.

The further unfolding of the introduction of the procedure poster showed that the design did not help the teacher to do this as a logical next step. Of course the previous events, which resulted in a lack of sense of purpose among students (it is doubtful whether they see learning about the procedure as a goal), made it even more difficult for the teacher. The teacher apparently decided to introduce the poster with the open question "what steps can you think of?" (Protocol 1H, line 5), which worked well in the sense that one by one the students came up with the intended steps (lines 7, 9, 14, 16, 18), directed by the teacher's questions (lines 5,6,8,13,15). The way students called out ‘yes, yes’ in line 20 to the teacher’s question ‘so you say test if it meets the criteria?’ indicated that students felt that their input is recognised (and matters).

In retrospect, although the teacher did point to the posters of the students on which they formulated the steps (line 6), it might have been easier for him and more logical for the students to have started with summarising the answers on the posters, as they were already there (instead of asking students to mention them again)

Protocol 1H

T: 1. Let’s see, you all wrote down ‘test the water in a lab’ [*reads the answers aloud*], very good, very good.
2. What else did I see [*reads aloud*] ‘measure acidity and the temperature of the water, ok, everybody is to have something like ‘test the water’.
3. And [*reads aloud*] ‘taking samples’, I see, I do not know about that.
4. Ok let’s start with the procedure.
5. What steps can you think of?
6. Just look at the questions you answered here [*refers to the posters*]?
S-1: 7. What sort of water
T: [*repeats*]
8. What sort of water’, very good, but can you be a bit more specific?
S-1: 9. What the water is used for
T: 10. What is the water used for? [*T. fills in for the students, writes this down on the procedure poster*]
Different students:
11. Should we write this down?
T: 12. Ok, ‘what is the water used for’,
13. And if you look at the next question you answered [*What do you think the water quality criteria would be in this case?*], what does that [*the water use*] determine?
S-2: 14. The quality criteria
T: 15. Very good. The criteria. The water use determines the quality criteria. [*Writes this down on the procedure poster*]. Ok, what next? What is the next step?
Different students together:
16. Test! You should test the water
T: 17. Test the water, very good [*writes this down*]
S-3: 18. To see if the water meets the criteria
T : 19. The criteria, very good. So you say ‘test to see if it meets the criteria’?
Different students:
20. Yes, yes

The teacher wrote this down, and read aloud ‘If it meets the criteria, is that ok?
The teacher wrapped up the episode as follows:

Protocol 1I

> T: Ok so far you have come up with .. [*reads steps aloud*]., agree? You can imagine that there are several more steps. In the next lessons we will add more steps to this poster, until we know how to determine whether this water meets the criteria.

The teacher was focused on arriving at ‘the right procedure’, instead of using student input. He, for example, did not include taking samples on the procedure poster, although five of the eight groups mentioned this as a step. In the evaluative interview, the teacher said he felt a kind of pressure to arrive at the proper procedure steps with the students. He did not know how to do this in a natural way and felt he needed to direct the evaluation and wrapping up too much.

Conclusions

Students showed to be generally motivated to get involved in the activities. They could easily answer the questions on the posters, and after a somewhat slow start (the introduction) they were involved in animated discussions and they were generally involved in the class discussions (besides some particular moments).
Based on the evaluation results of episode 1 it can be concluded that students understood that to judge water quality, one needs to 1. Determine the water function first as this determines the water quality criteria 2. Test the water.
However, the functions of episode 1 were not fulfilled as intended. Although students became broadly motivated to get involved in the instructional version of the practice, a certain sense of purpose grew in the course of the events, afterwards. This was due to, as argued above:

1. the design of activity 2 and the way of introducing the procedure-poster
2. the teacher did not functionally introduce the episode and did not explicitly invite students to play their part in the instructional version of the practice.

7.2.2 Episode 2: Does the water quality of this water sample satisfy the criteria for drink water? (phase 2)

Episode function

Students are to experience that their intuitive notions are not sufficient. (When asked: Does the water quality of this water sample satisfy the criteria for drink water?) They know that the water should be tested, but lack specific issue knowledge, which might be formulated as the following knowledge need: ‘What does the water sample contain?’ (function 2)

Main conclusions

The design was inadequate in achieving that students experience the intended knowledge need. However, the results of the evaluation showed that the intended knowledge need very probably could have been induced among students. The didactical structure of the design does not need to be adjusted.
Similar to episode 1, the teacher did not functionally introduce this episode. Furthermore, the teacher introduced the second episode immediately after episode 1 was finished. Because this unfolded so smoothly, students could not experience the episodic character of the story line.

Evaluation of the episodic teaching-learning process

Setting the stage for the episode question

The unfolding of this part showed that the teacher did not functionally introduce the episode. It can therefore be doubted, also in view of the unfolding of the previous episode, whether the students grasped the purpose of the activities in this episode beforehand as functional for learning about the procedure.

Unfolding of the stage setting
[Observation, video, protocol analysis]
The teacher turned smoothly to this episode immediately after wrapping up the first. In his introduction of this episode, the teacher introduced the case of the two water nets as an exemplary case in which it is important to check whether the water meets the quality criteria, like in the case of the rainbow fish (lines 3-4).

Protocol 2A

T (*immediately after wrapping up the previous episode*):
1. This is the start of a procedure to check if the water meets the quality criteria [*refers to the first steps on the procedure poster*].
2. You know, if you know the water function then you can establish the quality criteria.
3. Take for example the case of the aquarium. Rainbow fish need fresh water. It would be disastrous if the water contained too much salt, for example. The fish will die.
4. Some of you addressed for example the case of the two water nets.
5. There are residential areas, which have a two water net system.

S-1: 6. Yes, I know one

T: 7. Those areas have a water net for drinking water and one for what they call household water.
8. What do you think they use that [*household water*] for?

[*Different students call out, T repeats their answers*]
9. Water plants, flush the toilets, wash the car

S-2: 10. That area nearby over there [*points in a direction*], has it too!

T: 11. Nearby here?

Different students:
12. Yes, that eco-area

T: 13. Ok, next, you will all get a water sample of one of those two water types [*drinking water or household water*]. And the water has not been checked yet. It is from such area.

S-3: 14. Is it tap water?

T: 15. If you look at the activity 1.2, you get questions like 'do you dare to drink this?'.
16. Just see, with your group, if you can determine whether you can drink this water. Just think about that for a minute.

[*At this point, the water samples are handed out; the case description is in the student material.*]

The teacher involved students by checking their pre-knowledge about household water (line 5). Students were paying attention, and called out their ideas when asked. In that sense they acted involved and motivated [observations, video]. However, the teacher did not set the stage for the episode question as was intended this way. The episode question, 'Do you think this water (the genuine water samples of the drinking water net of this two water net system) meets the drinking water quality criteria?' was supposed to be *functional* in view of learning about the procedure. In order to bring forward this functionality clearly it is insufficient to announce simply that what comes next is the start of a procedure to check whether the water meets the quality criteria (line1). It is therefore doubtful if students experienced the functionality as was intended.

Addressing the episode question

The unfolding of activity 3 showed that students did not yet experience a knowledge need. Instead, they generally showed to be rather confident about what the next step should be: test the water by comparing the water samples with ‘good water’. They did not ask themselves at this point if they had enough specific knowledge to actually do this. So the *need-to-know* was not explicitly addressed.

Unfolding of activity 3
[Observations, video, protocols of two discussions of two groups]
The students were asked the following questions in the student materials:

Two water nets

These water samples are from a residential area with two water nets. One water net delivers water of drinking water quality. The other water net delivers water of lower quality. This lower-quality water is cheaper. Residents can use the lower-quality water to water their gardens, wash their cars or flush the toilet. In short, water uses which do not hold the risk that people will drink this water. In cases like this, the quality of the drinking water net is monitored extra carefully. It has happened that the two water nets were mixed up and people drank the lower quality water.

A. *Open the sample bottles. How does the water smell? How does it look? Maybe you can look at the water through a microscope. Do you think this water meets drinking water quality criteria? And how sure are you about that?*
B. *If you are not sure, want kind of information do you need?*

All groups (except the group who addressed this case already) read and discussed the description of the case of the two water nets. They all opened the water samples, and some students actually tasted the water. This caused some noise in the class. Following the suggestions in the textbook, students smelled the water and checked its clarity. Two groups asked for a microscope (also a suggestion in the textbook) and were very disappointed that there were none available. This indicates that the students were generally involved in the activity.

Discussions emerged about the water being drinkable or not and merely focused on smell, taste and how the water looked, not surprisingly, as this was what the textbook suggested they should test. The protocols of two group discussions typically unfolded, for example, as follows:

Protocol 2B, first group

S-1: 1. How does the water smell? I can’t smell it; someone else should try
S-2: 2. I don’t think that this water meets the criteria. I will cross off…
S-3: 3. Yes, but I smell the *bottle*. It’s just the bottle you smell. It’s such an industrial bottle.

Protocol 2C, second group

S-1:	1.	How does it smell?
S-3:	2.	Ieee [*as in disgust*], it smells of a factory. No the *bottle* smells, not the water!
S-4:	3.	You can smell the bottle, its true
S-1:	4.	Yes its true
S-2:	5.	How does it smell? Like plastic
S-3:	6.	I will write that down!
S-2:	7.	What does it look like?
S-3:	8.	Well..
S-1:	9.	Clear
S-4:	10.	It looks quite normal actually
S-2:	11.	Ok clear!
S-3:	12.	I don't think it meets the criteria
S-1:	13.	I am not sure, because at first sight....
S-2:	14.	Does not meet the criteria?
S-3:	15.	No I think not
S-1:	16.	I don't think so either
S-2:	17.	Why not?
S-3:	18.	Because, look there is something in here (*sees something floating*)
S-1:	19.	Yeah, there is something in there
S-4:	20.	Excuse me; in the tap water you could drink there were all kinds of stuff in there too!
	21.	So it does not say anything at all.
	22.	That just happened *[the stuff getting in*] when they put the water in the bottle it must have gotten in some time.
	23.	It can't just have dropped from the air in the bottle or something
S-2:	24.	I think it meets the criteria!
T:	25.	[*Walks by*] would you drink it?
S-2:	26	No!
S-4:	27.	No
S-1:	28.	I would
S-3:	29.	Me too, I'm really thirsty, ha ha
S-1:	30.	I am not sure

Of this second group, the students doubted very much the way the water smelled and later on the fact that 'something was floating in the water'. Based on this they concluded that the water definitely did not meet the criteria. Although S-2 said at one point that she thought that the water met the criteria (line 24), she would not drink it (line 26). They tried to compare its smell and clarity with tap water. When S-4 said 'it looks quite normal' (line 10) she probably referred to tap water.

The worksheets of four groups (16 students) were analysed. These worksheets (not all students of the same group wrote down the same answers) showed a variety of opinions, but most of them (9) were not sure whether the water met the criteria or not. Seven of those based their decision on the taste, smell and clarity, the other two referred to the argument that you do not know what might be in the water. The answers to question B varied from 'smell and taste' (1x) to 'pH ' (7x), microscope (2x), compare it with tap water (2x), make a chromatogram (1x), test it (1x). One student answered: "What the water contains". None of the students who were 'pretty sure' answered question B.

These answers indicate that the students did not experience the intended knowledge need at this point. They answered the question properly, but were not put in the position that they really questioned whether they can take the next step they suggested (test the water by comparing it etc.). As a result, they did not think further about the specific knowledge they might need in order to be able to take this next step. The unfolding of the evaluation confirmed this interpretation.

Evaluation of the episode question

The teacher tried to evaluate the episode as intended, but this was hindered by the fact that students did not yet experience a knowledge need.

Teacher C collected the opinions of the students on question A. Only the answers that the water is probably not drinkable came to the fore. Students showed to have no doubts by now. Not surprisingly, the students used arguments at the level of 'smell, clarity and taste' (protocol 2D).

Protocol 2D

T:	1. Ok [*reads question A aloud, he does not refer to the case*], who thinks that this water meets the quality criteria for drinking water?'
S-1:	2. No!
T:	3. Ok, and who thinks it does not?
[different students call out]:	
	4. No!
S-2:	5. It tastes bad
T [repeats]:	6. It tastes bad. Ok, are you sure about this or not? Why?
S-3:	7. It tastes really bad
S-4:	8. There's something floating in it
T [repeats]:	9. Something floats in it
S-5:	10. I think it smells different and it looks different
T:	11. It smells and looks different
S-6:	12. It tastes bad

The class-evaluation of the question: "What information do you need?", brought up a variety of suggestions for how the water might have been compared to "good water" (protocol 2E, line 5, 9, 11, 13): using a microscope; using a microscope to compare the water with tap water, see if a fish dies in the sample water. It might have been that the students who suggested using a microscope thought that using the microscope might solve the case (why would the textbook otherwise have suggested earlier they should do this?) Since there were no microscopes available they were not able to study the water this way. Only then, a knowledge need would have come forward, e.g. *what* other stuff or *which* bacteria to look for, since it is clear the students did not know the answer to such questions (lines 5,7). As things were, such a knowledge need had not yet emerged amongst students. Also, when the teacher asked how the comparison should be done (line 12), a student (S8 in line 13) answered very profoundly that one could use fish for this. The student clearly thought it a good suggestion and did not feel a need (yet) to find out more about how this could have been done exactly [observation].

Protocol 2E

T: 1. Ok, if you are not sure about this, what information do you need?
S-1: 2. pH!
S-2: 3. Microscope
T [*repeats the answers*]:
4. pH and microscope
S-3: 5. To see what other stuff is in there [*I think the student refers to the microscope*]
S-4: 6. Like bacteria
S-5: 7. But there are also bacteria in tap water
T: 8. So, to be sure if the water meets the criteria for drinking water, what should I know?
[*students start to discuss the question*]
S-7: 9. Compare the water with tap water under the microscope
T: 10. Compare with tap water
Different students:
11. Yes, yes!
T: 12. How can you compare it then?
S-8: 13. You put a fish in the water and see if it dies [*T. thinks he is joking, and S-8 is offended by this, he keeps saying that they* really *do this*]

Wrapping up

[Observations, video, protocol analysis]

Teacher C turned to the procedure poster. He tried very hard to get students to add the 'next step', which was expected to be a refinement of the step "test the water" with the knowledge need "what does the water contain?". Students showed to find the returning activity of putting a next step of the procedure on the procedure poster a little hard to place. When the poster turned up, students were guessing what the teacher wanted to hear. S-1 typically gives his answer in the form of a question (line 3)

Protocol 2F

T: 1. So, you guys, we had on our poster 'water use determines water quality criteria' and 'test the water',
2. How can I extend the procedure now?
S-1: 3. With harmful bacteria?
T: 4. Maybe you mean: 'what the water contains?', right?
S-2: 5. What does it say?
T: 6. If I test the water, what do I test? [*Teacher tries to get students to express the knowledge need 'what the water contains?'*]
S-2: 7. Ingredients
T: 8. Ingredients
S-2: 9. Bacteria!
T: 10. So, what the water contains?
S-3: 11. Yes!
T [writes this down]:
12. Ok?

This strengthened the impression that the students did not experience a knowledge need yet, they showed to be rather confident about what the next step should be (compare the water with good tap water). This was also confirmed by a question a student asked later (just before the bell rang):

Protocol 2G

> S: Sir, can we ask the people of that area [*with the two water net system*] if we can have two bottles of water, the special kind and the tap water?
> T: Why would you want to do that?
> S: To compare! Maybe we might test both of them.

Conclusions

The results of the evaluation showed that students were involved in the case of the two water nets and did come up with suggestions about how to test whether the water quality meets the criteria. However, although episode 2 unfolded in several ways as expected, its function was not fulfilled in the sense that activity 3 was not sufficient to raise the intended knowledge need. The intended knowledge need could very possibly have been evoked with extra activities (students could, for example, have been asked to execute their suggestions in order to experience that they need specific information for that).

In addition, it could be doubted whether the students' purpose was to learn about the procedure which people of the authentic practice apply in order to judge water quality, as they did not show to appreciate the procedural poster as functional. This was probably due to the unfolding of episode 1 and to the observation that the teacher did not functionally set the stage for episode 2 either. Besides this, as I mentioned before in episode 1, students and teacher were not used to a procedure as object for learning. This might have interfered with the intention of making the procedure object for learning by developing it.

Finally, the teacher did not make the episodic story line explicit for students, not being prepared for this. In retrospect, making students explicitly aware that an episode has concluded and a new one is to start might have added to the clarity of the story line and therefore to the students' sense of purpose.

7.2.3 Episode 3: What does the water sample contain? (start phase 3)

At this point phase 3 started (see table 5.1), covering episodes 3-7. Students were to extend their issue and procedural knowledge in progressive cycles, in the first instance in the direction of the raised knowledge need, until a satisfactory procedure was reached (see also figure 4.6). Episode 3 was the first of the five episodes of this phase.

Episode functions

3A Students satisfy their induced knowledge need: "What does the water contain?"
3B this will raise the need for taking the next step in this process: "How much of what is allowed in drinking water?"

Reviewing the episode functions

The previous episode showed that the students very probably did not experience a knowledge need yet. They showed to be rather confident about the next step and how it should be taken: the water should be tested by comparing it with tap water (using a microscope, doing an experiment with fish etc.). However, students were to learn in episode 3 what things are tested in reality in a case like this and how the water samples can be tested for these same things. Although this information did not directly fit the step students came up with in the previous episode, it might still have been the case that they appreciated this information as functional. It provided for information on how people solve the two water nets case in the authentic practice and how the same might be done in the instructional version of the practice.

Evaluation and conclusions

I will only present a summary of the evaluation details, but the conclusions in full of episode 3. For a detailed evaluation of the teaching-learning process, see Westbroek et a.(2005)

Although they did not have a knowledge need at the start of this episode, there are strong indications that students did experience the information presented in this episode as useful. All groups automatically and naturally used the information in solving the case [observations, video, work sheets] They generally did the tests, compared their results with the list of criteria (how much of what is allowed in the water), as part of episode 4, and formulated a conclusion (a quality judgment). So, the next knowledge need for a reference to compare the test results with occurred to students naturally. In fact, the students in this episode already largely satisfied this knowledge need. All groups considered the water quality as inadequate, due to the low acidity value.

Students expressed doubts, which referred to the accuracy of the tests and to their performances [observations]. They asked questions, which referred to these doubts and the teacher asked them to write down the different types of doubts in their workbooks. Twelve of the 16 students whose worksheets were analysed wrote down different doubts that refer to the accuracy of the tests and also to their performances:

> *Chloride test: is there a precipitation or not?*
> *The acidity results are not the same.*
> *We could not see the differences properly when comparing colours (pH test)*
> *We are not sure if all the chloride is out of the solution (if all the chloride had precipitated)?*
> *Did we fold the filter correctly [chloride test]?*
> [Several students were also not sure about the nitrite test that did not 'show colour':] *Was its concentration zero?*

The groups came up with the following test results:

Table 7.2: The test results poster.
(Completed in the third lesson)

Group	1	2	3	4	5	6	7	8
Chloride	25 mg/L	-	70 mg/L	30 mg/L	25 mg/L	10 mg/L	25 mg/L	20 mg/l
E-coli	negative	negative	negative	negative	negative	negative	negative	negative
Nitrite	0	0	0	0	0	0	0	0
Acidity Merck pH paper pH meter	4,8 5-6 3,9	4,5 3 4,5	5.4 paper between 5 and 6	4,5 3 4,0	5 3(paper)	4,75 5 4,5	5 3 4,0	4,2 4,5 3,5

The conclusion is: in spite of the incomplete unfolding of the previous episode, episode 3 unfolded mostly as intended and expected and its functions were fulfilled. Some points of episode 4 were already fulfilled. Students produced test results and understood that they should compare them with a reference. There were no indications that the design needs rethinking, although the teacher might have involved the students more in the stage setting.

7.2.4 Episode 4: Does the water quality meet the criteria for drinking water? (phase 3)

Episode functions

Students are to link their doubts to the following categories:

4A Trustworthiness of the limited list of parameters and norms.
4B Trustworthiness of the test results.

Evaluation and conclusions

I will only present a summary of the evaluation details, but the conclusions in full of episode 4. For a detailed evaluation of the teaching-learning process, see Westbroek, et al. (2005).

Although several things unfolded as expected, the functions of episode 4 were not fulfilled as intended. The design needs rethinking.
Besides this, it can be concluded that the description of the functions is inadequate.

Setting the stage for the episode question

[Observations, video]
Similar to the previous episode, the teacher (not explicitly prepared for this) immediately started with episode 4, without making explicit that a new episode had started. The teacher asked the students to discuss the questions of this episode (see below) without functionally introducing the episode.

Addressing the episode question

[Observations, video, protocol analysis, work sheets]

The expected doubts were raised about both the test results and the list of criteria in the previous section, but as students already solved the case, they did not feel the need to solve their doubts about the test results for *that* purpose.
Students were asked to discuss the following questions:

Question 5A: Does your water sample meet the above-mentioned quality criteria?
Yes/no, I am/ am not sure of that, because…

Question 5B: Suppose the water does meet the above-mentioned quality criteria, would you trust it and drink it?
I would /would not trust it, because…

Based on the worksheets that were analysed (4 groups, 16 students) it was concluded that question 5A did not make the students bring out doubts about the test results (see the evaluation of episode 3 for a list) as intended but that question 5B did make the students bring forward the expected doubts about the limited list of parameters and norms.
This suggests that the question 'Is the water clean enough for drinking' as the main evaluative question in solving the case of the two water nets (in order to learn about the procedure) needs rethinking. In retrospect, this question did not provide for a reason to check *all* parameters whatever the results were and draw conclusions on *all* parameters (which would include solving doubts). In fact, in the authentic practice the testing does not stop now that one parameter exceeds the norm. The students probably measured everything because they were told to do so, not because they felt the need: Why measure everything if the pH is already clearly exceeding the norm? In the authentic practice, all parameters are always tested and the accuracy of the methods is always reviewed, because if something is wrong it will e.g. provide for relevant information on a probable cause for this deviation. The authentic practice primarily addresses a different question: what is the water quality in view of its function?

Evaluation of the episode question

[Protocol analysis, video and observation]
In the evaluation, the teacher collected the students' doubts and wrote them down on the black board. He encouraged students to speak out and made the impression that he took the students answers seriously. He took time to listen and to verify his interpretation of what the students said. He also frequently repeated their words before writing down their remarks, which strengthened the impression that he took the remarks of students seriously. The students responded in a confident way and when they confirmed that the teacher had summarised their remarks properly, it sounded like "this is exactly what we mean". From this, it can be concluded that students experienced the summary of the teacher as their input. It also confirms that the students did have the expected doubts about the test results, but they were just not functionally addressed (in the sense that they needed to be solved for the intended purpose, put forward by question 5A).
Next, the teacher collected the doubts of the students about the list of parameters and norms, put forward by question 5B.

The teacher finally listed the doubts as follows on the 'Doubts poster' in front of the classroom:

Figure 7.1: List of doubts of the students

Do we trust the test results?	Do we trust the limited list of criteria?
Chloride test: precipitation? The acidity results are not the same Estimation mistakes when comparing colours Pouring the water through the filter: is all the chloride out? Did we fold the filter correctly [*chloride test*] Nitrite test tube shows no colour	Why these four? Should we not test the water on something else, more things?

The next step in the procedure, which was supposed to come forward at the end of this episode, is 'Judging'. Students had compared their test results with the reference and were asked if they could now judge the water. But in wrapping up this episode, the teacher interpreted the last step, judging, not as intended: not as judging the water quality, but as judging the test results and the list of parameters and norms (protocol). As this was actually what students were going to do in the next episodes, it did not interfere with the story line and it did not show to disrupt the teaching-learning process in that sense. So one step was not carried out explicitly, but that did not interfere problematically with the story line.
In the evaluation of the lessons, the teacher himself said that he found it very difficult to let *students* arrive at the 'right procedure steps', he felt he had to pull it out of the students. This type of interaction was more confusing to the students than clarifying [Protocol analysis, observations].

The conclusion is that students did have the intended doubts and they did see the logic of categorizing their doubts. However, the design of the activity needs rethinking. Question 5B did bring forward students' doubts about the list of water quality criteria, but students did not connect question 5A to their doubts about the test results (not surprisingly in retrospect).
Although students did not experience this episode as functional in light of their aim, the function description suggests that the function of this episode was fulfilled: students did link their doubts to the trustworthiness of the water quality criteria and the trustworthiness of the test results. However, this was *afterwards* instead of *beforehand* in light of their aim. I therefore conclude that this function description was inadequate, as it did not reflect how students, beforehand, experience the function of the episode in light of their aim. This problem of inadequate function descriptions was not limited to this episode. In retrospect the functional descriptions need rethinking. I will elaborate more on this issue in section 8.3.2.

7.2.5 Episode 5: Do we trust the list of tested parameter and their norms? (phase 3)

Episode function

The original episode function 5 was formulated as follows:
Students address the episode question by learning how the list of four parameters and norms to be routinely tested in this case is established.

Reviewing the episode function

The content-related motive to fulfil the episode function was supposed to have been triggered in episode 4. Students did come up with doubts about the list of tested parameters (Why these four? Are there no other parameters?) in episode 4. But, in retrospect, students had also already solved the exemplary case in episode 3 and therefore could not have felt the need to address these doubts for this purpose. However, it was not until episode 6 that it was realised that (because of this) both functions of episode 5 and 6 actually needed rethinking. This would have had consequences for the stage setting by the teacher. In retrospect, instead of setting the stage for the activities of episode 5 in view of the purpose of solving the case, he should have set the stage for the activities in view of the purpose of developing the procedure further.
Nevertheless, it could be that students still might have seen the point of critically evaluating the limited list of parameters as they based their decision on it. Also, students might have seen that solving their doubts by addressing the question ‘Do we trust the list of tested parameters and norms?’ would contribute to the development of the procedure (step: find out and decide if the choice of parameters and norms are accurate).

Main conclusions

Although episode 5 unfolded in some ways as expected, its functions were not fulfilled as intended. Students did learn about the official list of drink water quality criteria and proved to grasp at least part of the argumentation behind the choice for the four tested parameters. However, they did not beforehand experience the functionality of the activities in this episode, as necessary steps in the development of the procedure. In retrospect, the main cause of this problem can be traced back to the design of the activities. They did not adequately connect to the main purpose of the episode. Besides this, the teacher also did not functionally introduce the episode and activities sufficiently.
Also, it can be concluded that the description of the functions is inadequate.

Evaluation of the episodic teaching-learning process

The original functions of episode 5 were not fulfilled because of a cumulative effect of a number of causes. The stage setting of the episode did not unfold as intended and each of the designed activities was not connected for the students to the purpose of solving the questions raised in episode 4 (in retrospect).

Setting the stage for the episode question

The teacher did not functionally introduce the official list of drinking water quality criteria, in the sense that it was an answer to the students' question 'Are there not other parameters?', which was the intention. Probably because of this the questions 'Why are the four on this long list and why are the other parameters on this long list?' did not emerge among students, and the teacher did not check this.

The conclusion that the questions did not emerge among students is based on their confusion when they were supposed to start with the activities that addressed precisely these questions. Students showed a lack of sense of purpose that only grew as they progressed with the activities.

In addition, in retrospect, it can be doubted whether the questions 'Why are these four parameters on this long list and why are the others on this long list?' are the proper questions in the first place. After introducing the long list of drinking water quality criteria, the question 'But why are only these four tested?' was probably a lot more logical (again, in retrospect). If so, this means that activities 6 and 7 simply could not have been as functional as expected.

Unfolding of the stage setting

Immediately after the wrapping up of the previous episode, the teacher started to set the stage for the activities that addressed the main question of this episode. He did this very briefly, in a 'teacher lecture'. Although he referred to the doubts of the students, he did this in a sort of 'by the way' manner (line 2): by pointing at the poster on which they listed 'Why these four and why not others?' [video].

The teacher suggested that they were going to find out 'whether the list is trustworthy' (line 3). In line 5 the teacher referred to 'a story about the long list of drinking water quality criteria in the introduction of chapter 3, without referring to its function: it answers the students' question (raised in episode 4) 'Are there no other parameters?'. From a student's perspective [observation], this made the impression that the long list of drinking water quality criteria concerned a new topic, which they were now going answer questions about (lines 5 & 6).

Protocol 5A

T:

1. Let's see, what are we going to do?.
2. In the next chapters we will try to solve these [*points at the listed doubts*].
3. First we will, in chapter three, find out whether the list of criteria is trustworthy.
4. If you look at chapter 3 [*reads the title aloud*] 'can we trust the list, determined by the government, apparently, of criteria?'
5. What are you going to do? You can see a story about the long list of drinking water quality criteria and a lot of questions about it, if you look at pages 12, 13, 14, 15 and 16.
6. You are going to answer these questions using these information maps. You can find the necessary information in these maps. I will give you about 20 minutes to work on these questions.

The introduction of chapter 3 'Can we trust the list of water quality criteria?' did refer to the function of the long list of drinking water quality criteria. But apparently this did not to provide students with a sense of purpose (see below: 'addressing the episode question'). It could very well have been that the students did not read it, but the introduction of the teacher was probably much more directing anyway. The teacher also did not verify whether the intended questions that framed the activities 6-8 'Why are these four on the list? Why are the others on the list?' were raised and they probably were not in the perspective of the unfolding of activities 6 and 7.

Addressing the main question

Activity 6

Activity 6 was supposed to be functional for students in answering the question 'Why are the four parameters on the long list?'. Students were to answer the following questions by looking up the necessary information in special information maps:

6A: Why is acidity on the list? Why must the acidity be between 7 - 9,5?
6B: Why is chloride on de the list? Why is the maximum value for chloride set at 150 mg per litre?
6C: Why is nitrite on the list? Why is the maximum value for nitrite set at 0,1 mg per litre?
(E coli-bacteria was described as an example in the student materials)

It was expected that once students found out for each of these parameters in what sense it functions as an indicator for drinking water quality, the next question would become urgent for students: why are the others on the long list?

Unfolding of activity 6

[Observations, video, worksheets]

The unfolding of activity 6 showed that students asked many questions about activity 6 and what they were supposed to do. They showed lack of sense of purpose. Students had no problem in looking up the relevant information, but they did not see the difference between the question 'Why is X on the list?' and 'Why is the maximum value for X set at Y?' as they filled in the same answers on their worksheets. Students generally started to get irritated by this when activities 6 and 7 unfolded. Several students communicated their irritation to the teacher. In addition, in the questionnaire and the evaluative interviews several students said they did not understand why they had to answer the same question twice (see also section 7.5.1). The question was (in retrospect) too vague.

Furthermore, the worksheets that were analysed showed that two types of argumentation are mixed up, which (in retrospect) is not surprising because they were both in the information maps:

1. Each parameter is on the list because there is a specific health risk
2. Each parameter is also a general indicator for drinking water quality

The first argument (health risk) provided students with an answer as to why the parameter is tested, but it was not the proper (intended) answer.

The worksheets of the four groups (16 students) which were analysed showed the following answers (question 6A-C: see above):

Question 6A

The answers of three of the four groups (12 of 16) referred to acidity as an indicator:

1. *The value [of pH] changes when something poisonous is in the water, something might be wrong.*
2. *It is a warning. If the value is below or above what is allowed, there might be substances in the water'. One student added: 'Metals might be dissolved in the water.*
3. *The pH changes when poisonous substances are present.*

One group wrote down an answer, which comes down to 'it is unhealthy if the pH of the water is too low or too high'.

Question 6B

Only one group found out that too much chloride is an indication that something might have gone wrong in the production process, and that the water might be polluted. The other groups came up with answers that came down to 'too much chloride is unhealthy'.

Question 6C

Two groups answered that a high nitrite-concentration indicates that the water might be polluted. One group added 'with organic waste'
Two groups answered that nitrite is poisonous in too high concentrations (especially for babies).

When the groups finished activity 6 they immediately turned to activity 7.

Activity 7

In activity 7, students were asked to address the question 'Why are the others on the long list?' by answering the same questions as in activity 6 for a selection of other parameters and their norms.
[Observations and the video] It is hard to say whether students saw the purpose of activity 7. They learned about the long list of drinking water quality criteria but in view of their attitude, the design of activity 7 and what happened previous to this activity (the introduction by the teacher etc.), it is doubtful whether it was functional to them. Observations and the video showed that students really tended to loose interest now. The activities took up some time (25 minutes in total) and students acted bored and disinterested. Various conversations emerged that had nothing to do with the activity. There were no indications that the activity was too difficult for students, on the contrary. They did not ask the teacher for help when working on the activity. Besides that, the worksheets of the four groups that were analysed showed that most of these students did come up with the expected answers, although, similar to activity 6, they repeated their answers to the questions about the parameters when asked about the norms. Two students did not write down all the answers.

Activity 8

In retrospect, the purpose of activity 8 was rather vague and was not formulated from the perspective of the learning processes of the students. The purpose of activity 8 was that the teacher could evaluate activities 6-7 in a general way, avoiding extensive evaluations of all the questions without ignoring the work students had put into it.

In retrospect, as the activity actually did not have a clear function in the progression of the learning process of the students, the questions must have been rather out of context for them (and the teacher) even if the students had experienced activities 6 and 7 as functional. Instead of referring back to the *purpose* of activities 6-7, 'Did we answer the questions 'Why are these four on the list?' (activity 6) and 'Why are the others on the list?' (activity 7), the questions refer in a sort of general way to the type argumentation underlying the long list of drinking water quality criteria 'Did you stumble across argumentations underlying parameters that surprised you?'. The problems of the purpose of activity 8 was reflected in the way the 'discussion' floated in the air in the end, as if no one knew how to come to some sort of conclusion [observation, video]

[Observations, video, protocol analysis] Students did not have time to think about and discuss in their groups whether they stumbled across surprising reasons why certain parameters were on the list (activity 8). The teacher addressed activity 8 in a class discussion.

At first, there was little response but later on more students called out. A few students called out 'mercury', but did not mention why mercury surprised them. The teacher grabbed this opportunity to pose the question that was meant for activity 9: If mercury is so poisonous, why is not it routinely tested?.

Protocol 5B

T:	1.	How is it possible that mercury is forbidden, but not on the list [*of standard parameters to be tested*]?
S-1:	2.	Mercury just does not occur naturally in nature. It must be dumped or something
T:	3.	Sure, but ok, it might be dumped.
S-2:	4.	The test results will probably differ if there is mercury in the water
Different students:	5.	Yes, yes!
T:	6.	Yes [*positive evaluation*]
S-3:	7.	Yes, but still, then you do not know anything
T:	8.	S-2 is right. The four give a good indication.
	9.	Of course there could be poisonous substances in this water.
	10.	In principle, many poisonous substances, mercury is just an example.
	11.	They just cannot test the water on all these parameters. It would take too much time.
	12.	All sorts of bacteria could be in the water, viruses, there could be all sorts of things in the water, not only these four.
	13.	These four give a good indication. If one of these four does not meet the criteria, then they will look further.
	14.	A lot of poisonous substances are not being tested. That is too much trouble, to test them all.

The students showed they had understood that the four were good indicators (lines 4 and 5). The teacher only emphasised two reasons: that the four were 'good indicators' (lines 6, 8, 13) and that it would take too much time to test all the possible parameters (lines 11, 14). The teacher did not pay attention to the argument that the other parameters are not probable; something S-2 actually proposed (line2). He even suggested that the other parameters *are* possible, but it just would take too much trouble to test them all (lines 10, 12).

After a short teaching lecture about risk analysis, the teacher turned back to activity 8. One student mentioned iron and the fact that this colours the water brown. One or two students called out calcium, because it is actually healthy and has a minimum norm. The 'discussion' floated in the air, there was little more to say and consequently the original purpose (if there was any at all) of the activities drifted to the background [observation] After this, the teacher immediately turned to the evaluation of the episode (see below).

Evaluation of the episode question

The evaluation of the episode question (activities 9-10) showed that although there were strong indications that students did not experience activities 6-8 as functional in view of solving their questions, they grasped the idea that the four parameters gave a 'good enough' indication. Activity 10 (Why these norms?) was skipped.

Unfolding activity 9
[Observation, video, protocol analysis]
The teacher turned to the evaluative activities in a class discussion, without letting students discuss the questions first.
The teacher asked for the students' opinions (protocol 5C, line 11) and when given, asked them for arguments (line 13) as intended. Students could agree or disagree. Few students reacted. Only the argument that the four give a good indication was put forward. Also, S-6 said she does not trust it, but referred to the test results and the fact that the pH test results did not meet the criteria. In fact she suggested that this is an indication that there might be something else in the water (line 16), which was somewhat impatiently confirmed by S-7 (line 17) as if to say this is obviously what should be done. The arguments that the others are not probable and that it would take too much time to test them all to be completely sure, had been brought up by the teacher in the 'addressing the episode question'-part of the episode (protocol 5B, lines 4, 7, 8, 9, 14), but were not mentioned here.

Protocol 5C

T [*reads aloud*]:

	1.	"Can we trust the list of quality criteria?'
	2.	Why does the laboratory test the water only on E-coli bacteria, chloride, acidity and nitrite routinely?
S-2:	3.	Because those are the most important'
S-3:	4.	If the water contains certain substances, the pH will change
T:	5.	Yes, very good

	6.	You can say here, and I heard S-2 and previously S-1 mention this, that these four give an indication of the water quality.
	7.	If one of these four is not good, then they will look further.
	8.	[*Reads aloud*] 'Can you think of a situation in which the laboratory will test on extra parameters?', well that should be easy to answer.
S-4:	9.	If one of those four is not right
T:	10.	Exactly
	11.	[R*eads aloud*] You tested the four parameters yourself; do you consider that good enough? Who has an opinion on that?
S-5	12.	I think it is okay
T:	13.	You think it is okay, S-5, why?
S-5:	14.	Well if something else is in the water, then the others [the four parameters] will change. So then you will know that something is wrong.
S-6:	15.	Actually, I do not trust it, because the pH is too low everywhere [*refers to the test results*]
	16.	So we should look if there is something in the water.
S-7 [a bit impatient to S-6]:	17.	Well, that *is* exactly what they do

Finally, the teacher evaluated activity 10 ("Do you trust the norms?; Why these norms?") very shallow and a bit chaotic. It showed that both teacher and students did not grasp the question.

Wrapping up

The teacher wrapped up the episode by briefly recapitulating the steps of the procedure so far:

Steps of the procedure

1. Determine water function, this determines water quality criteria	What is allowed in drinking water? How much of this is allowed in drinking water? **Selection parameters and norms**
2. Test	What does the water sample contain? How much of this does the water contain?
3. Compare	
4. Find out about doubts	Do we trust the testresults? i. Did we perform okay? ii. Accuracy Do we trust the fact that only four **parameters** are to be tested? How are these four **parameters** selected? How are the **norms** established?
5. Judge	

He asked the students how the ‘what’ and the ‘how much’ might be called now in order to add the terms ‘parameter’ and ‘norm’ to the procedure poster. The teacher did this in a sort of by the way manner, emphasising that ‘we are not talking about a *new* step in the procedure here’. The teacher actually concluded himself that this was the case based on this lesson in the try-out class. In the try out class the adding of these terms was approached as a new step ‘What new step can we add to the procedure poster?’, which did not work. Students also did not consider new terms as a new procedure step.
However, in the student materials it is indicated that the next step should be formulated now. It confused the teacher and he seemed to think that a new step must be added here (although there *was* no new step). The teacher wanted to add the step ‘choose the right parameters and norms’, and tried to drag it out of the students, who did understand what de teacher was aiming at. They made the strong impression that they were guessing the next right step [observations].

Conclusions

Although students did learn about the long list of drinking water quality criteria and did grasp at least part of the argumentation behind the four tested parameters (see also section 7.4 ‘final test), there were strong indications that they had not experience the functionality of the activities as intended (beforehand). It can be concluded that although a great part of the learning goals of this episode were achieved, students did not experience a *need-to-know*. The main cause of this deviation of the scenario is in the design of the activities. The successive activities had a cumulative effect on the students’ (but also on the teacher’s) lack of sense of purpose. Furthermore, the teacher also did not functionally set the stage for the episode as was intended.
Similar to the previous episodes, the formulation of the function description of this episode suggests that they *were* for a large part fulfilled. As I mentioned before, this suggests that the function description is not accurate as it does not properly expresses the idea of functionality from the students perspective and how students should see *beforehand* how this episode and it activities will contribute to *their* aim. The concept of function and accompanying function descriptions need to be reconsidered. I will elaborate on this more in chapter 8.

7.2.6 Episode 6: Do we trust our test results? (phase 3)

Episode Functions

The original episode functions were formulated as follows:
Students address the episode question by developing a notion of reliability and accuracy of the used test methods and how this might influence their judgment.

Reviewing the episode functions

In order to solve the exemplary case of the two water nets properly, students were supposed to see the point of addressing their doubts by answering the question ‘do we trust our test results?’. Students were supposed to have developed such a content-

related motive already in episode 4, when their doubts were collected. This had not happened as intended however (see section 7.2.4). Students did have doubts about their test results, but as the pH test results clearly exceeded the norm, they had already and justly judged the water quality as unfit for drinking. Subsequently, at this point, the students did not need to solve their doubts about the test results to solve the exemplary case. They had already done that. Of course, this was also the case in episode 5, but the problem was not fully acknowledged until episode 6.
The original episode functions therefore needed to be adjusted to this situation. It was thought, that it could still be the case that students would want to solve their doubts as a need in itself, but more importantly as a necessary step in the development of the procedure, which was supposed to be the goal of the students. It was decided with the teacher that the 'solving doubts and as a result developing the procedure further' would be the adjusted function of episode 6 which should come forward in the stage setting of the episode.

Evaluation of the episodic teaching-learning process

Of this episode I will present a summary of the evaluation here. Because this summary is still extensive, I will present the full conclusions in a next section. For a detailed evaluation of the complete teaching-learning process, see Westbroek et al. (2005)

Although episode 6 unfolded in some ways as expected, its functions were not fulfilled as intended. Students did not beforehand experience the activities as functional in view of the goal to solve their doubts and as a result develop the procedure further. A number of related causes that can mostly be traced back to the design of the activities: in a similar way as the activities of episode 5, the designed activities did not reflect the purpose of solving doubts as a necessary step in the development of the procedure.

Observations and protocol analysis of the teacher's introduction showed that the teacher did refer to the students' doubts about the test results, but not as intended. He merely read them aloud from the poster after referring to the previous episode in which the other type of doubts was solved. The teacher did not refer to the purpose of the episode: these doubts needed to be addressed by answering the question 'Do we trust our test results' because it formed a step in the procedure to be developed. He also did not distinguish between the doubts concerning the students' performances (addressed in activity 11) and the doubts about the accuracy of the tests (addressed in activities 12-14). The students listened attentively to the teacher, but as the students did not speak out at this point, it was hard to say whether they saw the point of the episode.

The short introduction of chapter 4 in the students' material also referred only very briefly (in the last sentence) and a bit indirectly to the usefulness of solving the doubts in the development of a procedure (this was the new, adjusted function):

When you performed the tests, you had to follow prescriptions without knowing what all the steps in the descriptions were for. Maybe you could not always see the colour *of the solutions very well. On the other hand, maybe you had a different test result than the other groups. This is why you might not be sure about your test results. Do you trust your test results? And can you draw a conclusion on whether the water is drinkable or not?*

Addressing the episode question: Do we trust the test results?

Activity 11

The intention of activities 11A and B was that the students would solve their doubts concerning the tests and their performances: 'Did we fold the filter correctly? Did all the chloride precipitate? Pouring the water through the filter: has all the chloride been filtered out of the water? And: the test says there is no nitrite, can that be correct?'. Students were to think in a general manner about strategies to test the tests and their performances in activity 11A before solving their doubts and reconsider their test results in 11B in the light of the development of the procedure (as students had already solved the case in episode 4). Therefore, question 11A was merely an introductory question to activity 11B.

In summary, as activity 11B was skipped by the teacher (who apparently did not see the purpose of activity 11 himself), the purpose of activity 11 was lost. In retrospect, it is questionable whether activity 11B would have evoked the intended functional discussion (in the light of developing the procedure), because the question that students were to discuss was too vaguely formulated and did not reflect the purpose of the activity.

However, activity 11A did unfold as intended and students did come up with the expected general strategies to test tests and performances. Also, there are indications that the students did feel insecure about their chloride test results in view of their performances on the tests when they were to evaluate their chloride test results as was assumed (see below). It therefore would have been possible to evoke a need to solve their doubts about the chloride test among students (see the evaluation of activity 13).

The analysed worksheets of 16 students, the class evaluation of activity 11 A and specific results of a question on the final test indicate that activity 11 A unfolded as intended.
The teacher asked the students to think in their groups about question 11A, 'How do you find out whether a test and the way you carried it out actually measures what you set out to measure?'.
The observations and video show that students generally were involved in discussing the question with their group-mates in a general manner. They came up with the expected, general, answers.
Two (8 students) of the four (16 students) groups of which the worksheets were analysed, wrote one of the two expected types of answers down in their worksheets: (1) Perform the same test with better equipment (group 1) and (2) Perform the same tests on tap water, of which you know the composition, and compare that with the actual results (group 4). This is actually the same strategy some students suggested in episode 2, when thinking about how they could find out whether the water meets

drinking water criteria. When the teacher collected the students' answers in the evaluation of activity 11A, the two strategies were put forward. Students showed that they were involved. The students did not raise objections or pose questions. This indicates that the strategies were generally accepted as proper answers. This impression is strengthened by the answers students wrote down on the test question (see section 7.4)

Activity 12
In activity 12, students were expected to solve their doubts about the accuracy of the colorimetric nitrite and pH tests in the light of the development of the procedure (as they already solved the case). Students were supposed to see the point of explicating their existing notions of the concepts variability and position and apply them in the evaluation of the accuracy of the nitrite and the pH test results. Again, in certain ways, the activity unfolded as expected, but its functions were not fulfilled in the sense that students had not experienced its functionality beforehand: achieving their aim to arrive at the procedure. The main reason, besides that the teacher did not set the stage for the purpose of the activity, was that the design of activity 12 was inadequate. As a result, students did learn about variability and position, but in a general way and not in the functional way as described above.

Unfolding of activity 12
The protocol of the introduction of activity 12 showed that the teacher did link this activity to the colorimetric nitrite and pH tests, but in an indirect way. He explained to the students what they were supposed to do and how they were supposed to do it, but he did not explain why they had to do that. As a result, the teacher did not set the stage for the *purpose* of the activity as intended: answer the activity-question 'Did we established the test results adequately for nitrite and pH as we could not see the colours very clearly (for example)'. That would have made the purpose of the activity much clearer. S-1's remarks in lines 9-10 were a first indication that students lacked a sense of purpose and that this was the result of how activity 12 was designed. S-1 put into words very accurately what activity 12 was actually all about, putting the tubes in the right order. Of course, the students were to estimate the unknown solutions of tubes 11 and 12, but putting them in the right order basically does that. It was not a functional activity in the sense that this did not reflect why students did the pH and nitrite tests: to find out whether the values of these specific parameters were within the norm. The students should have experienced that if it is hard to see the differences between the tubes (the solutions), the test results show variability.

The further unfolding of activity 12 showed that students did possess a notion of the concept variability but they did not use it in combination with the concept position to review their test results . This might simply be because they were not asked to do this. As a result, the evaluation later on remained on the level of accuracy of these types of colorimetric tests, and did not reach the level of considering whether the nitrite and pH tests were accurate enough for this case considering their variability and position.

The observations and the video showed that all groups started with the activity. Nearly all students were involved in animated discussions on the right order of the test tubes and thought of tricks that might help them (like putting a white paper behind the tubes, to see the colour better). The teacher explained that the groups should let him check their calibration sequence before they started with estimating the concentrations of the unknown solutions in test tubes 11 and 12. They were also to write down their estimations of solutions 11 and 12 and hand it over to the teacher.
As expected students frequently called out that it was difficult to see what the right order was. Not all the groups came up with the same sequence. They generally made the impression that they were involved in the activities in a concentrated manner. All the results of test tubes 11 and 12 were taken in and put on the black board. They showed the expected variability. In the evaluation, the teacher used the estimations of 11 and 12 to explain the deviation as the difference between the maximum and minimum value. The greater the difference, the greater the variability. He did not specifically direct students to think about the deviation in the pH test results, but made a more general link: these types of tests (like pH and nitrite) are never 100% accurate.
Students showed to have grasped and accepted the general idea that the inaccuracy of this type of colorimetric test results is reflected in the variability in those results, as was intended. This is based on the profoundness in the way they call out 'yes' together with the fact that in this class students were not afraid and did not hesitate to speak up when they did not understand it. Furthermore, students applied the idea of variability as a measure for the accuracy of test results easily in the next activity in which they discuss the chloride test results. This also indicated that they grasped the idea of variability.
The teacher concluded with referring to the pH and nitrite Merck kit test results, but activity 12 only consisted of the question 'Do you consider these type of tests accurate or not?' (activity 12), without including position and the question if they were accurate enough in this case.

In summary, in activity 12 students learned about accuracy and variability in a sort of general way. Activity 12 did not combine the two concepts of variability and position in one functional action: decide if a solution within the norm, based on test results. Then students would have had to decide whether the test results were accurate enough, based on their variability and their position with respect to the norm. Instead, students experienced first that if a test method is inaccurate (like the used colorimetric test method) this is reflected in the extent of variability. The unfolding of activity 13, in which the accuracy of the chloride test results should have been evaluated, confirmed the idea that activity 12 could not have been functional for students. Students primarily evaluated the test method in terms of variability and they really needed the teacher to direct them afterwards again to the case by including the position of the chloride test results in the decision about the case. The design needs to be reconsidered at this point.

Activity 13
The purpose of this activity was to address the accuracy of the chloride test and to solve possible doubts, by determining the deviation in the test results and their

position with respect to the norm, in order to decide whether the results were accurate enough to judge the water quality properly.
The unfolding of activity 13 showed that students did not experience the functionality of the activity beforehand. It was only *afterwards* that they evaluated their chloride test results for this specific case with help from the teacher. Of course, this is for a large part an accumulative effect caused by the way activity 12 unfolded, but it was also due to an inadequate design of activity 13.

Unfolding of activity 13
The worksheets of the 16 students which were analysed indicated that students generally had a sense of variability, and that the idea of position did not emerge. Furthermore, the protocol of the evaluation of activity 13 showed how the students apparently needed help from the teacher to direct their thoughts to the level of the exemplary case of the two water nets. Students were to discuss two questions A and B in activity 13:

Question 13A: What do you think of the accuracy of the chloride test? Explain
Question 13B: Do you think that this test [chloride] was accurate enough?

The 16 worksheets show the following answers:

Question 13A

Fourteen of the 16 students wrote that the test was inaccurate. Ten referred to variability**:** 'The values differ too much', 'the answers differed very much from 10 to 70', 'one group has 10 and one group has 70', 'varies a lot'. Four students referred to inadequate performances in their answers: the test is not very accurate, because the filter was probably not folded correctly or the residue (the chloride precipitation in the filter) was hard to weigh. Two students found the chloride test more accurate than the colorimetric tests. One student thought it was more accurate (than the colorimetric probably) although 'they [*the test results*] differ somewhat'. One student wrote down that this test is more accurate because it was not a colorimetric test.

Question 13B

None of the 16 students referred to the position of the test results. The expectation was that students would realise that although the test results showed a great deviation, the results were well within the norm (position). So in this case, the deviation did not matter so much and the test was accurate enough. Fourteen of the 16 students thought the test was not accurate enough in this case. Ten of these fourteen students wrote down that too many things went, or might have gone, wrong (again referring to their performances), four referred to the great variability in the test results.
The remaining two students thought the test was accurate enough. One student answered that everybody did it the same way and that although there were some extreme values, several groups had approximately 25; the other student answered that most results were pretty much the same (except for some extreme results).

The evaluation showed that the teacher needed to help the students by specifying the question. He needed to emphasise that the question was about this particular case (thereby explaining ‘accurate enough or not?’ and adding: accurate for what?) to direct students to an evaluation of the chloride test results for the specific case.

Finally, the unfolding of activity 13 and its evaluation showed two indications that students were insecure about their performances with respect to the chloride tests and felt that this influenced their test results (this issue was supposed to be addressed in activity 11B). The first indication was the students’ answers on their worksheets. Ten of the 16 students whose worksheets were analysed wrote that they thought the chloride test was not accurate enough because of their performances (‘too many things went wrong’). The second indication forms the evaluation of activity 13. A discussion started about how a group carried out the chloride test, because their test result was extreme. As I already mentioned in the evaluation of activity 11, this showed that it would have been possible to evoke among students a need to solve their doubts about the chloride test.

In summary, students did have a notion of variability and position, but activity 13 (like activity 12) did not invite students to review their test results in view of the exemplary case and was therefore not functional for students. When student were to evaluate their chloride results in activity 13, they only spoke of variability and accuracy. It was not until the teacher specifically asked them whether the inaccuracy of the chloride is really a problem in this case, in view of the chloride norm, that students called out that it does not matter because the test results are well below the norm (position) in spite of their variability.

Activity 14
Activity 14 consisted of the following questions:

14A: What does the prescription of the E-coli test say about its accuracy?
14B: Do you consider the E-coli test accurate enough? I think the E-coli test was / was not accurate enough, because…

The purpose of this activity was to address the accuracy of the E-coli test, solve possible doubts about the test and decide how this might have affected their judgment (had the acidity results been within the norm). It was expected, that the roughness of the test would come forward as a problem: the test not only gives just a very rough indication (positive or negative) the turning point also coincides with the norm.
Two problems emerged with the unfolding of activity 14. First, students had not expressed doubts about this test before (unlike the other tests). The unfolding of activity 14 showed that students did not have doubts about the accuracy of this test, and that the questions of activity 14 did not *make* them question the accuracy of the test in terms of roughness of scale. Most students answered that the test is accurate, because only two clear results are possible: positive or negative. The teacher evaluated the activity with a teaching lecture. Activity 14 proved to be inadequate in solving the didactical problem of how to make roughness of scale together with position with

respect to the norm a problem for students, so that they will see the point of reconsidering their E-coli test results from this perspective. On the contrary, students tended to see the roughness of scale of the E coli test as a sign of its accuracy.
Secondly, in retrospect, activity 14 was not in line with activities 12 and 13 in the sense that different concepts (roughness of scale together with position with respect to the norm) than the concepts of variability and position played a role in being able to reflect on the E-coli test results. This break was not recognised before and therefore underestimated in the design of the activity and in the preparation of the teacher. As a result, the function of this activity was not fulfilled either, but for other reasons than in the previous activities.

Immediately after this, the teacher turned to setting the stage for episode 7 (see below) and the overall evaluation of the episode question 'do we trust the test results', and how this effected judging the water quality, which was to link all the activities together, was skipped.

Conclusions

The adjusted function of this episode was: 'Solving doubts and as a result developing the procedure further'.
Students learned about general strategies as a way to evaluate the test methods. They also learned about the concepts of variability and position and how these can be applied when deciding the accuracy of test results. However, they understood this only *after* completing activities 11 (strategies) and 12 (variability and position). To the students it was not functional knowledge, necessary to solve their doubts in order to be able to solve the exemplary case and (finally) find out about the procedure. Therefore, although the learning goals were at least partly fulfilled, the teaching-learning process had not unfolded as intended.
Similar to the previous episodes, it can be concluded that the function description of this episode was inadequate. It does not cover the idea of functionality from the students perspective and how students should see *beforehand* how this episode and it activities will contribute to their aim. The concept of function and accompanying function descriptions need to be reconsidered. I will elaborate on this more in chapter 8.

Furthermore, similar to episode 5, the fact that students apparently did not experience the activities of episode 6 as functional in the light of achieving their intended aim to develop the procedure further is caused for an important part by the design of the activities, due to the following reasons:

A. In view of the reactions of the students (see Westbroek *et al.* (2005) for details) it can be concluded that even if points B-D (below) had been in order, students still could not have experienced the activities as functional in light of their aim. Especially not in the sense that is intended by a problem-posing approach: before getting involved in the activity.
Similar to the activities of episode 5, the activities just did not follow from the assumed motives (if they were there) of students because of the vague and general

character of the questions. The activities were not designed as functional activities in light of such motives; they were not embedded in an overall purpose.
Apart from this, the assumptions about the students' intuitive knowledge and type of motives proved to be correct. It can therefore be concluded that the didactical structure of this episode is adequate, students do have the expected intuitive knowledge and content-related motives, and these can probably be evoked, but the activities of episode 6 just did not achieve this. The problem is the design of the activities.
I will discuss this issue further in chapter 8.

B. The students' primary aim proved to be to solve the case, which they did in episode 3, and not to learn about the procedure. Because of this, it was very difficult for the teacher to adjust the episode function and make its content functional for students (besides the fact that he did not show to try this, see D).
The students' purpose of the instructional version of the practice is discussed further in chapter 8.

C. The main evaluative question 'Is the water clean enough to drink?' did evoke a motive for further testing or for solving doubts (when the pH is so clearly too low). Episode 6 was designed as functional in solving the case; students felt they already solved the case. In the authentic practice, however, all parameters are always tested. And accuracy of the test methods is always an issue. I will elaborate more on this in chapter 8.

D. Finally, the teacher also did not functionally introduce this episode, in spite of its adjusted function, which was explicitly discussed with the teacher. Furthermore, the way he guided and directed the activities and their evaluations showed that he did not always saw the point of the activities in light of the episode function himself. This was probably due to the inadequate design of the activities.

7.2.7 Episode 7: Does the water quality meet the quality criteria for drinking water? (phase 3)

Episode functions

The functions of episode 7 were formulated as follows:
In order to learn about the procedure:

- 7A students bring the exemplary case of the two water nets to a close and
- 7B feel the need to consider the usefulness of the procedure in the other exemplary cases (because they expect this to be the case).

Evaluation and conclusions

I will only present a summary of the evaluation details, but the conclusions in full of episode 3. For a detailed evaluation of the teaching-learning process, see Westbroek et al., (2005).

Although the episode unfolded for a great deal as intended and expected, its functions were not fully fulfilled as intended. This was mostly due the design. As a result, at the end of episode 7, the intended content-related motive did not emerge: students did not consider applying the procedure to the other cases to be a next functional step in their purpose to learn about the procedure.

Students were able to fill out the report and provide their judgment. They generally were able to retrieve the steps of the procedure from the report. The case of the two water nets was not explicitly closed in the evaluation of this episode, but there are indications that students felt that with the report the case was clearly closed (see also how students looked back on this episode in section 7.5.1). However, students were not motivated in episode 8 to consider whether the procedure applied to the other cases discussed in episode 1 (orientation phase). The intention was that this motive would follow from the purpose to learn about the procedure. It would have bridged this episode to the next one, thereby closing the circle of the story line.
Finally, similar to the previous episodes, it can be concluded that the description of the functions of this episode was inadequate. It did not properly expresses the idea of functionality from the students perspective and how students should see *beforehand* how this episode and it activities will contribute to *their* aim. The concept of function and accompanying function descriptions needed to be reconsidered. I will elaborate on this more in chapter 8.

7.2.8 Episode 8: To what extent does the procedure we used apply in the other exemplary cases of the orientation phase? (phase 4)

Episode functions

In order to learn about the procedure, students considered whether they could apply the procedure they explicated for the case of the two water nets to the other cases. They also reflected on what they had learned: was it worthwhile?

Main conclusions

Although students generally were able to fulfil the activities properly, and proved to have learned about the procedure, they did not experience this episode as a last, reflective step that was functional in the light of learning about the procedure. From the evaluation, it appeared that the design of the previous episodes was inadequate in generating such a motive. Moreover, the design of episode 8 was not clearly connected to such a motive. In retrospect, its function was not in line with the authentic practice of testing and judging water quality. Students were suddenly given other roles in a different type of activity. Probably because of this, they had difficulty in understanding how this activity was related to the previous ones. In addition, it can be concluded that the description of the functions was inadequate.

Evaluation of the episodic teaching-learning process

The teacher did not functionally set the stage for episode 8, which in retrospect can be traced back to the design. The description of the activity in the design and the materials was vague and not functional.

Setting the stage for the episode question

Unfolding preparation of the stage setting

[Observations, protocol] The teacher, immediately after evaluating episode 7, continued to set the stage for episode 8. Similar to what happened between episodes 1 and 2 and between 3 and 4 it was not made explicit that a new episode had started, which might have added to the clarity of the episodic story line

The teacher merely told the students what he expected them to do. The emphasis lay on ‘What are we going to do and how are we going to do this ’, and not (as intended) on the purpose of the episode (‘Why are we going to do this?’). Students did not act motivated to address the activities of episode 8. Also, in the questionnaire and evaluative interviews some students explicitly expressed that they found this last activity irrelevant (see section 7.5.1).

In lines 3-4 of protocol 8A, for example, the teacher referred to the orientation phase by mentioning the different water types. However, he did not explicitly refer to the fact that the students had now finished the exemplary case of the two water nets as was intended. Nor did he point out that they arrived at a procedure, which might have been useful in general for judging water quality. This was the intended overall purpose of episodes 2-7: to arrive at a procedure by which one might solve all the cases.

The episode question of episode 8, “Can we now apply this procedure to the other cases?”, was not put forward by the teacher in the stage setting of this episode nor in the evaluation of the previous episode.

It was indeed put forward in the textbook, but not very clear (see below: “Keep in mind what you have learned about the procedure”). The introduction of the chapter that covered episode 8 did not refer to the function of the activity of episode 8 either. In spite of the preparation, this might have confused the teacher. The text of the introduction was as follows:

Activity 16: Water with a different function

At the start of the module, you have seen different situations in which the water quality is routinely tested and judged. For this activity, you have to select one of these cases. Design a report-form for the case you select and write a short manual for the lab assistant who will have to solve the case. Keep in mind what you have learned about the procedure. You can find information [on the different water quality criteria] *on the Internet.* [A list of Internet sites followed]

Students made a distracted impression and had difficulty with understanding what exactly was expected of them. They did not act very motivated to get started (see below). In response to the questionnaire, when asked about the logic and usefulness of the activities in the module, some students mentioned the main activity of episode 8 as ‘not relevant’, ‘this should be skipped’ (see section 7.5.1). This also contributed to the conclusion that students did not experience this episode as a logical next step.

Protocol 8A

T:

1. Ok, attention please. What are we going to do the remainder of the lesson?

2. We are going to work on chapter 6.
3. In chapter 6, remember we had all those different waters at the start of these lessons?
4. Surface water, swimming water, all those different samples? [*does not wait for an answer*].
5. Well, in this last chapter, we are going to make our own report -form.
6. What do I want you to do? I want you to make with your group a report-form that someone else can fill in for a water type I am going to give to you.
7. We will give you a water type, surface water swimming water

S-1: 8. Aquarium water!

T: 9. So, in chapter 5 you filled in a report and in chapter 6 you are going to make such a report.
10. And as you know by now, every water type has its own specific water quality criteria.
11. In the back of your workbook you will find Internet sites were you could find that kind of information.
12. You will hand in this assignment next week.
13. So, next week, we want from you a report-form like the one in chapter 5 for the water type we will give you now.
14. We also want you to write a manual for someone who has never done this before and who has not taken these lessons.
15. Suppose I tell my students tomorrow that my students of class X did a really nice water research, and that they are going to do it too.
16. So the idea is that you make a sort of manual for them to use and an accompanying report-form.

S-2: 17. The report-form, do we have to make it like the one here [*refers to the report in chapter 5*]?

T: 18. Yes, like that, only for a different water type.

Students call out, they react as if they are not sure what is expected of them.

T: 19. Yes, you are going to write a manual for someone who is going to test that water.
20. Someone who does not know anything about what we did here.
21. Every group makes such a report

Addressing the episode question

[Observations, written assignments] Students slowly started to work on the assignment. They generally did not show to be very much inspired. Discussions on various topics emerged, but rarely about the activity. The teacher called the class to order a few times and urged them to work. However, all groups understood (as expected) that they had to find out about the specific water quality criteria, which applied to their case. They were all able to find those within the lesson. Some groups started with the report and the manual. They all finished it at home.

The report

All groups produced a report. Seven of the eight groups produced a report with the specific water quality criteria on it, one lacked this information. The one that lacked the water quality criteria and six of these seven reports were accurate in the sense that they contained all the necessary information. One was incomplete; it only contained a table for the test results and space for a water quality judgment.

The manual

The accompanying manual students had to write differed with respect to the type of information. Some of the groups just briefly mentioned the steps of the procedure, like a list they reproduced. Others gave detailed prescriptions of the necessary tests without writing down the steps of the procedure that should be followed (determine water function etc.). One of these last groups even found on the Internet specific tests that are used for testing aquarium water and added the brand names and descriptions in their manual. Yet another group wrote a manual for filling in the report-form, 'Here you can fill in the test results etc.' The conclusion is that the activity appeared to be too vaguely formulated for the teacher and the students. Moreover, they way the students wrote down the procedure steps as a list they reproduced adds to the conclusion that students were not used to consider a procedure as a topic to learn about. It was already concluded in the evaluation of episode 1 that when students have to learn a procedure, they usually are supposed to read it in their books and be able to reproduce it, instead of developing it themselves as intended in this module.

Evaluation of the episode question

Students handed in the assignment the next lesson, before the final test started. The teacher did not discuss it in class. He did not evaluate nor wrap up episode 8 and he did not look back with the class on the lessons. The teacher, busy with the organisation of the final tests of the learning results (see section 7.4), which took up the whole lesson, also skipped the final question of episode 8. This question was also the last question on the poster of activity 1: 'Did you find in worthwhile, looking back, to learn how they do this in practice in chemistry class?'.
By explicitly discussing this question in class and collecting students' opinions, this question should have contributed to the students' feeling that their input mattered.

An impression of the students' appreciation of the theme and what they learned is provided by the questionnaire and student interviews, in which students look back on the lessons (section 7.5.1).

Conclusions

Similar to the previous episodes, the formulation of the function description of this episode was also inadequate. It did not properly expresses the idea of functionality from the students perspective and how students should see *beforehand* how this episode and it activities will contribute to *their* aim.

Based on the evaluation results, it can be concluded that students mainly *experienced* episode 8 as not very relevant and not functional for the purpose of learning about the procedure (or any other purpose) and reacted as if they were bored by it, although students were able to fulfil the task and learned about the procedure by applying it in another case. From this I conclude that they considered episode 7 more as the final episode. A number of causes might have contributed to this:
1. The design of the previous episodes was inadequate in triggering the motive to apply the procedure to the other cases in order to learn about the procedure. The basis for such a motive should have been laid in the orientation phase and the step in which

students focussed on the exemplary case of the two water nets, in order to learn about the procedure that might be applicable to all cases. This part of the design needs rethinking.

2. The design of episode 8 was not adequate. The description of the activity was not functional, but vague and confusing for both students and teacher. Apart from that, making a report-form was probably too easy for students as it was merely a copy of the report of chapter 5 and writing a manual was probably not clear enough (observing the different interpretations of students).

Furthermore, the activities did not match with the authentic practice. Students were suddenly given other roles and were to carry out another type of activity. I will discuss this issue in Chapter 8

All this left the teacher with the difficult job to make the activities functional. The result was that the teacher emphasised, when setting the stage, what the students were supposed to do and how, without explicitly going into the purpose of the activity.

7.3 Evaluation of the teacher role and interaction

In this section the teacher role and interaction is evaluated. The main question that needs to be answered is: were the designed interaction structures implemented by the teacher as intended, did they lead to the intended interaction patterns in the classroom and did students as a result feel that their input mattered as a driving force in the teaching-learning process (*attention for student input*-d)?

The implementation of interaction structures was expected to help the teacher pay proper attention to the student input, which was supposed to contribute to the students' feeling that their input mattered (section 4.6). The idea of structuring the teacher role by interaction structures evolved from the evaluation of the second version of the design. The teachers involved in this second trial, not being prepared for this, had difficulty with paying proper attention to the student input.

For each episode, an intended interaction structure was formulated (see chapter 5 and appendix 1). The teacher was to pay attention to student input when setting the stage for the main question and to make sure that the students always bridged the content-related outcomes of the previous learning activity (their input) with its successive learning activity. Furthermore, the teacher was to pay proper attention to student input when evaluating the main question of an episode. The idea was that this should be done in such a way that the complexity of the expected content-related progression matched the complexity of the interaction structure. As a result, it was expected that the specific interaction structures would offer the teacher clues about what content-related input he might expect from students and should consider theirs, and what he should add to the content-related progression.

Figure 6.2 in the previous chapter shows in broad outlines the decision scheme that is followed in the evaluation of the interaction structures to answer the main question.

In short, for each episode the expected complexity of the content-related progression was first compared to the difficulty students had with coming up with the intended

input in order to decide if this matched the complexity of the interaction structure. The evaluation presented in section 7.2 served as an information source.
For example, in section 7.2.1 it is described how students are able to come up with and discuss the first steps of the procedure, when asked about it in activity 1 (information sources: observations, protocols and posters) and also in the evaluation of episode 1 (information source: protocol of the teachers audiotape). None of the groups had difficulty with this question and the teacher merely had to collect their input, as expected. In this case the difficulty students had with coming up with the intended input (no difficulty) matched with the expectations and with the interaction structure collect input.
Next, the teacher’s actions were evaluated: did he implement the interaction structure as intended? Did he adequately handle student input? This evaluation was mainly based on the protocol of the teacher’s audiotape (verbal interaction): did the teacher build on the previous episode by using the student input? Did he collect, summarise or categorise student input when evaluating, as intended? The main information source was the evaluation as presented in section 7.2. Additional information was collected from observations and video. The teacher’s actions were viewed from the perspective of the students: e.g. was his tone of voice or his attitude inviting, or maybe very directing or judging from their perspective?
When the teacher implemented the interaction structure as intended, the third question was: Are there indications that students were involved in the teaching-learning process in which their input is handled adequately? This question is answered by looking at the way students reacted. Conclusions are based on indications such as: did students make the impression that they wanted to call out, maybe acting impatient to do so, wanting to share their opinions and answers? Did they react enthusiastically or not, or maybe neutral? What was their tone of voice? Bored? Enthusiastic? What was their body language? Did they act distracted, bored, or did they sit up straight and look involved (sparkle in their eyes)?
Finally, the most difficult question is addressed: did students feel that their input mattered? The answers to the first three questions above will already contain indications for this fourth question. If, for instance, the students were able to deliver the required input and the teacher was able to adequately deal with their input, this as such can be taken as a contribution to their feeling that their input mattered. Even more so, if there are indications that the students were involved (question 3). Of course, each such indication in itself does not prove anything, but to the extent that they reinforce one another a firmer conclusion can be drawn. Some further indications to add to this answer concern the way students reacted at those moments of the teaching-learning process where the teacher verified with the students if he had correctly understood them. Such indications are based on field observations, video and the teacher’s audio tape (tone of voice).

The evaluation of the teacher role is presented in section 7.3.1. In section 7.3.2, conclusions will be drawn about the main question: Does the idea of implementing interaction structures to structure the teacher role live up to its aims and expectations?

7.3.1 Evaluation of the realised interaction structures

Episode 1

When setting the stage and when evaluating episode 1 the teacher only had to collect the student input. Students were mainly to make their already existing knowledge explicit and give their opinions. The unfolding of episode 1 shows that this estimation of the complexity of the content-related progression was correct, students had no difficulty in coming up with the intended input (section 7.2.1).

The evaluation of episode 1 showed that although the teacher did not functionally introduce the episode, he tried to involve students in the stage setting, by asking for their input (section 7.2.1). He asked if they could come up with more exemplary water types, which was not as intended, but he collected their input as intended (7.2.1, protocol 1A).

Furthermore, in the evaluation part, the teacher was merely supposed to collect student input on the steps of the procedure they already had formulated. The unfolding of the evaluation part however showed that the teacher *directed* the students' answers very much as if he was preoccupied with getting to the right steps of the procedure, rather than the student input. In section 7.2.1, it was concluded that the teacher felt obliged to do this because students lacked a sense of purpose. In turn, the students reacted involved in finding out what he was aiming at and when they understood this (in protocol 1H students answered with increasing confidence), they reacted involved calling out the steps (lines 7, 14, 16,18). The teacher verified whether he had understood their input correctly, upon which several student called out "yes, yes", in such a profound manner that it can be concluded that they at least felt that their input was recognised. It cannot be concluded, however, whether students felt that their input *mattered* at this point.

Episode 2

The unfolding of episode 2 showed that the complexity of the content-related progression students experienced was as expected. Students had no difficulty with explaining why they thought the water did or did not meet the drinking water quality criteria and they suggest several ways to test the water further (see section 7.2.2: analysis of the worksheets, and protocols 2D, 2E, 2G). Although the design needs rethinking, as the actual intended content-related motive did not emerge, the interaction structure " Collecting Input" was appropriate.

The teacher tried to involve students in the stage setting, but not in the intended way from the perspective of the content-related progression. He checked the students prior knowledge, but, similar to episode 1, did not functionally introduce the episode, that is he did not indicate how it contributed to the purpose of developing the procedure, by using the first steps students came up with as input (section 7.2.2, protocol 2A). Observations and the video showed that the students acted involved: they generally sat up straight, paid attention to the teacher and each other when two students told the class that they knew about places that had two water nets.

In episode 2, it was expected that the teacher only had to collect the students' opinions on whether the water was of drinking water quality or not, and he did just that.

Protocols show that he asked the students for their opinions and repeated their input to confirm (protocol 2D, lines 1,3, 6, 9; protocol 2 E, lines 1, 4, 8, 10, 12; protocol 2H, lines 1, 2, 4, 6, 8, 12). Several students called out, and the way they did this, enthusiastically, sitting straight [audio tapes and also observations and video] indicates that they were involved. The way several students called out in protocol 2E line 11: "Yes! yes!" And protocol 2D: "yes!", when the teacher verified whether he understood them correctly, indicates that they at least felt that their input was recognised. However, the result was the same as in episode 1: although the teacher implemented the interaction structure as intended and students acted involved, it could be seriously questioned whether they felt that their input mattered in the progression of the process in order to achieve a common purpose. The intended knowledge need did not emerge and students were not asked to pursue their suggestions on how to find out if the water is drinkable.

Episode 3

The way episode 3 unfolded showed that the complexity of the content-related progression students experienced was as expected. All groups produced the test results and smoothly moved on to what was supposed to be the next step: compare the test results with a reference. The teacher only needed to ask students to describe this step explicitly (see section 7.2.3). The interaction structure "Collecting Input" proved to be appropriate.

When he set the stage for episode 3, the teacher did not involve students by using their input of the previous episode. That did not to interfere with the students' general sense of purpose however. They generally showed they appreciated the usefulness of the information on what people of the authentic practice test the water on in this case as they readily applied it in the case they were working on (section 7.2.3).
The teacher evaluated the episode as intended, asking the students for the third step of the procedure (compare) and several students called out "compare!". There were no specific indications that students felt more or less involved by this, or whether they felt that their input mattered for some common purpose.

Episode 4

Although the content-related progression did not unfold in every way as intended and expected, and although the design of episode 4 needs rethinking, the students did not have trouble with the content-related progression. The teacher did not need to direct students or help them more than expected. They were merely put on the wrong track by the type of questions (see section 7.2.4). The teacher categorised the input of students, not because it was too difficult for students, but because it saved time and students were expected to appreciate the logic of the categories, which they did. Several students called out with a distinct: "yes!", when the teacher suggested categorising their doubts concerning 'Do we trust the test results?'. And later on when the teacher suggested that the second type of doubts concerned the choice of the limited list of tested parameters ('Why these four and why not others?'), several students again called out 'yes' in an emphatic way. Categorising input proved to be the appropriate interaction structure.

Although the teacher did not actively involve students in the stage setting, he evaluated the episode as intended: collecting students input, repeating their answers every now and then to confirm and categorize their input. Students acted involved and they called out their answers. The way they answered the teacher's question 'Do you think we can summarise this as 'Are the test results ok?', several students called out a profound yes!, also indicates that they were involved. From this, it could be concluded that the teacher really wanted to know what the students had to say. He asked for their input in an inviting, questioning manner. Not at all directing. He verified whether he understood correctly what the students meant and, when the students confirmed, he immediately wrote down their doubts in the way they expressed them. Students, in their turn, made the impression that they really wanted the teacher to understand what they meant by calling out and readily confirming the teacher when he correctly interpreted them. This indicates that the students felt that their input was taken seriously. Whether they also felt that their input mattered for a common purpose is more difficult to conclude. The teacher did refer a few times to later chapters where the doubts would be addressed.

Episode 5
Students did not have difficulty with the complexity of the activities, when they worked on activities 6 and 7 in the 'thinking-sharing phases'. Section 7.2.5 showed that students were able to find and discuss information on the parameters on the long list of drinking water quality criteria. Apart from the design being inadequate, the complexity of the type of content-related progression students experienced matched with the expectation and with the complexity of the interaction structure 'Thinking-sharing-exchanging'.

Section 7.2.5 showed that the students and the teacher lacked a sense of purpose because of the way the activities were designed. In addition, the activities did not show much variation. All this led to the students losing interest in this episode. They generally acted bored and disinterested as described in section 7.2.5. This probably influenced the evaluation part. The teacher was to guide students in the Exchanging-interaction towards the full argumentation. This exchanging- interaction consisted of three activities. In activity 8, students summarised the types of argumentation behind the parameters and their norms. This included the ones that were tested, and the activities 9 and 10, in which students were to exchange their ideas on why these four were tested and not others. The intention and expectation had been that the teacher merely would have to collect input and summarise the full argumentation as an extra step. The teacher did ask for students' opinions as intended (section 7.2.5, protocol 5C), but the response was not very substantial. Few students called out and only one argument was put forward (7.2.5, protocol 5C). The conclusion is: although the teacher did not use student input when setting the stage, he tried to pay proper attention to students input in the evaluation part. The unfolding of the episode and the lack of sense of purpose among students, however, caused students not to react in an involved way and probably not feeling involved in the teaching-learning process. It is therefore impossible to conclude that students felt that their input mattered for the progression of the process.

Episode 6
Students did not have difficulties with the activities. They were able to come up with strategies in activity 11A (11B was skipped). They were able to come up with the expected input in activity 12 and generally showed to have accepted the idea of variability as an indication for accuracy in activity 13. Activity 14, about the E-coli test, was the only activity they did not grasp, but this can be traced back to the inadequacy of the design. Although the design was inadequate and the episode not functional for students, the complexity of the content-related progression the students experienced was in line with what was expected and therefore matched with the complexity of the interaction structure (section 7.2.6).

The fact that the evaluation of episode 6 was skipped (section 7.2.6) indicated that students, as well as the teacher, lacked a sense of purpose like in episode 5. Only, in contrast to what happened in episode 5 where the students showed to be disinterested, in episode 6 their lack of sense of purpose did not interfere with their involvement [observations, video]. Students generally acted involved, especially in activity 12 (accuracy of the colorimetric Merck Kit tests). They discussed their findings and were willing to join the discussion when the teacher evaluated the activity (see the description of the unfolding of activity 12 in section 7.2.6). Students generally paid attention. The same happened when the chloride test results were discussed. Based on this it could be concluded that students reacted and felt involved. It is questionable, however, that they felt that their input mattered for the progression of the process.

Episode 7
The unfolding of the episode showed that students did not have difficulty with the activity, which was expected. They generally discussed the report together, acting more or less involved (see section 7.2.7, observations). It can therefore be concluded that the complexity of the interaction structure matched the complexity of the content-related progression that the students experienced, and that the goal to involve all students more or less equally in the activity was achieved.
The teacher did not actively involve students in the stage setting, but he did so in the evaluation part. The teacher introduced the reporting as an activity to finish the exemplary case, but he evaluated question 15B: “See if you can retrieve the procedure steps from the report”. The teacher asked them to retrieve the procedure steps from the report in a very directing way and the students generally paid attention and were willing to do this [observations]. Based on this, it could be concluded that students were involved in the unfolding of the evaluation part. It is very doubtful however that they felt that their input (retrieving the steps) mattered for finishing the case, because they were not asked to formulate some general conclusion or judgment that would have finished the case. Thus, apart from the activity of reporting, it is questionable if the students had a sense of purpose in this evaluative part.

Episode 8
The unfolding of episode 8 showed that, although the design needs rethinking, basically students did not have any real difficulty with this type of activity. It can be concluded that the complexity of the interaction structure matched the complexity of

the content-related progression and the purpose to involve all students more or less equally in the activity.
The teacher skipped the evaluation of episode 8. Therefore, conclusions whether the interaction structure contributed to the involvement of students, and whether students felt that their input mattered in the light of the purpose cannot be drawn.

Finally, in the evaluative interviews most of the students expressed that they generally felt that their input mattered (except for two of the eleven students that were interviewed). The remaining nine students generally referred to episode 4. One student referred to the addition of steps to the procedure. This is further elaborated upon in section 6.8.

7.3.2 Conclusions

The teacher generally tried very hard to involve students by asking for their input and by really trying to understand and appreciate their input. However, he did not involve students properly as intended when he set the stage and he sometimes skipped the evaluation.

The aim to make students feel that their input mattered is closely linked to a sense of purpose: matter for what? Although the design was not always adequate, section 7.2 showed that the way the teacher set the stage often did not provide students with the clear sense of purpose that they were suppose to have.
The unfolding of the setting the stage of episode 1 (section 7.2.1) is a typical example of that. Although in the scenario it was explained in detail what the stage setting should have been in episode 1, the descriptions of the interaction structures do not specifically indicate (in retrospect) for the teacher how to functionally introduce an episode (a set of activities about the episode question). The interaction structure only referred to the importance of a sense of purpose in a very general way and apparently did not sufficiently provide the teacher with clues.

> *When setting the stage for a question the teacher always makes sure that the students bridge the content-related outcomes of the previous learning activity with this successive learning activity. By doing this, the teacher makes sure that the question will logically emerge from the previous learning activity. The important thing is that students must experience this logic. The teacher has to verify all the time if this is the case. He/she can do that by letting students formulate 'the bridge' as well as the question themselves first, in their own words. Or he/she can pose the question and ask the students to explicit the logic of the question (the bridge). (Note: the learning activities are designed in such a way that explicit attention has been paid to the logic of the successive questions.)*
>
> (See appendix 1)

The lack of sense of purpose is not an uncommon situation for both students and teacher given the traditional school culture. In fact, it was one of the challenges the teacher had to face: how to teach adequately in a teaching-learning process in which every activity builds on the previous one and directs to the next (see section 6.3).

Episode 7 proved to be another example of this. The evaluation did not clearly contribute to the purpose of closing the exemplary case; no conclusions were drawn and discussed, the only activity the students actually did was retrieving the procedure steps from the report. However, they acted involved. Their involvement could be explained (as in some of the previous episodes such as episode 6) from the perspective of the traditional school culture: students are used to doing activities that do not have a clear content-related purpose for them. This traditional approach interferes with the approach that activities should be functional from the students' perspective and that their input should matter in achieving their aim. The latter approach asks for different interaction patterns (see also section 3.3). In this case, students answered the teacher because the teacher wanted them to (in this case the design 'told' the teacher to do this). Maybe students thought the teacher had some hidden purpose, like checking their knowledge for example. So, the teacher might have acted according to the interaction structure prescribed and, for example, have collected input as was intended and students might have acted involved, but it is questionable whether they felt that their input *mattered* in the progress of the *process* in which they were to achieve a certain goal.

In summary, section 7.3 shows that, although interaction structures helped the teacher to pay more attention to the students input, it did not sufficiently address the full scope of the problem of making students feel that their input matters. The difficulty is that students need a real sense of purpose to feel that way. A purpose they actually are contributing to. I will elaborate on this more in chapter 8.

7.4 Evaluation of the results on the final test

After the last lesson, all the students took a written final test. The test was composed of three main questions. In appendix 2, the test questions are presented together with the answering criteria for the three different categories. The construction of the test, the data collection and analysis is described in section 6.4.3. I will discuss the students' results and how this can be linked to the realised teaching-learning process.

Two questions turned out to be inadequate (2B and 3A). Question 2B does not properly reflect the step 'considering the variability and position of the test results'. Only one test result per parameter is presented on the test instead of a number of test results that show variability. The students learned in episode 6 that the accuracy of test results shows in the extent of their variability (see section 7.2.6). Although for some students the hint 'use your experiences with the Merk kit tests' was apparently enough to involve the accuracy of the phosphate test in their answer, question 2 probably (and justly) confused others.
It can also be questioned whether question 3A was an appropriate question as it only assessed the students' ability to reproduce the procedure as a list of steps they could master by root learning. In fact, question 3A rather assessed the for students familiar

learning goal (see also sections 7.2.1 and 7.6) 'be able to reproduce the list of procure steps', and not the learning goal 'be able to develop and reflect on the procedure'. The questions will not be included in the evaluation.

The students generally did well on the test. This is a further indication that the design did guide students (in general) to the intended goal to gain insight in the procedure for judging water quality.

The test results

All twenty-seven students of the focus class took the test. Table 7.2 shows an overview of their test results.

Table 7.2 Relative test scores (%) of the students' performances on the final test (n=27)

Score / Question	1A	1B	2A	2C	3B
A	4	37	52	52	48
IcS	74	44	41	7	37
F/I	22	19	7	41	15

A: Adequate
IcS: Incomplete, but sufficient
W/I: wrong/insufficient

Most of the students gave a incomplete but sufficient answer to question 1A. All of these incomplete answers are of the type: the four parameters give a sufficient impression of the drinking water quality. In retrospect, this is not surprising. They reflect the incomplete argumentation that came forward in the evaluation of episode 5 (see section 7.2.5). The full argumentation behind the choice for the four parameters that were to be tested was not explicitly discussed and summarised as such in the realised teaching-learning process.

Question 1B shows that most students were able to find the relevant information, the same way they did in episode 5, and write down 'incomplete but sufficient' answers or complete answers. The incomplete answers are all correct, but lack a full argumentation as to why all the substances other than methyl bromide are not probable.

Students scored well on question 2A. This shows that they understood that quality criteria such as 'healthy for humans', or 'healthy for animals' are reflected in the parameters and norms. It is not very surprising as the unfolding of episode 1 showed that students (as was expected) have intuitive knowledge of this relation between water quality criteria and water use that could easily be activated (see section 7.2.1).
Answers like 'because the criteria for ditch water are different than for drinking water' were considered incomplete but sufficient, because it did not say why this was the

case (different types of water use). The following answer was considered wrong, because it was too unclear: ‘people drink drinking water; there is more danger that someone gets ill by the water. That’s why they have stricter criteria for ditch water’. Maybe the student meant that people do not drink ditch water and therefore there is less chance that they get ill from ditch water. However, she then should have concluded that ditch water meets less strict criteria.

Most students answered on question 2C that Mei Ling should use a new test or a better test and do the test again, which reflects their answers on activity 11A of episode 6 (see section 7.2.6).
In the case of question 2C there were little ‘incomplete but sufficient’ answers, but a few students gave two answers of which one part was wrong and one part was correct. These answers were marked as wrong answers and they were all as follows: ‘repeat the test a few times and see if the results differ (wrong) or do the test with a different (new/better/Merck Kit) test (correct)’

Question 3B shows that most students are able to come up with what would be the relevant type of information on the report (3B).

It can be concluded that students scored generally well on the final test, which is a further indication, besides the evaluation of the episodes and especially episode 8, that they did acquire insight into the procedure for judging water quality as intended. The incomplete answers of students on question 1A reflect what the teacher evaluated in the teaching-learning process.

7.5 Retrospect by students and teacher

In this section it is discussed how the students and the teacher look back on the teaching-learning process.

7.5.1 Retrospect by the students

In order to get an idea of the students’ appreciation of the teaching-learning process students were given a post-trial questionnaire (inspired by Kortland, 2001, p154).
Also, the 11 students whose worksheets were analysed and who were audio taped, were interviewed in four groups of 2 students and one group of 3 students. This gave them the opportunity to clarify their answers, thus providing for additional information. The questions of the questionnaire and the evaluative interviews together cover roughly three themes: 1. How students, in retrospect, generally appreciated the teaching-learning process, 2. How they have experienced the didactical structure of the teaching-learning process: as clear/logical or not. 3. [only in the evaluative interviews, as it was considered too difficult to capture this theme in a questionnaire] whether they felt their input mattered in the process.

The questionnaire and the evaluative interviews

In appendix 3, the students’ questionnaire is presented. The first part of the questionnaire (questions 1-11) was meant to get a general impression of the students’

appreciation of the teaching-learning process. Questions 6 and 7 were specifically meant to get an impression of how the students had appreciated the unit in comparison to the normal chemistry lessons. The second part was meant to get a general impression of the students' appreciation of the clarity of the teaching-learning processes and how they perceived the logic and usefulness of the activities. Unfortunately, the last two questions about episodes 7 and 8 got lost when photocopying the questionnaire. In order to still gain some information on how the students had appreciated these last two episodes (functional or not), some of them were asked about this in the evaluative interviews.
Before the students filled in the questionnaire, its purpose, the type of questions and the fact the responses would be handled anonymously and only for research purposes, were briefly explained to them in an introduction.

Students' perception of the teaching-learning process

The questionnaire (n=22) and the evaluative interviews (n=11) show that students, looking back, generally appreciated the teaching-learning process.

1. General appreciation
The overview in table 7.3 of the questions 1, 2 and 9 of the questionnaire indicates that students generally appreciated the unit as interesting and important and did not find it too difficult.

Table 7.3 Student answers on questions Q1, 2 and 9. n=22

Question	Score				
	1	2	3	4	5
Q1: How interesting did you find the topic "judging water quality"?	2 (very interesting)	12	8	-	- (not interesting at all)
Q2: How important do you consider learning about this topic?	4 (important)	15	1	2	- (not important at all)
Q9: Do you think the unit was difficult or easy?	2 (very easy)	16	4	1	- (very difficult)

The results from the questions 1-11 of the questionnaire and the first three questions of the evaluative interviews show that students positively appreciated the teaching-learning process.
The students' answers to the questions 4-8, 10 and 11 specify what students appreciated or did not appreciate about the water quality lessons. On question 4, students answer the following. The topic 'judging water quality' was frequently mentioned as an important topic. Also, the feeling that they were given the opportunity to act independently and inquire issues repeatedly comes forward in both the questionnaire and the interviews (Q10, eight students; Q6, six students; Q7, two students). The feeling that the activities were all very clear and that one learned a lot without noticing (Q10, four students; Q6, three students) are typical answers that repeatedly came up. Some students also referred to a sense of purpose they

appreciated (Q7, six students). Especially in the evaluative interviews, students frequently refer in this way to experiencing content relative motives as a positive point (e.g. “ You know all the time why you do things”). All the students described that they learned a lot (question 3). The majority referred to the procedure when asked what they learned, e.g.: “How you can test water?”.

Negative points that came forward were the lack of clarity (Q10, four students; five students wrote they found the last activity not clear (Q 11, three students wanted more clarity) and that they did not have enough time (Q11, five students want more time).
Some students said they thought it all went too fast and that they were not given enough time to finish the activities properly. Also, some students found the tests more boring than regular chemistry tests.

Most students said that they appreciated the water quality lessons (Q6: nineteen of twenty-two, three liked it less) and the water tests (Q7: fifteen of twenty-two, one liked it equally and six liked it less) more than regular chemistry lessons and regular chemistry experiments.
Students wrote that they especially appreciated the feeling that you could discover things yourself (six), an answer that also came up in the evaluative interviews:

> You are much more independent in this project. Normally the teacher shows you exactly what to do and here we could find out ourselves

Also, some students said that they understood it better (three). In the evaluative interview, three students referred to a sense of purpose, e.g.:

> You know all the time why you do things, like the tests
> [student adds:] Yes, like the tests we are doing now [the student refers to some experiments they just did in the regular chemistry lesson] I don’t know what we are doing. Ok, I do it, but for what? What does it mean?
>
> I enjoyed it better than the book, because now we knew what we were doing with those tests. What you are actually doing.

The same type of answers were given when students were asked what they thought was good about the lessons in question 10.
Eight students answered that they appreciated that you could determine a lot yourself and four students appreciated that it was all very clear. One student answered: “You learned a lot and you did not even notice it”.
On the other hand, when asked what they thought was not good about the water quality lessons, four students mentioned that it was sometimes unclear what you had to do. In particular, five students mentioned that they did not see the point of the last activity (making a report and a manual, episode 8).

Of the fifteen students who wrote they appreciated the tests more, six also refer to a sense of purpose, e.g.:

We were doing real tests, with a purpose

Again, the answer that they could find things out for themselves more (two) comes forward. The six students, who appreciated the tests less, mostly say that they just thought the test not very spectacular and a bit boring.

Finally, when asked: "What do you think needs to be changed or improved about the unit?" (Q11), ten students answered plainly "nothing", five wanted "more time" and three wrote that it should be made more clear all the time what you are supposed to do.

2. Clarity

When asked about their general appreciation and especially compared to the regular chemistry lessons, several students already referred to the clarity and some to the lack of clarity they experienced in the teaching-learning process. Questions 12-15 specifically addressed this topic. The answers of the students to the more specific issues show that the students did generally experience the teaching-learning process as clear and logical, but not always and not all the time.

Students generally thought that it was clear what they were supposed to do and how after the first episode. On question 12, "Was it clear to you after the introductory chapter 1 (up to activity 3) what you were going to do the next lessons and how you were going to do this?", most students (eighteen) answered positively, three negatively and one did not know.

Students generally said that they appreciated the activities as logical. Only some specific activities were mentioned as being illogical, concerning the same activities that were mentioned in the evaluative interviews. On question 13, "Were there any activities that were not logical to you, of which you did not understand why you had to do it?", 10 students answer negatively. Twelve students answered positively. They referred to a limited number of activities: five students referred to the activities about the long list of drinking water quality parameters and norms (activities 6, 7 and 8):

S-1: Some of the questions were asked twice.
S-2: For example why nitrite is in the list, they would ask the same question twice.

Four students mentioned the manual and report-activity (episode 8). One mentioned the making of the posters in activity 1. Two just wrote down "yes".

Almost all students answered positively to question 14, "When doing the tests, did any doubts rise with you about the tests and/or the test results?", and mentioned one or two of the doubts. Two just answered "no"

The majority of the student answered positively to question 15: "Did the activities of chapter 4 help you diminish these doubts (if you had them)". None of the students

explained, except for one student: “Yes, because then [*activities of chapter 4*] you are going to compare and check”.
On the question if students thought at any point ‘Why are we doing this?’ the same activities were put forward.

3. Feeling that your input mattered
The questionnaire did not contain a question specifically addressing this theme. The results of the evaluative interviews show that at least most (S-7 & S-8 were accidentally not asked) of these students felt that their input mattered in some sense.

[The interviewer is indicated with ‘Q’]
Q: Did you feel that your answers were used in the formulation of the steps of procedure?
S-1 and S-2 said they did, and when asked further:
Q: Or did you feel that these steps were actually already there?
S-1 says: No no, because you did all the questions yourself, with the class. Sometimes the teacher directed a bit.
S-2: Yes, like, you should think more in this direction, but besides that, I think it was well..
Q: Your answers mattered?
S-1 & S-2: Yes

Q: Did you feel that your input mattered?
S-3: Yes, I did more than in regular chemistry lessons.
Q: Can you be more specific?
S-3: Yes, I just thought that you know when thinking about the procedure, you could use your own information. And also, the mistakes you could have made when doing the tests, you could mention them all [refers to episode 4].
S-4: Yes and it was all written down.

S-9, 10 and 11 all think that their input did matter:
It was all written down and then we really got into it
When asked: So it mattered what you had to say?
All of them answer distinctly:
Yes.

The answers refer to episode 4 in which the teacher wrote down all the doubts students came up with. And these doubts (student input) formed the basis for the next two chapters. It is therefore not surprising that students put this forward as a particular example of their input having mattered.

7.5.2 Retrospect by the teacher

In this section the teacher’s appreciation of the teaching-learning process and especially his role is discussed. This retrospect is based on an evaluative post-trial interview and was structured around three main questions:

4. What is your general impression, looking back, of the lessons?
5. How do you look back on your role as a teacher, especially compared to the role you usually fulfil in chemistry lessons?

6. Your aim was to improve your practice by participating in his project, how do you look back on that?

The evaluative interview provided the following results. The teacher said he generally appreciated the teaching-learning process as very positive. He especially appreciated the way the activities built on each other, raising questions among students, and the way he was expected to pay attention to their input. These two aspects he appreciated most and he wanted to try to implement them in his practice to improve this. On the other hand he generally found his teaching role too demanding, partly because he had to pay real attention to the student input and partly because he experienced the didactical structure and especially the development of the procedure, in contrast to the students, as to having one's hands tied.

Question 1: What is your general impression, looking back, of the lessons?
When asked this question, the teacher explained that he generally liked the lessons, the way everything unfolded. He also had the impression that the students generally liked it and really understood what they were doing. He appreciated the aim to let students develop the procedure, and said that this is an important and new type of thing for them to learn. But he also thought the lessons too demanding for both himself and the students (see below). He especially referred to the students (and himself) not being used to developing a procedure as object of learning.

> (…) I think the students are the least familiar with having to develop a procedure. But I thought that this [developing a procedure] was very useful and it makes it all also very real. So I think it went well, and that students enjoyed it, and are able to do it. But on the other hand it was a bit much. For them but also for me.

Question 2: How do you look back on your role as a teacher, especially compared to the role you usually have in regular chemistry lessons?
The teacher saw two major differences in his role compared to his usual role: it was much more demanding in two ways: 1. in the sense that he had to think about a lot of practical things and 2. in the sense that he had to pay serious attention to student input. The latter he also considered as something that improved his practice (question 3).

About the organisational side of his role, the teacher said the following:

> It is too demanding. You can never be enough prepared. It is not that I am a lazy or easy person in that sense, but you really have to organise a lot to be able to do these lessons. You have to, if you want to do this properly, get all the water samples of all these cases, you have to do all that stuff with all these posters. That really is a drawback of this method.

About the fulfilment of his role, the teacher said the following:

> (..)..I do not claim to be such a good teacher, but I do not think every teacher could do this. (..) I do not know how to put this exactly, but if you are a

> traditional teacher, then you cannot do this, because you really need a lot of response from the students. You need quite a lot of interaction. I mean, otherwise the whole flow of the story line just comes to a hold every time. And then you have to get the thing going again. I think that is it. It has to become a sort of interactive story, don't you think? And if you are used to just telling your story, check the homework and so on, then it will not work.

On the question if it is hard to react adequately to the students' responses, the teacher said:

> Yes, yes. And on top of that, when students are used to this way of teaching [traditional] then you cannot suddenly teach in a completely different way. Students would not understand what is happening. They are not used to speak out like that.(..) Therefore, I do not think that this [way of teaching] is for everyone. Maybe the younger teachers, they are more used to different teaching methods.

Question 3: Your aim was to improve your practice by participating in his project, how do you look back on that?
The teacher mentioned two major things he said he had learned: the importance of trying to understand and pay attention to the answers from the students and how to create coherence in the activities by posing the right questions and build in evaluations. He has planned to rewrite a chapter of the regular chemistry book from this perspective.

The teacher finally added the following about question 3:

> I think that trying to get the answers from the class and using the questions as a way to go through a story line are the most important things. And with that I mean that you actually do not have to explain a lot, you let them figure out a lot themselves. Yes, I think that is a really good way of teaching. (..) And I have learned to build in evaluative moments. Sort of summarising moments in order to secure that line. You learn to see the story line, what the goal is and to work towards that goal. That is something I really miss in the regular methods. It often is a sort of bunch of topics and everything that has something to do with each other in whatever way is put together in a chapter.

7.6 Conclusions

In the third version of the design, the three characteristics of meaningful were operationalised as follows (see also section 4.6):

Context and need-to-know

a. Students should become broadly motivated by the purposes of the authentic practice, by presenting them a selection of appealing cases of the authentic practice, to adopt their role in its instructional version and find out how people in the authentic practice judge water quality by simulating the authentic practice (that is: solving an exemplary case).

b. The students' intuitive knowledge of such a procedure is used to design a problem-posing teaching-learning process, thus creating an *instructional* version of the procedure of the authentic practice.

Attention for student input

c. Student input contributes during the whole teaching-learning process towards the common purpose of developing the procedure for judging water quality.
d. The teacher directs interaction at the level of the stage setting and evaluation of *each* of the episodes of the problem-posing story line, according to the respective interaction structure (see chapter 5 and appendix 1).

The criteria based on which it was decided to what extent the design properly embodied these characteristics of meaningful and therefore could be seen as an exemplary case of meaningful chemistry education, were:

Context and *Need-to-know*

A. Are students broadly motivated by the exemplary cases and therefore the purposes of the authentic practice to adopt their role and get involved in its instructional version as intended?
B. Do students experience the intended knowledge needs at the intended moments?
That is, do students experience at the intended moments that they need specific knowledge to take the next step in light of their purpose? Do students at the intended moments experience that such specific knowledge asks for a refinement of a procedural step?

Attention for student input

C. Do students feel that their input (including the presentation of their own findings) contributes to a common purpose?
D. Are the designed interaction structures implemented as intended, do they lead to the intended interaction patterns in the classroom and do students as a result feel that their input matters as a driving force in the teaching-learning process?

Based on the evaluations in the previous sections of this chapter, the conclusion can be drawn that the third version of the design was more meaningful to students than the second version. It increasingly met the criteria for meaningful chemistry education. Designing an instructional version of the authentic practice (section 7.2) and the implementation interaction structures (section 7.3), the two main new principles in the third version, proved to contribute to the students increasingly experiencing the functionality of activities, beforehand, in light of their purpose and to their feeling that their input mattered in this process.

Students were generally motivated by the exemplary cases of the authentic practice, generally wanting to learn how such cases are solved.

The flow of the activities was generally more meaningful to them in the sense that their content-related motives were induced as expected (although not always) and as a result students felt the purpose of the flow of the activities more than in the second version. The students' questionnaire and the evaluative interviews strengthened this conclusion (see section 7.5.1). Students expressed in different ways their appreciation of the logic of the general story line of the teaching-learning process. The evaluation of the third version of the design therefore did not to give cause for adjustments at the level of the didactical structure. It can therefore be concluded that at this level the didactical structure of the design was adequate (see figures 4.6 A and B) and that the intended story line of the teaching-learning process was consistently designed.
Of course, the evaluation results also show that at the level of the activities the design was not always adequate. The activities were not always functional for students. This was due to inadequacies in the design of certain activities and of episodes and episode functions (as evaluated in section 7.2).

With respect to criterion C, students did feel, more than in the second version, throughout the teaching-learning process that their input contributed to a common purpose. Their purpose, however, was not so much the procedure but the solving of the exemplary case.

Furthermore, the teacher structurally paid more attention to the input of students. The story line of the teaching-learning process unfolded in a much more interactive way than in the second version. As a result, students were more involved and increasingly felt that their input mattered.
However, the evaluation results also show that student input was not always properly addressed and students did not always act in an involved way. This was due to inadequacies with respect to the teacher's role: its design, but also the way the teacher interpreted his role as evaluated in section 7.3.

In this section, I have summarised the main conclusions on the adequacy of the third version of the design: what worked, what did not and why not? In the next, I will point out some points for improvement and formulate some implications. In chapter 8, I will, amongst others, answer the research question and discuss the contribution of this study to didactical theory.

A. Were students broadly motivated to adopt their role and get involved in the instructional version of the practice as intended?

Although students adopted the purpose to learn about how people monitor water quality, they saw the solving of the exemplary case as sufficient for this purpose. The students did not adopt the intended purpose of learning about a general procedure to monitor water quality for all the cases presented to them in the orientation phase (phase I).
This was due to:

Aa. Students did not get a clear orientation on developing a procedure for all cases.

Ab. The teacher did not invite the students to play their part in the instructional version of the practice as intended, which also made the distinction between the authentic practice and its instructional version less clear.

B. Did students experience the intended knowledge needs at the intended moments?

This criterion was not always met.

Ba. Activity 2 of episode 1, for example, needs rethinking, as it does not contribute to the introduction of the poster with the procedure and to the invitation to students to play their part (see also D). The activities of episode 5 and 6 were not functional partly because of their design, but also because the students had already solved the case in the previous episode. The main driving question of the instructional version of the practice needs rethinking. At various points, the repeating question throughout the lessons was formulated as: 'Can we decide if this water is drinkable [can we drink it?] or do we need extra information?' And *this* question had finally been answered. Students probably did not feel the need to think further. In the authentic practice, laboratory chemists always test all the parameters in order collect all information that is needed on the water quality: what is the quality of this water with respect to the norms for it?

Bb. The reflective activity of episode 8 still was not functional for students. The motive to develop a procedure for judging *drinking* water quality did emerge, but only with the purpose to solve the particular case with the two water nets (which was satisfied, see Ba) and not with the purpose to see if the procedure might be applied to the other cases. As discussed, the main driving question needs rethinking. Besides, another cause might be in the orientation phase. Students did not get a clear orientation on developing a procedure for all cases. Students also did not get a clear focus on the case of the two water nets as an *exemplary* case to arrive at such procedure. It might be necessary to add an extra activity in which students focus on the function of solving an exemplary case in order to arrive at a procedure, as a functional step in the process.
Yet another cause might be the design of the activities of episode 8. The activities asked of students to take other roles, which formed a break in the teaching-learning process. Until episode 7 they had the role of interested students who were imitating the laboratory chemists. Episode 8 asked students to take the role of the designer of a manual and report for a fellow student.

Bc. Implicit shifts between the authentic practice and its instructional version did occur. At certain points, the instructional version of the practice could have been among emphasised more. This would have added to a sense of purpose to the students: to find out about how people solve such cases in the authentic practice.

Bd. The teacher often did not functionally introduce (and evaluate) the activities of an episode in the light of the episode question they were supposed to address. This contributed to the students' lack of sense of purpose at certain moments in the teaching-learning process (e.g. episode 1).
One cause might have been that the teacher was not used to functionally introducing and evaluate activities in a sequence where every activity builds on

the previous one. It was identified in chapter 6, section 6.3, as one of the challenges for the teacher. That the teacher still had difficulty with functionally introducing and evaluating activities of an episode after specifically being prepared for this in the preparation trajectory is probably due to the persistence of the regular school culture both students and teacher are used to. In the regular school culture it is very common that activities do not have a clear content-related purpose for students, let alone that the activities functionally build on each other in order to achieve a certain content-related goal.

Be. It was underestimated that students and teacher were not used to developing and reflecting on a procedure as a learning object. The traditional approach of learning about a procedure by root learning of lists of procedure steps (or features) interfered with this new learning goal.

C. Did students feel that their input contributed to a common purpose?

Ca. This part of the *attention for student input* characteristic actually coincides with part B: because students felt more that what they did actually contributed to the process and therefore their purpose of solving the exemplary case, they also felt more that their input contributed to a common purpose.

When addressing the problem of paying proper *attention to student input,* a distinction should be made between doing this on the one hand in the design of the activities, so that students feel (beforehand, during the activity and afterwards) that what they *do* contributes to their purpose, and on the other hand in de design and implementation of a matching teacher role (see D). I will elaborate on this in chapter 8.

D. Did students feel that their input mattered as a result of properly implemented interaction structures?

Da. The interaction structures did not cover the whole scope of the problem of how to make students feel that their input matters (see the conclusions in section 7.3). To increase the sense of purpose among students, they needed to experience how each *activity* contributed to the functions of an episode, which in turn contributed to the functions of a specific phase. The structure of the more complex interaction structures (thinking-sharing-exchanging, episodes 5 and 6), did not reflect this embeddedness of purposes and therefore did not give the teacher clear clues on how to handle student input. I will elaborate more on this in chapter 8.

Also, the influence of the traditional school culture was underestimated. It interfered with the approach that activities should be functional from the students' perspective and that their input should matter in achieving their goal (see also Bd). The latter approach asks for different interaction patterns.

Db. Students said in the evaluative interviews that they appreciated the autonomy of choice and independency they were given in the water quality lessons.

Dc. The teacher experienced in retrospect (and he also mentioned this in evaluative discussions after completing a lesson) his role as very demanding; especially when he had to secure the interactivity of the story. He also experienced the didactical structure and especially the development of the procedure, in contrast to the students, as restrictive.

The main problems that emerged from the evaluation of the design are the following:

- To achieve that students adopt their role in the instructional version of the practice and the connected purpose is to a certain extent still problematic.
- To achieve that students experience the activities as functional beforehand, in light of this purpose (which is now theirs) is to a certain extent still problematic at the level of the design of activities.
- To achieve that students feel that their input matters in the process is to a certain extent still problematic in the sense that students can only feel that their input matters if they have some sense of purpose: matter for what? The characteristic attention for students input should meet the criterion: students feel that their input matters *in the process of achieving their purpose*, instead of: students feel that their input matters.
  The evaluation showed that the design of interaction structures did not cover the full scope of this problem.
- Finally, all the three problematic issues raise the question of what the role of the teacher is in all this, and how the new approach differs from traditional school culture. That leads to the question ho the problem of the interference of the traditional school culture in this new approach can be solved.

These four problems will be discussed further in chapter 8.

Chapter 8

Conclusions and discussion

8.1 Introduction

In this chapter, I will answer the research question and discuss the contribution of this study to didactical theory with respect to the design of meaningful chemistry education.

The aim of this study is to contribute to the question of how to involve students more in the process of learning chemistry by making chemistry education more meaningful to them. The students' lack of involvement is in chapter 2 for a large part ascribed to three main problematic features of chemistry education: traditional school chemistry has the character of rhetoric of conclusions, the sequence of its content is incoherent at both concept and curriculum level and it shows a lack of attention for student input. Three widely applied potential solution strategies are identified: the use of a *context*, a *need-to-know* and *attention for student input*. These three 'characteristics of meaningful chemistry education' are adopted as solution strategies. They are expected to contribute to solving the problematic features and to involve students more in the learning of chemistry as follows.

Context
A well-defined, for students' recognisable context for concepts provides the use of the concepts with a distinct function, and thereby makes students' use of the concepts meaningful and motivating. The 'rhetoric of conclusions' feature can be avoided, when the emphasis is shifted from 'getting an overview of the conceptual products of chemistry' to the 'functional use of concepts in relation to a certain relevant, recognisable context'. A *coherent development of concepts* can be achieved in this way, as they are to be used with a 'distinct function and meaning'.

Need-to-know
Addressing students' questions on a need to know basis, which also implies building properly on their existing knowledge, provides for an increasing involvement of students in the teaching-learning process as they will see the point of what they learn every step of the way. This characteristic, together with the first characteristic of a well-defined context, can provide for the development of a *coherent emphasis.* The question 'why are we learning this?' is answered by this need-to-know approach. The first and second characteristics are both to realise a coherent sequencing of the content.

Attention for student input
The third characteristic is closely related to the second characteristic: if the aim is to really incorporate the need-to-know approach in the design of a teaching-learning process, then 'real attention for student input' is inevitable. In a successful *need-to-know* approach, students have more insight into and experience the functionality of 'what comes next'. As a result, the teacher has more opportunity to pay real attention to their input, which now should become a driving force of the content-related

progression. Consequently, students will *feel that their input matters*. Obviously, this characteristic addresses the problem of 'lack of students input'.

A proper implementation of the three characteristics of meaningful is expected to result in the students being generally motivated, seeing the point of every activity in the teaching-learning process and having the feeling that their input matters. The question of how to involve students in learning chemistry can now be rephrased as: 'how to properly implement the three characteristics of meaningful?'. Moreover, as argued in chapter 2, questions such as the following become relevant: Do students *really* know what they do and why they do it every step of the way, and how can this be linked to the design? (this is an implication of the *need-to-know* characteristic); Do they feel that their input matters and how can this be linked to the design? (*attention for student input* characteristic). To answer such questions means that the teaching-learning process should be object of study in a process called developmental research. In chapters 3 and 6 this research strategy has been discussed in detail.
Because of the intensive and detailed character of this type of research, one module is chosen as a case study of meaningful chemistry education. For several reasons described in section 2.4 and inspired by projects like *Globe*, the *Evolution of Water*, *Green* and several others, it was decided to choose 'water quality' as the theme of the module and specifically 'testing and judging the quality of surface water in the neighbourhood' (Becker *et al*., 1997; Howland *et al*., 2002; Rivet *et al.,* 2000). The central, situated question of this research was:

What is an adequate teaching-learning process in a module about judging water quality for initial chemistry education (students: 14-15) that properly embodies the three characteristics of meaningful: a motivating context, a proper need to know and a proper attention for student input?

The situated research question has been addressed in three research cycles with three successive, increasingly adequate, versions of the design. Each time the realised teaching-learning process has been evaluated in detail using a scenario of argued expectations of every teaching-learning activity. These expectations formed the argumentation as to why certain design decisions were expected to give content to the characteristics of meaningful chemistry education. They functioned as hypotheses to be tested (see section 3.3). How the first two research cycles have led to the third version of the design is discussed in chapter 4. It illustrates how each research cycle has led to the formulation of specific didactical problems that emerged from the problem of properly operationalising and implementing the three characteristics of meaningful. Each time this led to new ideas and operationalisations and to a new version of the design and the scenario. In chapter 5 the scenario of the final, third version is presented. This third version is evaluated in order to decide to what extent it can be considered as a case of meaningful chemistry education. In chapter 6, the methodology of evaluation is discussed in detail and the evaluation itself is presented in detail in chapter 7.

In this chapter, I will answer the situated research question and reflect on its contribution to didactical theory. It is structured as follows. In section 8.2, I will answer the situated research question by reconstructing the three research cycles in very broad outlines. It summarises the argumentation underlying the final (within the scope of this study) operationalisation of *context*, *need-to-know* and *attention for student input*: an instructional version of an authentic practice in which the interaction in the classroom is directed by the implementation of interaction structures. This final operationalisation and the specific didactical problems that still emerged from the evaluation, form the answer to the situated research question.

Next, in section 8.3, I will reflect on the contribution of this study to didactical theory. This is on one hand the design heuristic that emerged from the situated research question in section 8.2 and on the other hand the conclusion that, in retrospect, the idea of functional embeddedness has actually been the central directing principle in my attempt to design meaningful chemistry education. I will end this chapter with some final remarks on further research in section 8.5.

8.2 The research question

In this section, I will answer the situated research question by reconstructing the three research cycles in broad outlines (in sections 8.2.1-8.2.3 respectively). My aim is to show how the three characteristics, their interrelatedness and subsequently the criteria that were used to evaluate their effectiveness evolved in the three successive research cycles. This reconstruction forms the argumentation on which the final interpretation of the three characteristics of meaningful (within the scope of this study) is based: an instructional version of an authentic practice and the implementation of interaction structures.

Sections 8.2.1-8.2.3 are structured as follows: each section starts with a summary of the respective operationalisation of the three characteristics of meaningful. Next, the accompanying criteria are presented on which these operationalisations are evaluated. Finally, the main evaluation results are presented and the main implications for the next operationalisation of the three characteristics that emerge from this are summarised.

8.2.1 The first research cycle

The first research cycle is presented in more detail in section 4.4.

The three characteristics of meaningful

In the first version of the design, the three characteristics of meaningful were operationalised as follows (see section 4.4.1):

Context:

A relevant and recognisable driving question that is expected to appeal to students
In this case: Is the water in our neighbourhood clean enough?

Need-to-know:

A set of sub-questions that students are expected to need for answering the driving question. The sub-questions are derived from the driving question by putting oneself in the position of the students keeping in mind the learning goals.

Because it is a learning goal that students master the procedure for judging water quality, some of the sub-questions can, in retrospect, be seen as somewhat reflecting the steps of this procedure (see figure 4.1 for an overview of the steps), although at the time this was not the explicit aim. See table 4.2 for a general outline of the first version of the design.

Attention for student input:

This characteristic is operationalised in two ways:

a. Students experience certain autonomy of choice, in particular with respect to which water they want to test.
b. Students present their own findings to their classmates. In a final reflective class discussion, guided by the teacher, the conclusion emerges that they all basically followed the same procedure.
   The insight that each group basically followed the same procedure and the realisation that all groups contributed to this insight should add to the students feeling that their input matters.

Criteria

To establish if and how meaningful chemistry education was achieved, the realised teaching-learning process was evaluated according to the following criteria (see section 4.4.2):

1a *Context*:

Does the driving question appeal to students in the sense that they relate it to their daily life? Are they motivated to answer it?

1b *Need-to-know*:

Do students experience what they learn by answering the sub-questions as contributions to answering the driving question?

1c *Attention for student input*:

Do students feel that their input matters in the sense that a. they experience and appreciate that they can make their own choices and b. that their input (their findings) contributes to achieving the learning goal of making a common procedure explicit?

Evaluation

The evaluation of the first version shows that criterion 1a (*context*) was met, that of criterion 1c (*attention for student input*) at least part-a was met and that criterion 1b(*need-to-know*) was not met (see section 4.4.3).

Students were very motivated to answer the driving question, as they considered it an important question (criterion 1a). They were also excited by the fact that they could choose their own water and present their own findings and they appreciated this autonomy of choice (criterion 1c, part-b). However, criterion 1b was not met. Students did not see the point of many sub-questions and as a result did not experience what they learned by answering the sub-question as contributions to answering the driving question. Also, the final class discussion did not unfold as intended. Students did not explicitly conclude that a common procedure was followed. It has to be concluded that the procedure (the learning goal) was not a well-integrated part of the teaching-learning process and that the insight and realisation that all groups contributed to the learning goal of making the common procedure explicit was not achieved. Part-b of criterion 1c was not met.

Implications

In order to solve the problems with the *need-to-know* and to better integrate the procedure, a problem-posing approach was adopted as a refinement of the need-to-know (Klaassen, 1995; Kortland, 2001; see section 4.4.4). This approach was expected to solve the problems as follows. A problem-posing approach basically means that the teaching-learning activities are designed in such a way that students are put in a position in which they feel the need and see the point, in light of some broad motive (e.g. to answer an appropriate, top-down established, driving question) to extend their knowledge in a certain direction. A problem-posing approach is primarily a way of emphasising that in the design and in the evaluation of the teaching-learning process special and detailed attention is paid to the creation of such a broad motive and connected 'knowledge needs' or 'content-related motives' among students (Klaassen, 1995). In the first version of the design this was clearly not achieved.
The challenge for the second version was to design teaching-learning activities in such a way that they build on the prior one and induce a content-related motive in the direction of the next teaching-learning activity, and so on. The designed teaching-learning activities should thereby, evidently, make proper use of the already existing interests and intuitive notions students have. So, instead of using a set of sub-questions to frame the teaching-learning process this should now be done by the content-related motives of students.

The procedure was expected to become an integrated part of the teaching-learning process as follows. *Firstly*, students should develop a broad motive for the purpose of explicitly developing the procedure for judging water quality (how do people do this?; this was expected to contribute to *attention for student input*-part-b). *Secondly*, the procedural steps were used to evoke content-related motives by using the intuitive knowledge of students about what would be the logical next step and by using the quality control question 'Would you now drink the water, do you think it meets the criteria?'. Each next procedural step was explicitly summarised in an extra activity as a story line. *Finally,* the procedure (which should be the purpose of the students) was explicitly reflected upon in a concluding activity. See figure 4.3 for the adjusted overview of the successive procedural steps.
A problem-posing approach has implications for *the attention for student input* characteristic as well. Apart from the aspects of autonomy of choice and contributing to a common purpose, student input should now become an important driving force in the process. The teacher should pay attention to the student input, as far as it is an important driving force of the teaching-learning process, e.g. when students, at certain

points in the teaching-learning process, express their content-related motives as questions.

8.2.2 The second research cycle

The second research cycle is presented in more detail in section 4.5.

The three characteristics of meaningful

Based on the conclusions of the first research cycle, the three characteristics of meaningful are given content in the second version of the design as follows (see section 4.5.1):

Context:

A relevant and recognisable driving question which is expected to appeal to students.

In this case: Is the water clean enough?

The additional ‘in our neighbourhood’ is left out of this version, as all students are now to produce drinking water from surface water and test whether the water quality of their product meets the criteria for *drinking* water. They are not testing and judging surface water in the neighbourhood to see whether the water meets the criteria for its intended use.

Need-to-know:

Within an overall broad motive (corresponding with the driving question) a series of connected and nested content-related motives should be triggered among students. This should be done by using students’ intuitive knowledge of what would be the next logical step of the procedure and the quality control question to induce the need for a next necessary procedure step. This way the students’ content-related motives frame the sequence of activities and the procedure can be expected to become a truly integrated part of the teaching-learning process.

See figure 4.4 in section 4.5.2 for the didactical structure of the second version of the design: a problem-posing story line in which the intended content-related motives are made visible.

Attention for student input:

a. Students have certain autonomy of choice, in particular a choice of the surface water from which they are to produce drinking water.
b. The input of all groups is necessary to achieve the learning goal of making the common procedure for judging water quality explicit. Students develop the insight that each group basically follows the same procedure and realise that all groups contribute to this insight throughout the teaching-learning process. This insight is made explicit in the concluding reflective activity.
c. The teacher should pay attention to student input, as far as it is an important driving force of the teaching-learning process, e.g. when students, at certain points in the teaching-learning process, express their content-related motives as questions.

Criteria

The criteria for the evaluation of the effectiveness of the solution strategies needed to correspond to the new operationalisation of the *need-to-know* and therefore the *attention for student input*.

2a *Context*:
Does the driving question appeal to students in the sense that they relate the driving question to their daily life and does this result in them being broadly motivated to answer it?

2b *Need-to-know*:
Is it possible to trigger the intended content-related motives at the intended moments in the teaching-learning process among students in light of their broad motive to answer the driving question?

2c *Attention for student input*:
Do students feel that their input matters in the sense that

a. they experience and appreciate that they can make their own choices?
b. they experience throughout the teaching-learning process that their input contributes to achieving the learning goal of making a common procedure explicit?
c. they experience that their input, in the form of their content-related motives, is now an important driving force in the teaching-learning process?

Evaluation

The evaluation of criterion 2a (*context*) and 2b (*need-to-know*) shows that students felt broadly motivated by the driving question and that the intended content-related motives were triggered among students (or could have been) at some points in the teaching-learning process. However, this was not always the case (see section 4.5.3). Students still did not see the point of some activities, such as the summarising activities and the concluding activity in which students were to reflect on the procedure. At some points they lost track of the driving question altogether, for example, when they had to purify surface water and produce drinking water. This activity was included to let students choose their own surface water (which was expected to contribute to the *attention for student input* characteristic) but it interrupted the problem-posing story line. It can be concluded that the didactical structure as presented in figure 4.4 still was not adequate.

The evaluation of criterion 2c part-a shows that students still experienced and appreciated the autonomy to choose their own water and present their own findings, although the activity of producing drinking water quality from surface water interrupted the problem-posing story line and undermined the *need-to-know*. For the *attention for student input* characteristic it was an adequate operationalisation of part-a. However, with respect to criterion 2c part-c it can be concluded that in the designed teaching-learning process no structural explicit attention was paid to the role of the teachers and how they could have made students feel that their input mattered as a driving force in the teaching-learning process. As a result, the teachers often tended to largely ignore student input. In line with Lemke's findings (1990), the teachers relied on certain types of interaction patterns (or dialogue structures) that are common in traditional science teaching where the teacher is largely responsible for the content-related progression (see section 4.5.4). This interfered with a problem-posing teaching-learning process in which student input forms a driving force. Consequently, at this level criterion 2c was not met: the type of interaction patterns prevented that students sufficiently felt that their input mattered in the process.

Finally, the fact that students did not see the point of the activities in which the procedural steps were summarised and of the concluding activity in which the procedure was reflected upon shows that the procedure still was not a well-integrated part of the teaching-learning process. The problem was not that students did not have the expected intuitive knowledge of the procedure. The evaluation shows that they did and students proved to have experienced the logic of many procedural steps. They just did not see making the procedure explicit as their goal. Therefore, part-b of criterion 2c was still not met.

Implications

The evaluation has the following implications for the interpretation of the three characteristics of meaningful and their interrelatedness (presented in more detail in section 4.5.4).

With respect to the *context* and *need-to-know* characteristic, it was concluded that the idea of a driving question and the broad motive to answer such a question (*context)* did not sufficiently direct what content-related motives could be triggered among students at which moments (*need-to-know*). The relationship between *context* and *need-to-know* (broad motive and content-related motives) needed to be strengthened in this sense. This led to the idea of establishing an instructional version of an authentic practice[12] (Bulte *et al*., 2002).

In the way that I am going to use the term, a practice involves a central, characteristic procedure (or activity) with distinct purposes and aims (in this case: judging water quality in order to decide whether it meets the criteria for its use). The characteristic procedure that is followed in such a practice is functional for achieving these purposes and aims. In this case, that would be the water quality judging procedure that is used by chemistry analysts in a lab using standard methods.

The basic idea is that such a characteristic procedure can be used to strengthen the relationship between *context* and *need-to-know*, thereby making the procedure an integrated part of the teaching-learning process, if the following conditions are met.

a. Students should appreciate the purposes and aims of the authentic practice, by presenting them appealing exemplary cases of the practice (see section 5.3.1 for the cases used in the water quality lessons). This contributes to a broad motive to learn about the practice by simulating one of the exemplary cases (in this case: the example of the two water nets, see section 5.3), in the sense that it relates what is going to happen in chemistry class to society at large[13].

[12] It is inspired by Van Oers' (1998) and Van Aalsvoort's (2000) interpretations of the concept 'context' as a 'communal enterprise or practice'

[13] Van Oers and Van Aalsvoort emphasise this aspect of enculturalization as an aim in itself.

b. Students should have intuitive knowledge of the characteristic procedure that is followed in the authentic practice and appreciate its functionality in achieving its purposes and aims. The evaluations of the previous versions proved that students do have such intuitive knowledge of the procedural steps. This intuitive knowledge and appreciation of 'what would be a logical next step' can be used to induce content-related motives among students. That is, in order to take a next logical step students will at times find that they need more detailed information of a certain kind, which can be obtained by consulting the authentic practice. At other times, such information (e.g. the parameters to be tested) will induce the need for a refinement of such a procedural step. Examples of this can be found in section 4.6 and, in more detail, in chapter 5.

Framed as a heuristic for the design process, it can be said that the task is to determine the structure of functional activities of the authentic practice and to transform this structure into a didactical structure of *good reasons* for students to perform actions in the instructional version of this authentic practice. If this is done properly, students see the point of every activity in the instructional version of the practice beforehand, in light of achieving their aim. Although achieving this is not a straightforward process, this design heuristic is at any rate much more directive and supportive than the idea of a driving question, and can be expected to lead to a more cohesive didactical structure.

With respect to the *attention for student input* characteristic, the findings with the second version of the design show that the interaction between the teacher and the students should be structured at the level of *each* activity and within a format that matches the complexity of the content-related progression of the respective activity.
It is concluded that the teacher should be given explicit guidelines for how to direct interaction in order to achieve this. The teacher should, for example, give students the opportunity to express the content-related motives that emerged from an activity (to be addressed in the next activity, and so on) in such a way that they feel that by doing this they drive the teaching-learning process. These considerations have led to the design and implementation of the interaction structures (see section 4.6 and appendix 1).

The interaction structures were designed as follows. The teaching-learning process of the third version was divided into episodes. Every episode can be seen as addressing a question, the 'episode question', in one or more activities depending on the complexity of the question. Such an episode-question is, in fact, a content-related motive and reflects what the episode is about (except for the first episode in which students are to develop a broad motive). The basic structure of each episode is the following:

Setting the stage for the episode question
Addressing the episode question
Evaluating the episode question

An interaction structure covers a whole episode and the teacher directs the interaction in the part of preparing a question, but especially during the evaluation. The type of student input, its complexity, determines the type of interaction. A typical guideline for the teacher is, for example, that s/he can suffice with collecting students' opinions, because s/he can rely on the fact that students are capable and see the point of expressing the intended opinions. In more complex episodes the teacher can be provided with a guideline such as that s/he now needs to summarise or categorise student input, maybe in a question that reflects the students' content-related motive, because students can not be expected to be able to do this themselves (but they should see the point of the teacher doing this).

Finally, with respect to the autonomy of choice aspect of the *attention for student input* characteristic: in an instructional version of the authentic practice, students will have less autonomy of choice than in the second version of the design, because the instructional version of the authentic practice now dictates which cases are relevant to present to the students. It is up to the designer/research to decide which selection of cases would be appealing and appropriate to present to the students to broadly motivate them and which specific case would be interesting and appropriate as an example to solve (see section 5.3.1 for such argumentations).

8.2.3 The third research cycle

The third research cycle is presented in detail in section 4.6 (third operationalisation of the three characteristics), chapters 5 (the scenario), 6 (the final test: methods of evaluation) and 7 (the evaluation).

The three characteristics of meaningful

In the third version of the design the three characteristics of meaningful were now clearly interrelated and given content as follows (see section 4.6.1 and chapter 5):

Context and *Need-to-know*:

The reconceptualisation of (the connections between) the characteristics *context* and *need-to-know* led to the idea of designing an instructional version of an authentic practice.

Students should become broadly motivated by the purposes of the authentic practice by presenting them a selection of appealing cases of the authentic practice. They should become motivated to adopt their role in its instructional version and find out how people in the authentic practice judge water quality by simulating the authentic practice in a sense (that is: solving an exemplary case). The students' intuitive knowledge of such a procedure is used to design a problem-posing teaching-learning process, thus creating an *instructional* version of the procedure of the authentic practice.

The didactical structure of the third version of the design is presented and explained in section 4.6. Figures 4.6.A and B show respectively the complete didactical structure of the third version of the design and the didactical structure of phase 3. These figures show how the characteristic procedure is used to induce content-related motives.

Attention for student input

In the third version this characteristic is operationalised as follows:

a. Student input contributes throughout the whole teaching-learning process to the common purpose of developing the procedure for judging water quality.
b. The teacher directs interaction at the level of setting the stage and evaluation of *each* of the episodes of the problem-posing story line, according to the respective interaction structure (see chapter 5 and appendix 1).

Criteria

3a *Context* and *Need-to-know*

Are students broadly motivated by the exemplary cases and therefore the purposes of the authentic practice to adopt their role and get involved in its instructional version as intended? And do students experience at the intended moments that they need specific knowledge to take the next step in light of their purpose? Do students at the intended moments experience that such specific knowledge requires a refinement of a procedural step?

3b *Attention for student input*

a. Do students feel that their input (including the presentation of their own findings) contributes to a common purpose?
b. Are the designed interaction structures implemented as intended, do they lead to the intended interaction patterns in the classroom and do students as a result feel that their input matters as a driving force in the teaching-learning process?

Evaluation

With respect to criterion 3a (*context* and *need-to-know*) the following can be concluded (see also section 7.2 and 7.6).

1. Students were motivated by the purpose of the authentic practice through the selection of appealing exemplary cases. They considered water quality control a very important issue. However, the evaluation also shows that students did not explicitly adopt their roles and the specific connected purpose of making the common procedure explicit. This was due to the design of the activities and to the fact that the teacher did not explicitly invite the students to do so in the stage setting of the instructional practice.

   Students gradually became aware that ‘finding out how people solve such cases’ was the purpose and when they did, they *were* motivated to adopt it. However, in the course of the events they generally saw solving the exemplary case of drinking water quality as their main purpose (instead of making the common procedure explicit).

2. In comparison to the second version, students' content-related motives were more induced as expected and intended in light of their aim (to solve the case). As a result students felt the purpose of the flow of the activities more than in the second version. It can be concluded that the evaluation of the third version of the design does not to give cause for adjustments at the level of the didactical structure and that at this level the didactical structure of the design is adequate (figures 4.6 A and B). The intended story line of the teaching-learning process is consistently designed.
3. The design of the activities needs refining. The fact that some of the content-related motives were not induced as intended is due to the design of the activities. The evaluation shows that the intended content-related motives *could* have been triggered had the activities been properly designed, for example because they emerged at other moments in the teaching-learning process. Especially in the unfolding of the more complex episodes, which constitute of a series of activities, students tended to loose track of the purpose of their activities. As a result, they did not experience the activities of these episodes as functional *beforehand*, in light of achieving their purpose, but rather afterwards.
4. Students (and, in retrospect, the teacher too) also had difficulty in seeing the point of activities in which students formulate the procedure steps to be written down on a poster. Students also did not see the point (beforehand or afterwards) of the concluding activity by which they were to reflect on the common procedure. It can be concluded that the students still did not consider the procedure as the learning goal. In other words: the procedure still was not sufficiently integrated in the teaching-learning process.

With respect to criterion 3b (*attention for student input*), the following can be concluded (see section 7.3).

A. With respect to part-a, students did feel, more than in the second version, throughout the teaching-learning process that their input contributed to a common purpose (only not making the common procedure explicit but solving the exemplary case).

B. With respect to part-b, the teacher structurally paid more attention to the input of students. The story line of the teaching-learning process unfolded in a much more interactive way than in the second version. As a result, students were more involved and increasingly felt that their input mattered.

   At some points the teacher indeed implemented the interaction structures as intended (section 7.3, e.g. episode 7). However, although students acted involved they could not have experienced that their input mattered for their purpose, because the evaluation of criterion 3a showed that they did not see the point of the specific activity at the time.

C. The more complex interaction structures did not provide the teacher with sufficient guidelines as to how to direct the interaction at the level of each activity. It has to be concluded that interaction structures do not cover the full scope of *attention for student input*: students should feel that their input matters *for their purpose.*

These findings suggest that the operationalisation of the *attention for student input* characteristic needs rethinking. In retrospect, criterion 3b part-b is in fact a result of a properly worked out problem-posing approach. The design has proven to be more adequate in triggering the students' content-related motives in light of a, for them, more clear purpose (compared to the second version of the design). As a result they also feel that what they do (their input) drives the teaching-learning process and therefore *matters* in achieving their purpose.
So, designing a proper instructional version of a practice which meets the criterion of students seeing the point of every activity beforehand, in light of achieving their purpose *is* actually: paying proper attention to student input in the design of a teaching-learning process.
Interaction structures can then be considered as providing the teacher with guidelines for his/her role in the instructional version of the authentic practice, specifically for directing the interaction properly in such a teaching-learning process.

Sections 8.2.1 (first research cycle) and 8.2.2 (second research cycle) ended with 'implications'. I will not do this here, because the empirical part of this study ends with the third research cycle and there is no fourth version of the design. Section 8.3 describes what might be appropriate implications of this last research cycle.

8.2.4 The answer to the situated research question

With respect to the situated research question it can be concluded that the didactical structure of the design as presented in section 4.6 (figures 4.6A and B) was adequate. The ideas of an instructional version of an authentic practice and the implementation of interaction structures have contributed to the creation of meaningful chemistry education. These ideas (including the problems that emerged from the third evaluation) can be seen as informing a design heuristic for other designers of modules of meaningful chemistry education. Obviously, this heuristic does not provide for algorithmic decisions, which you only have to 'execute' to end up with a meaningful teaching-learning process. They rather serve as a specified heuristic scheme of design steps and criteria for decisions. In practice this means that a detailed, situation specific and intensive study of the teaching-learning process is needed to establish if, how and to what extent meaningful chemistry education has been realised in a design. This study is a first exploration of this idea and several aspects need further study, especially with respect to the specific didactical problems that emerged from the third research cycle:

- To achieve that students adopt their role in the instructional version of the practice and the connected purpose of finding out about the procedure (the 'stage setting') is to a certain extent still problematic.
- To achieve that students experience the activities as functional beforehand, in light of this purpose (which is now theirs) is to a certain extent still problematic at the level of the design of activities, especially in the more complex episodes.
- The teacher role (including interaction structures) in all this (e.g. setting the stage and directing interaction) is to a certain extent still problematic

8.3 The contribution to didactical theory

In this section I will reflect on the contribution of this study to didactical theory. Apart from the conclusions of section 8.2 on the design heuristic, this contribution lies in the conclusion that the operationalisation of the three characteristics of meaningful can, in retrospect, be seen as governed by one single principle, which I will call the principle of functional embeddedness. In retrospect this idea can be seen as directing the operationalisation of all three characteristics of meaningful throughout this study, in the first instance rather intuitively but later on more explicitly. It therefore embodies, as will be argued below, the interrelatedness between the three characteristics in the final version of the design.

Section 8.2 shows that the idea that students should experience the functionality of their activities in light of their aims was from the start implicitly present in the idea of what would be meaningful for students. Therefore they were also present in the evaluation criteria of all versions of the design. In every research cycle it became more and more explicit as a central principle, increasingly directing the operationalisation of the three characteristics of meaningful until it was finally explicitly described as such, as I will now illustrate.

1. The first version of the need-to-know criterion was:

   Do students experience what they learn by answering the sub-questions as contributions to answering the driving question?

   In other words: answering the sub-questions was expected to be *functional* for students in order to answer the driving question. Already in the first evaluation the notion was present that students should experience this functionality of the sub-questions beforehand, in light of achieving their aim to answer the driving question, instead of afterwards. However, it was not explicitly expressed in the scenario, but came to the fore at some points in the evaluation in order to point out that the design was not living up to the criterion.

2. The second version of the need-to-know criterion was:

   Is it possible to trigger the intended content-related motives at the intended moments in the teaching-learning process among students in light of their broad motive to answer the driving question?

   The idea of functionality was now captured in the aim to make the students' content-related motives the driving force of the teaching-learning process: students were to see the point of extending their knowledge in a certain direction in light of their broad motive. In other words, they were to experience the activities in which they extend their knowledge as *functional* in light of being able to answer the driving question (their general/broad motive).

3. The third version of the need-to-know criterion was:

   Are students broadly motivated by the purposes of the authentic practice to adopt their role and get involved in its instructional version as intended and do students experience at the intended moments that they need specific knowledge to take the next step in light of their purpose? Do students at the intended moments experience that such specific knowledge requires a refinement of a procedural step?

   In the third version the idea of functionality was explicitly and structurally present in the scenario, and it specifically refers to the *activities*[14] instead of the knowledge involved; in version 1: 'what they learned'; in version 2: 'extend their knowledge in a certain direction'. The students are to adopt their role and the connected purpose of the instructional version of the practice *as their own*. So students should see the point of adopting the purpose of the instructional version of the practice beforehand, as way to pursue their basic interests (e.g. good water quality is important for life in general). In other words: the purpose of the instructional version of the practice is to be *functionally embedded* in the basic interests of the students. Next, because they are to have a clear goal and intuitive knowledge of how to get there (the procedure, as a kind of advance organiser), students are to see the point of every episode to come, beforehand, in light of their aim to achieve this goal. So, the purpose of each episode, i.e.: addressing the episode question, was (or should have been) functionally *embedded* in their overall goal. In the more complex episodes, which consisted of more than one activity, the students were to see the point of *every* activity beforehand, in light of the purpose of the respective episode, i.e.: answering the respective episode question. In other words: these activities were (or should have been) functionally *embedded* in the purpose of answering the episode question.

   Note that thus creating a proper instructional version of an authentic practice, in which activities are functionally embedded in an overall purpose, matches well with the operationalisation of the *attention for student input* characteristic. For if properly designed, student will feel that what they *do* contributes to the process of achieving their goal, in other words, that their input matters in the process of achieving their purpose.
   The characteristic *attention for student input* is also operationalised by the teacher properly directing the interaction in such a teaching-learning process. Interaction structures should provide the teacher with guidelines for how s/he should do this in such a way that students feel that their input is taken seriously and is recognised as a driving force. In fact, this means that the teacher should hand over responsibility for the teaching-learning process to the students when appropriate.

[14] In line with Van Oers the functional unit of a teaching-learning process is now explicitly an *activity*, and not (for example) '*knowledge* students need' (Van Oers, 1998) .

Based on this it can be concluded that the idea of functional embeddedness should be considered as the main directing principle in designing cases of meaningful chemistry education. The three characteristics can finally be rephrased as follows.

Context
The purpose of the instructional version of the practice should be *functionally embedded* as much as possible in the values, goals or interests of students or of society at large.
Criterion:
Do students experience achieving this purpose (and taking up the connected student role) as a way to pursue their more basic interests?

Context, *need-to-know* and *attention for student input*
Activities should be *functionally embedded* as much as possible in the purpose of the instructional version of the practice (now being the students' purpose).
Criterion:
Do students experience the activities as functional beforehand, in light of achieving their purpose and do they therefore feel that their input matters?

The teacher role
a. The teacher role should contribute to inviting students to play their part in the instructional version of the practice by reducing the tension between top-down set learning goals and students' basic interests as much as possible.
b. The interaction structures should direct interaction between teacher and students at the level of each activity in such a way that students are given their part of the responsibility for the progression of the process as intended.

Criteria:
Does the teacher fulfil his/her role as intended? Does this result in
a. a diminishing of the above described tension?
b. the students feeling responsible for the progression of the learning process?

In summary, I see it as a contribution to didactical theory to make the idea of functional embeddedness an explicit directing principle in the design of a meaningful teaching-learning process in which students are motivated, experience the functionality of their activities beforehand in light of their purpose and feel that their input matters in the process of achieving their purpose. One could conclude that all I have found is that 'meaningful' can simply be replaced with 'functional embeddedness', which might raise the question of how this simple replacement can be considered a contribution to didactical theory. I will therefore explain why I do not consider this conclusion a trivial one.

First of all, it is a more directive design principle than 'meaningful' or 'need-to-know'. The analyses of the exemplary projects in chapter 2 show this. They can all be seen as operationalising 'need-to-know' differently. The idea of functional embeddedness, however, places more specific constraints on the design, which, of course, only truly directs the design when they are made in the process of

developmental research. Only then it is possible to study whether the design has sufficient didactical quality, i.e.: lives up to its aims.
Secondly, the principle of functional embeddedness not only brings certain didactical issues to the fore which are more generally recognised in science education research and partly coincide with the problematic issues that emerged from this study, but it also *connects* them:

1. **External functional embeddedness of learning goals in basic interests and goals.**
   The overall learning goals of the teaching-learning process are to be functionally embedded in the basic interests, goals and values of the students or of society at large.
   The first problematic issue emerging from the third research cycle (see 8.2.4) actually refers to this level of functional embeddedness: to achieve that students adopt their role in the instructional version of the practice and the connected purpose of finding out about the procedure (the 'stage setting'). In other words: how to solve (or reduce as much as possible) the tension between top-down set purposes and students' basic interests, values and goals
   I will elaborate at this level of functional embeddedness in section 8.3.1.
2. **Functional embeddedness of activities in the learning goals of the teaching-learning process.**
   The teaching-learning activities should be properly functionally embedded in the overall learning goals of the teaching-learning process.
   The second problem emerging from the third version of the design can be seen as referring to this level of functional embeddedness: to achieve that students experience the activities as functional beforehand, in light of this purpose (which is now theirs) at the level of the design of activities, especially in the more complex episodes.
   I will elaborate more at this level of functional embeddedness in section 8.3.2
3. **Functional embeddedness of the teaching-learning process in the curriculum**.
   The overall learning goals of the teaching-learning process are to be functionally embedded in the preceding and succeeding teaching-learning processes of the curriculum.
   This level of functional embeddedness was beyond the scope of this study. However, in section 8.3.3 I will elaborate at this level of functional embeddedness.
4. **A functional fine-tuning of the teacher's activities to the above.**
   The teacher's activities should support the students' activities on all levels of functional embeddednes.
   The third problem emerging from the third research cycle referred to this issue: the teacher role in all this, e.g. in setting the stage and directing interaction, was still problematic.
   I will discuss this issue in section 8.3.4

8.3.1 The tension between top-down set purposes and students' basic interests, values and goals

To make students see the point of a teaching-learning process can be considered as a

widely accepted problem in science, chemistry and physics education, as such a purpose is always established top-down by the designers. In the ideal situation students see the purpose of a teaching-learning process as a way to pursue their basic interests, values and goals (in other words: as a properly functionally embedded purpose). In this case their motivation will be maximal because the top-down established purpose would truly become *their* purpose. In a less perfect situation students should at least understand the relevance of the top-down set purpose and why it is thought to be important for them to learn about. In this study and in the literature two strategies seem to emerge that can be seen as addressing this level of functional embeddedness: authenticity and autonomy of choice. These strategies are to avoid or reduce the tension between the top-down established goals and the students' basic interests, values and goals.
I appreciate these strategies as worthwhile to embed functionally learning goals in interests, values and goals of students that lie outside the classroom. I want to stress, however, that these strategies at the same time may negatively influence the second kind of functional embeddedness, namely that of the activities that students are going to perform in the classroom. These activities are to be functional from their perspective to achieve their learning goals. Somehow an optimal balance between these two levels of functional embeddeness should be found, as I will discuss below.

The *first* strategy that I will discuss is the strategy of increasing the authenticity of the teaching-learning process. Different projects[15] use this strategy. In this study, for example, the use of an authentic practice, its characteristic procedure and appealing exemplary cases can be seen as a way to apply the strategy of authenticity. But this study also showed that in order to design truly functional embedded activities from the perspective of the *students*, the teaching-learning process is to be shaped by the content-related motives of the students, thus creating an *instructional version* of the authentic practice. From the perspective of functional embeddedness, the following dilemma emerges: the more the purpose of the teaching-learning process resembles an *authentic* purpose, and the more the teaching-learning activities resembles the authentic activities (because how can such an authentic purpose otherwise be achieved?), the more difficult it will be to achieve that *students* experience these activities as functionally embedded in the purpose of the instructional version of the practice and the less they will experience the activities as authentic in the sense of appreciating that their activities are going to contribute to achieving their goals.

There is a wide range of projects in which there is an attempt to reach a certain sense of authenticity among students by letting them produce something that is actually used, and therefore judged by *others*, outside the classroom: an article in a newspaper, test results that are presented to other schools or officials in relevant institutes. This strategy is often used to motivate students by giving them the feeling that their product

[15] Some of these projects (as described in chapter 1) also use authenticity to achieve that students get a more realistic and honest view of chemistry or science in society and also that students develop certain 'authentic' competencies (such as: arguing scientifically (Driver , Newton & Osborne 1999).

is actually useful to others and not only for grading, ending up in the teacher's drawer (Edelson, 1998; Becker, *et al.* 1997**,** Van Rens, 2005; Rivet *et al.*, 2000; Roth, 2003b). Several of these studies show that students are motivated by this set up: they feel that they are involved in something real (Van Rens, 2005; Rivet *et al*. 2000; Roth, 2003b). At this point the above described dilemma comes to the fore, because the more the purpose and activities resemble authentic purposes and activities, the more this will put pressure on the aim to design a teaching-learning process in which activities are truly functionally embedded in the students' overall purpose.
By applying this strategy the students will be thrilled and motivated to become involved in something real, and the above described tension between the top-down purpose and students basic interests will probably be reduced, but will they see the point of getting involved in every activity beforehand, in light of achieving their aim? In other words: will the activities of such a teaching-learning process be authentic from the *students'* perspective?
Edelson, for example, promotes authenticity in the sense that he wants students to get engaged in activities that resemble doing science as much as possible by involving them in research projects addressing their own questions. He emphasises that students should truly *want* to find out about a scientific question, with the same passion as a real scientist, because it is this passion that drives the process (Edelson, 1998). At the same time, Edelson acknowledges as a problem that students are not scientists and before they are able to formulate a proper scientific question they are to be passionate about, and to address the question in a scientific way, students are to accumulate knowledge, master tools and techniques etc.:

> Before engaging in any investigation, students will need to accumulate enough knowledge to pose well-framed questions. In the course of conducting investigations, they will need to master the tools and techniques that allow them to generate and analyse meaningful data.
>
> (Edelson, 1998, p319)

From the perspective of functional embeddedness, the problem Edelson touches on embodies exactly the above described dilemma and can be rephrased as: how to achieve that students are motivated to answer a scientific question, and how to achieve that they see the point of the knowledge, tools and techniques in light of answering their scientific question?

The *second* strategy to diminish the above-described tension that emerges from the literature is to give students autonomy of choice as much as possible, specifically about what they want to achieve and how they want to do this. The argument is that autonomy of choice will increase ownership of the students over their learning processes in science or chemistry class, and this will motivate them (Roth, 2003b). This also raises a dilemma. The more students are given autonomy of choice in deciding on their learning goals and ways to pursue them, and the more functionally embedded these goals obviously will be in the students' basic interests, the greater the risk will be that students are not able to achieve their goal through functionally embedded activities (the second level of functionally embeddedness). When given true autonomy of choice it will ask of the students that they design their own functionally embedded teaching-learning process. Especially in the case of such complex subjects as chemistry, physics or science this is probably too much to ask of them. Again, in giving students certain autonomy of choice to reduce the tension between their basic interests and goals and top-down established learning goals, somehow an optimal balance between the two levels of functional embeddeness should be found.

Finally, the most clear example of a teaching-learning process in which both authenticity and autonomy of choice are used to reduce the tension between top-down set goals and students purposes is described in Contradictions in "Learning Communities" (Roth, 2003b). Roth reports on a project in which students studied a creek in their neighborhood as increasingly full participants in a community of environmentalists (that was the goal).

From the perspective of functional embeddedness, Roth's project embodies in my opinion both the above described dilemmas. Because: how to give on the one hand students true autonomy of choice with respect to what they want to achieve and how they want to do this and, on the other hand, achieve that they experience their activities as functional in light of their purpose. And how, in light of this autonomy of choice, to achieve that students truly come to understand the environmentalist community they are to participate in and the complex environmental problems that this community deals with?
In the examples of the unfolding teaching-learning process that Roth presents, the dilemmas become visible. On one hand, Roth describes how all students felt eager to contribute to the practice of the environmentalists and were highly motivated to undertake action. Also, students were to decide themselves which questions they wanted to address, planning their own activities or, in other words what their overall purpose is (autonomy). On the other hand, Roth gives examples of how environmentalists, teachers and parents provided students with guidance and information on a 'need-to-know basis'. All the described interventions, however, seem to have more the character of the teacher (parent, environmentalist) explaining a new theory, top-down, than that it reflects a for students functional activity in light of their aim (on which they were to decide themselves).

> Thus when students in a group decide that they want to look at the relationship between the frequency of different invertebrates and stream speed (…) I also ask students how they want to measure stream speed. When they suggest measuring how long it takes for a floating object to move a certain distance I ask them whether they think that there are differences between floating objects of different materials (..) Although they may begin with wild guesses, they learn in the course of their investigations that a piece of Styrofoam may be pushed by the wind, or an orange may get stuck in shallow parts of the creek.
>
> (Roth, 2003b)

The teacher might very well have made the students see the point of this intervention *afterwards*, but they certainly did not see it beforehand, in light of their purpose to

find out about the relationship between the frequency of different invertebrates and stream speed (considering their wild guesses). In that sense I doubt that students had autonomy of choice as far as 'control over their actions and with it, their own learning' (p19).
Finally, Roth asked the students to write down what they thought they had learned. He presents the answers of one student. It is a list of seemingly arbitrary procedural and issue knowledge, but none of what is listed reflects a coherent view of what this student sees as her role in the authentic practice and her contribution. This example could be seen as exemplary of how difficult it is to achieve learning goals such as 'students truly come to understand the environmentalist community they are to participate in and the complex environmental problems that this community deals with', when students are given great extent of autonomy of choice. It then becomes very difficult for them to become involved in functionally embedded activities and achieve such desirable (by the teacher/society) learning goals.

8.3.2 The functional embeddedness of the activities in the learning goals of the teaching-learning process.

The second level of functional embeddedness is that of the activities in the learning goals of the teaching-learning process. This turned out to be difficult to establish, especially in the more complex episodes, and was still problematic in the third version of the design.
The questions are: what are the problems with this level of functional embeddedness and how might it be improved? I will discuss this with respect to the functional embeddedness of the phases and episodes in the overall purpose of the instructional version of the practice (A). In section B, I will discuss this with respect to a deeper level of functional embeddedness of the activities of the more complex episodes in the overall purpose of the respective episode.

A. Phases and episodes: the function descriptions

The evaluation of the third version of the design showed that the function descriptions of the phases and episodes were not adequate. Although the starting point of the third version of the design had been that the episodes and their respective activities should be functional for students beforehand, in light of their purposes, almost all function descriptions of the phases and the episodes do not express this idea of functionality. Some reflect learning goals, others are just descriptions of activities (e.g. episode 1: 'Students are to get a clear view of the purpose and the unfolding of the lessons and their role in it'; or episode 5: 'Students address the episode question by learning how the list of four parameters and norms to be tested in this case is established').

The students were not confronted with these descriptions, but they formed the basis for preparing the teacher to properly direct the setting of the stage and the evaluation of each of the episodes and the design of the activities, as these were to fulfill the episode functions. The mixing-up of the episode function descriptions with the teacher's/designer's learning goals and with activity descriptions obviously influenced the design of the activities and the way the teacher was prepared for his role. It may be partly due to this mixing up that at some stages in the third version of the design, especially in the more complex episodes 5 and 6, students did not experience the episodes and their activities as functional (e.g. estimating the concentration of the unknown solutions of copper sulfate in episode 6).

If the goal is to design a teaching-learning process of functionally embedded activities there are two ways that might help the designer to explicitly focus on this level of functional embeddedness. The first one is to use a distinct format for the functional descriptions. To focus on the idea of functionality, it might help if the function descriptions have the following basic structure: 'In order to achieve X, we are/I am going to do Y'. Secondly, it could help to make an explicit distinction between 'didactical functions for the teacher/ designer' on the one hand and 'didactical functions for the students' on the other hand.

The didactical functions for students should describe the students' perception of the role the episode (or phase) will play in achieving their aim, *before* getting involved in the episode (or phase). Whereas the didactical functions for the teacher/designer describe the teacher's perception of the role the episode (or phase) will play in achieving his/her aim, before getting involved in the episode (phase).

For example, for students the didactical function of an episode might have been: *In order to* be able to judge the water quality of the exemplary case (common overall purpose of phases 2 and 3) *we are going to* solve our doubts about our test results (episode 6). In turn, the common function of phases 2 and 3 is embedded in the overall purpose of the instructional version of the practice: '*In order to* find out how people monitor water quality (the overall purpose), *we are going to* judge the water quality of an exemplary case ourselves.
The didactical function for the teacher of the same episode might be: 'In order to achieve that students learn about the procedure which is followed in the exemplary case, I am going to involve students in activities in which they learn about the concepts variability and position of test results and how they should use these concepts when judging the accuracy of their test results. I will do that by using their doubts about the test results as means to functionally introduce these concepts in such a way that they see the point of every step beforehand, in light of their purpose, and feel that their input matters'.

Appendix 4 (tables A, B and C) presents an overview of the episodic function descriptions of the third version as well as proposals for a possible fourth version in which a distinction is made between didactical functions for the students and didactical functions for the teacher, all formulated according to the proposed format of a functional description.

Note that, from the perspective of functional embeddedness, it is rather obvious, if not inevitable, to make a distinction between didactical functions for the students on the one hand and for the teacher/designer on the other hand. However, as logical as it seems, I have not encountered other cases in which the same sort of distinction is made. The didactical functional phases Kortland (2001) distinguishes in the problem-posing teaching-learning process (on which the phases of the second and third version of the design were based), for example, are in retrospect didactical functions for the teacher/designer (see also chapter 4). Kortland's phase description 'Students experience a knowledge need when they apply their intuitive knowledge to an exemplary case' (phase 2) reflects what the teacher wants to achieve, not what the students are going to do, because they see the point of doing so beforehand, in light of their purpose.

Also, Leach and Scott's concept of learning demands, as discussed in chapter 3 (Leach & Scott, 2002, 2003), rather reflect the didactical functions for the teacher/designer than for the students. Leach & Scott introduce the concept of learning demands as 'challenges for the learners'. They describe the purpose of identifying learning demands as follows*:*

> The purpose of identifying learning demands is to bring into sharper focus the intellectual challenges facing learners as they address a particular aspect of school science.
>
> (Leach & Scott, 2002, p126)

When Leach & Scott speak about a sharper focus on the challenges facing the learners, this sounds more like something what the *teacher* wants students to achieve and not what *students* think they are going to do will contribute to *their* purpose. In other words: on the didactical functions for the students.

B. Activities: the basic structure of the an activity

The evaluation of the third version of the design showed that students did not experience the functionality of many of the activities. This was especially the case in the more complex episodes (5 and 6). The question is: how can this part of functional embeddedness be improved?

These episodes were more complex because they had a deeper level of functional embeddedness: their episode questions were addressed in a *series* of what were supposed to be functionally embedded activities (instead of one activity covering one episode-question). This deeper level of functional embeddedness had apparently not been explicitly paid attention to and was not designed properly.

In retrospect, the activities of episode 6, for example were designed with a learning goal in mind rather than with the motives and aims of the *students*: 'students need to learn about the concepts variability and position of test results and how they should use these concepts when judging the accuracy of their test results'. As a result, the question 'What do students need to know first in order to be able to answer their question *afterwards*?' tended to direct the design, instead of the question: 'what would students see as a logical next step in the sense of seeing that what they are going to *do* will contribute to their purpose (in this case: solving their doubts)?' My preoccupation as a researcher/designer with the former question made that the two concepts ended up being introduced separately and not functionally combined in a for students properly functionally embedded activity.

Looking at it this way, the attention of the designer should explicitly be directed at this level of functional embeddedness. Apart from making a proper distinction between didactical functions (see A), it might help to make the basic structure of an activity explicit: this structure is the same as the structure of an episode: students orientate on each activity in light of their purpose (stage setting of the activity question), they get involved in the activity (addressing the activity question) and evaluate afterwards whether their purpose is met (evaluating the activity question). This is actually in line with activity descriptions of Vygotsky (1981) and Leonte'ev (1978), who both stressed that all these three stages of an activity should be attended to. If explicit attention is paid to the structure of the activities then the nesting of these activities in the episodes can be made visible. This might serve as a tool for properly designing all levels of functional embeddedness in the teaching-learning process. Of course, paying explicit attention to such a basic structure in the design process, does not in itself provide answers as to *how exactly* the activity should be designed in order to achieve that student do experience it as a functional activity beforehand, when involved and afterwards, in light of their purpose.

Version A in the figure 8.2 shows the basic structure of an episode, involving one activity. In the more complex episodes setting the stage for the episode question might need a functionally embedded extra activity (version B). The teacher must now focus on this functional embeddedness. The same might occur for the evaluation-part. Version C has an extra 'functionally embedded activity' in this part. Of course, all kinds of variations are possible, depending on the complexity of the episode.

Figure 8.2 ***Versions of episode structures from the perspective of functional embeddedness.***

Basic structure:	*More complicated structures (e.g. episodes 5 and 6):*	
Version A	Version B	Version C
Setting the stage	**Setting the stage**	**Setting the stage**
Addressing the question	Setting the stage	**Addressing the question**
Evaluating the question	Addressing the question	**Evaluating the question**
	Evaluating the question	Setting the stage
	Addressing the question	Addressing the question
	Evaluating the question	Evaluating the question

This outlining of the more complex episodes also has implications for the more complex interaction structures (see section 8.3.4).

8.3.3 Functional embeddedness of a module in the curriculum

In this section I will discuss the third level of functional embeddedness: a module in the curriculum. The overall learning goals of the teaching-learning process are to be functionally embedded in the preceding and succeeding teaching-learning processes of the curriculum. Such a curriculum should be more coherent and relevant from the perspective of the students than the current curriculum (and from the perspective of the teacher/designer/society for that matter).

This level of functional embeddedness was beyond the scope of the actual empirical research. Therefore, I will only explore here, in very broad outlines, some of the main issues and questions that arise from such an aim.

The questions that emerge are: what would be a suitable set of authentic practices, from which instructional versions could be created and how to form a curricular story line in which these practices are properly functionally embedded in overall goals? This would mean, for one thing, that a distinction should be made between goals and didactical functions for the students and goals and didactical functions for the teacher/designer /society, leading to two descriptions of the curricular story line, each from one of the corresponding perspectives. In such a curriculum, the students should experience the relevance of the goals of particular modules in the curricular story line *beforehand* in light of their purposes. This means that the introduction of the curricular story line (or story lines) forms a special problem in the stage setting of the school subject chemistry, similar to the level of one module. Students have to adopt the overall purpose as their own purpose as much as possible.

Some first attempts were made to extrapolate the idea of involving students in ‘instructional versions of practices’ at the level of the curriculum (Bulte *et al*., 2003; Lijnse & Boersma, 2004). The attempts give an idea of how instructional practices might be functionally connected for students in a sort of curricular story line. The examples described in this section are only meant as illustrations. Instead of describing curriculum ideas in very general terms, it was chosen to illustrate ideas with real examples in order to avoid abstractions.

In the following, several general *types of practices* are identified which might be suitable for creating instructional versions and could possibly be functionally embedded in a curricular story line.
The practice that was used in this study (Judging the quality of water) can be generalised to a type of practice in which the quality of a product is evaluated. This type of practice exist in several institutions and, because they address basic needs, it seems possible to functionally embed this type of practice in more basic interests, values and goals of students or society. The practices in which the quality of a product is evaluated might therefore function as a starting point for a curricular story line.

Examples are: institutions evaluating food products, or consumer organisations judging consumer products.
The story line could start as follows. In our society we deal with all kinds of consumer products. So what does it take to evaluate whether a product is good enough?
Let us have closer look into *practices in which the quality of a product is evaluated*: for example, the quality of water, the quality of marmalade, and the quality of products for personal hygiene. These practices might be structured in a similar way as in the module of water quality. While studying this type of practice, students gradually may wonder: we evaluate this product, but how are these products made? What if a product is not good enough but we want it to meet some criteria? This can provide the students with a motive to proceed to the next set of modules about *production practices*, to be functionally embedded in the story line this way, e.g. the production of drinking water that is below the nitrate norm, while ground water contains too high an amount.

Several practices exist in which production takes place: companies producing potable water, products for personal hygiene, and all types of other consumer products. Production practices that synthesise or purify synthetic or natural components for such products can also be identified: e.g. detergents in shampoo, and super-water absorbent particles in diapers. Within these production practices the question can rise: What to do if an ingredient of the product is not available or too expensive? Then we have a motive to get involved in practices in which ingredients for products are synthesised, e.g. the detergent for shampoo. Students are provided with a motive to get involved in the process of synthesising certain ingredients.

While studying *this* type of practice, students may gradually wonder: what to do if we cannot synthesise what we wish, or if we need to investigate alternative routes? Such questions can provide students with overall content-related motives for getting involved in, and therefore functionally embed *research practices* which deal with fundamental questions. In research practices scientists search for a better understanding of our natural world, either because of pure curiosity or because a production process raised some fundamental questions to be resolved. Experts can also be involved in practices in which they provide advise to their consumers, mainly have to communicate with a target group, or to teach .

Integration of issue knowledge and procedures about all these practices can come together in a fourth type of *practice, in which professionals have advisory functions, communication functions or educational functions*. In these authentic practices, professionals need to integrate knowledge from different disciplines, several procedures and attitudes. The project could, for example, involve solving a crime, the detectives, the forensic department, asking advise from a group of experts, communication with the public etc. In such a project, other school subjects can be involved: languages, social studies, biology etc. Such projects can be a final project in the curriculum line, but also planned at intermediate stages.

In summary, the *types of practices* which can be identified and might be used to form a functionally embedded story line through content-related motives are the following:

- Practices judging the quality of a product
- Production practices
- Research practices
- Advisory, communication and teaching practices.

The presented examples here are only used for illustration. Different choices of subsequent practices and types of practices can be made, depending on different curricular ambitions, on different levels of student abilities or student motivation, and / or different societal needs. Rethinking the curriculum in terms of practices and functional embeddedness can thus provide for a different set of argumentation, which chemistry to choose for which circumstances.

To conclude, a curriculum based on adapted authentic practices directed by the principle of functional embeddedness also creates opportunities to link the subjects biology, chemistry and physics (and possibly also other subjects) in a natural way. In authentic practices there often is no sharp distinction between the subjects.

8.3.4 Functionally embedding the teacher role

In this section I will briefly discuss the teacher role. The question is: what is an adequate teacher role in a teaching-learning process which is directed by functional embeddedness? Such a teacher role has two aspects:

A. The teacher has a role as a participant in the instructional version of the practice.
The teacher might, for example, take the role of an intermediate between the authentic practice and its instructional version: students can consult the teacher as a representative of the authentic practice for information.
Some projects provide the teacher in a natural way with a more coaching role by making the students produce something that is assessed by others (and not by the teacher), such as an article about the results of a research project in a genuine journal (e.g. Van Rens, 2005).

B. The teacher has a specific role in directing interaction in the teaching-learning process.
The interaction structures were intended to provide the teacher with guidelines for how to direct interaction: collect input, summarise input, categorise input etc (see section 4.5.4, chapter 5 and appendix 1). However, if serious attention is paid to the basic structure of an activity, leading to episode structures as presented in figure 8.2, then a more complex interaction structure such as Thinking-Sharing-Exchanging needs rethinking. In that case, it would be more logical to design interaction structures which cover an *activity* instead of an episode, though the teacher should of course also direct the interaction of the overall stage setting and evaluation of the respective episode.

This leads to more detailed guidelines for the teacher role than Leach and Scott's concept of learning demands. As I explained in section 8.3.2, what Leach and Scott define as learning demands ("challenges for the learners") can be interpreted as didactical functions for the teacher/designer (Leach & Scott, 2002; 2003). Leach and Scott explain also how identifying learning demands directs the teacher role. They distinguish between an 'authoritative' and a 'dialogic' teacher role. When learning demands are at stake, for example when new ideas are to be introduced, the teacher is mostly in charge of the content-related progression and s/he should adopt a more authoritative teacher role according to Leach and Scott. The learning demands tell the teacher the kind of learning problems s/he can expect among students when (for example) introducing a new idea. S/he can use this knowledge when explaining and providing feedback to students' questions. At other moments in the teaching sequence, when no specific learning demands are at stake, the teacher can adopt a more dialogic role. At these moments, the students are more in charge of the content-related progression, for example by getting involved in peer discussions. The teacher is now more in the background, guiding rather than directing the content-related progression. In this situation the concept of learning demands prescribes a more prominent role for the students in the teaching-learning process, because the teacher now knows when s/he can adapt a more dialogical role and leave the content-related progression to the students.

No doubt this will contribute to the students' involvement in the teaching-learning process and lead to a better balance between not underestimating and not overestimating students, and the ways in which they are able to contribute to the progression of the teaching-learning process. Useful as it is, though, I think it is still too coarse grained. The principle of functional embedded activities and matching interaction structures directs the design of the teacher role differently, on a detailed level. When the activities and episodes are properly functionally embedded in the instructional version of the practice, the interaction structures provide the teacher with guidelines for structuring interaction at the level of each activity. This approach results in a teacher role that is tailored to the content-related motives of the students. As a result, the students are much more in charge of the content-related progression altogether. This is a consequence of designing the teaching-learning process primarily in such a way that *students* see the point of every step they take and feel that their input matters in every step. The teacher in his/her role should support this process.

Finally, the traditional school culture might interfere with the new approach (see also section 3.3 and 4.5.4), especially when the teacher is to hand over a part of the responsibility for the teaching-learning process to the students. This problem is also acknowledged more widely in the literature by researchers who have tried to implement new teaching approaches. I discuss this issue here because it is part of the teacher role, being the director of the process, to avoid that traditional school culture interferes with the new approach. Tiberghien, for example, explicitly argues that the teacher handing over responsibility for the teaching-learning process to the students means changing the rules of the game, and it is part of the teacher role to do this *explicitly* when the stage of the teaching-learning process is set (maybe in a class discussion). She refers to such rules as 'the didactical contract' between teacher and students (Tiberghien, 2000).

Cobb and his colleagues also emphasise the importance of the teacher making sure that certain norms that direct the teaching-learning process are made explicit (Cobb *et al*., 2001; Cobb, 2002; DiSessa and Cobb, 2004). They specifically pay attention to the importance of the teacher ensuring that the normative interaction between students and teacher (in a math class) is explicitly discussed, but they do not address it as something that the teacher should make clear in the 'stage setting' of a teaching-learning process, but rather as something that s/he should pay attention to throughout the whole teaching-learning process. Cobb and colleagues use the term socio-mathematical norms (such as 'what counts as a sound mathematical explanation'). These norms should always be negotiated between the teacher and the students according to Cobb & DiSessa. They are part of what students learn about math. DiSessa & Cobb stress the relevance of the teacher paying explicit attention to these norms and when designing mathematics education:

> (..) all designs involve assumptions about the nature of classroom interaction and discourse. The theoretical category of socio-mathematical norms orients the designer to both explicate these assumptions, thereby further specifying the design, and to examine the type of mathematical dispositions that the design appears to support.
>
> (DiSessa & Cobb, 2004, p99)

8.4 Concluding remarks on further research

In the first chapter the problematic situation of chemistry education was sketched. The aim of this study has been to contribute to the discussion and provide for new approaches to solve the problematic features of chemistry education. The idea of designing instructional versions of authentic practices, directed by the main principle of functional embeddedness together with implementing interaction structures, have proved to be potentially useful. In this sense, this research has contributed to answer the question 'How to make chemistry education more meaningful to students?'. Therefore, this research is a first attempt to explore ideas. It gives rise to a wide spectrum of new, more specific questions.

For example, with respect to the functional embedding of an instructional version of a an authentic practice in the basic interests and goals of the students: how to find a good balance between achieving this level of functional embeddedness and at the same time achieving that the activities of the teaching-learning process are properly functionally embedded too?

On the curriculum level specific questions emerge from this study, which were already tentatively explored in the previous section: what would be a suitable set of authentic practices from which an instructional version might be created and which might form a functionally embedded curricular story line? How would such a curricular story line be described in terms of didactical functions and goals for the students on the one hand and didactical functions and goals of the teacher/designer/society on the other hand? A project on this issue will also start in the near future (Lijnse & Boersma, 2004).

The teacher role in the instructional version of the practice needs further attention. At this moment a research project did start that focuses explicitly on the teacher role in teaching-learning process such as the teaching-learning process in the water quality design. The focus of this study is on what it means for a teacher to hand over responsibility to the students in a highly structured functionally embedded teaching-learning process (Mathijssen *et al*. to be published).

Thinking about contexts and concepts in terms of practices and activities in relation to concepts is also explored in the discussions on the chemistry education reform in the Netherlands (chapter 1). These discussions show that thinking in practices and actions helps to make a clear and argued distinction between what would be a suitable curriculum for two different levels of upper secondary education, HAVO and VWO[16], by establishing what would be *typically relevant* practices for HAVO students and what for VWO (Pilot & Bulte, 2005a).

At an international level, related problems and results of research and development are discussed, bringing together expertise on context based chemistry education from different countries in a symposium that focuses on teaching-learning processes in meaningful chemistry education (Pilot & Bulte, 2005b).

It shows that this research not only resulted in some answers on the problems of chemistry education, but also gave rise to new challenging questions and potential solutions.

[16] VWO (‘Voorbereidend Wetenschappelijk Onderwijs’) is pre-university education; HAVO (‘Hoger Algemeen Vormend Onderwijs’) is senior general secondary education. The latter type prepares students for higher vocational education (“Hoger Beroeps Onderwijs’; HBO), such as mechanical engineering or teaching at primary schools.

References

Acampo, J. (1997). *Teaching electrochemical cells.* A study on teachers' conceptions and teaching problems in secondary education. Utrecht: CD- ß Press.

Banet, E., & Núñez, F. (1997). Teaching and learning about human nutrition: A constructivist approach. *International Journal of Science Education 19*(10), 1169-1194.

Barker, V., & Millar, R. (1996). Differences between Salters' and traditional A-level chemistry students' understanding of basic chemical ideas. Research paper 96/05. York: department of Educational Studies, University of York, UK.

Barker, V. (1999). Students' reasoning about chemical reactions: what changes occur during a context-based post-16 chemistry course?. *International Journal of Science Education 21*(6), 645-665.

Becker, M., Congalton, R., Budd, R., & Fried, A. (1997). A GLOBE collaboration to develop land cover data collection and analysis protocols. *Journal of Science Education and Technology 7*(1), 85-96.

Bencze, L., & Hodson, D. (1999). Changing practice by changing practice: toward more authentic science and science curriculum development. *Journal of Research in Science Teaching, 36*(5), 521-539.

Bennett, J., Hogarth, S., & Lubben, F. (2003). A systematic review of the effects of context-based and Science-Technology-Society (STS) approaches in the teaching of secondary science. EPPI-centre, London.

Bennett, J., & Holman, J. (2002). Context-based approaches to the teaching of chemistry: What are they and what are their effects?. In: J.K. Gilbert (Ed.), *Chemical Education: Towards Research –based Practice.* Dordrecht, The Netherlands: Kluwer Academic Press: 165-184.

Bennett, J., Holman, J., Lubben, F., Nicolson, P., & Prior, C. (2002). Science in context: The Salters' Approach. Paper presented at the Second international IPN-YSEG Symposium: Context-based Curricula, October 10-13, 2002, Kiel, Germany.

Brown, A.L. (1992) Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *The Journal of Learning Sciences, 2 (2),* 141-178.

Bruner, J.S. (1962). *On Knowing*. Cambridge: Havard University Press.

Bulte, A.M.W., Carelsen, F., Davids, W., Morelis, H., Pilot, A., Velthorst, A., & De Vos, W. (1999). Dilemma's in de schoolscheikunde [dilemmas in school chemistry]. *NVOX, 24*(6), 289-291.

Bulte, A.M.W., Klaassen, C.W.J.M., Westbroek, H.B., Stolk, M.J., Prins, G.T., Genseberger, R., De Jong, O., & Pilot, A. (2002, October). Modules for a new Chemistry Curriculum, Research on a Meaningful relation between Contexts and Concepts. Paper presented at the Second international IPN-YSEG Symposium: Context-based Curricula, Kiel, Germany.

Burns, R.B. (2000) *Introduction to research methods.* London: Sage Publications Inc.

Campbell, B., Lazonby, J., Millar, R., Nicolson, P., Ramsden, J., and Waddington, D. (1994). Science: The Salters' approach: A case study of the process of large scale development. *Science Education, 78*(5), 415-447.

Carr, D. (1998). The art of asking questions in the teaching of science. *School Science Review, 79*(289), 47-50.

ChemCom (2002, October). ChemCom Contribution to the second International IPN-YSEG Symposium. Paper presented at the Second international IPN-YSEG Symposium: Context-based Curricula, 2002, Kiel, Germany.
Cobb, P. (2002). Modelling, symbolizing, and tool use in statistical data analysis. In K. Gravemeijer, R. Lehrer, B. van Oers and L. Verschaffel (Eds.), *Symbolizing, Modelling and Tool Use in Mathematics Education* (pp.171-195). Dordrecht, The Netherlands: Kluwer Academic Publishers.
De Jong, O. (1990). *Rekenen aan reacties*. Utrecht: CD- ß Press.
The Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.
De Vos, W. (1985). *Corpusculum Delicti.* Utrecht: CD- ß Press.
De Vos, W., Bulte, A.M.W., & Pilot, A. (2002). Chemistry curricula for general education: Analysis and elements of a design. In J.K. Gilbert (Ed.), *Chemical Education: Towards Research –based Practice* (pp.101-124). Dordrecht, The Netherlands: Kluwer Academic Press.
De Vos, W., & Pilot, A. (2001). Acids and bases in layers. *Journal of Chemical education*, 78, April: 494-499.
De Vos, W., Van Berkel, B., & Verdonk, A. H. (1993). Het isolement van de schoolscheikunde [The isolated position of school chemistry]. *Tijdschrift voor Didactiek der bètawetenschappen, 11 (2),* 98-109.
De Vos, W., & Verdonk, A.H. (1990). Een vakstructuur van het schoolvak scheikunde [a structure of school chemistry]. *Tijdschrift voor Didaktiek der ß-wetenschappen*, 9, 97-108.
DiSessa, A.A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *The Journal of the Learning Sciences, 13*(1), 77-103.
Driessen, H.P.W., & Meinema, H.A. (2003). *Chemistry between Context and Concept. Designing for Renewal.* Enschede, SLO, Netherlands National Institute for Curriculum Development.
Driver, R., Guesne, E., & Tiberghien A. (Eds.), (1985). *Children's Ideas in Science.* Milton Keynes, Open University Press.
Duit, R. (1999). Conceptual change approaches in science education. In W.Schnotz, S.Vosniadou, & M.Carretero (Eds.), *New perspectives on conceptual change* (pp.263-282). The Netherlands, Amsterdam: Elsevier Science.
Edelson, D.C. (1998). Realising authentic science learning through the adaption of scientific practice. In B.J. Fraser, & K.G. Tobin (Eds.), *International Handbook of Science Education, part I* (pp.317-331). Dordrecht, The Netherlands: Kluwer Academic Publishers.
Eilks, I., Parchmann, I., Grasel, C., & Ralle, B. (2004). Changing teachers' attitudes and professional skills by involving teachers into projects of curriculum innovation in Germany. In B.Ralle, & I.Eilks (Eds.), *Quality in practice-oriented research in science education* (pp.29-40). Shaker, Aachen, Germany.
Eijkelhof, H.M.C., & Kortland, J. (1988). Broadening the aims of physics education. In P.J. Fensham (Ed.), *Development and Dilemma's in Science Education* (pp. 282-305). London: Palmer Press.

Eijkelhof, H.M.C., & Wierstra, R.F.A. (1986). Effecten van straling en risicoafweging: een onderzoek naar kennis en attituden van leerlingen van 5-HAVO. *Tijdschrift voor didaktiek van de β-wetenschappen, 1*

Fensham, P.J. (1999, August). Science content as problematic; issues for research. Paper presented at the annual ESERA conference, Kiel, Germany

Geddis, A.N. (1991). Improving the quality of science classroom discourse on controversial Issues. *Science Education 75*(2), 169-183.

Goedhart, M. J. (1990). *Meten: normen en waarden.* Utrecht: CD- ß Press.

Gräber, W., & Bolte, C. (Eds). (1997). *Scientific Literacy. An International Symposium.* Kiel: Institut für die Pädagogik der Naturwissenschaften.

Gravemeijer, K.P.E. (1994). Developing Realistic Mathematics Education. The Netherlands, Utrecht: CD- ß Press.

Grice, H.P. (1975). “Logic and conversation”. In P.Cole & J.L.Morgan, (Eds.), *Syntax and Semantics: Vol.3, Speech Acts,* New York: Academic Press.

Hart, C., Mulhall, P., Berry, A., Loughran, J., & Gunstone, R. (2000). What is the purpose of this experiment? Or can students learn something from doing experiments?. *Journal of Research in Science Teaching, 37*(7), 655-675.

Haselager, N. (2001). *Basisboek milieumetingen.* De achtergronden van onderzoek naar chemische verontreinigingen bij lokale milieuproblemen. Chemiewinkel Utrecht.

Howland, D., Larsen, & Becker, M. (2002). GLOBE- The science behind launching an international environmental education program. *Journal of Science Education and Technology*, *11*(3), 199-210.

Jansen, K., Kerkstra, A., Meijer, M., Pilot, A., & Slob, M. (2002a). Superslurpers, nu nog beter!. Handleiding voor leerlingen. Utrecht: Universiteit Utrecht, sectie Chemiedidactiek. 30 p. (document #02-01).

Jansen, K., Kerkstra, A., Meijer, M., & Pilot, A. (2002b). Superslurpers, nu nog beter!. Docentenhandleiding. Utrecht: Universiteit Utrecht, sectie Chemiedidactiek. 37 p. (document #02-03).

Jansen, K., Meijer, M., Slob, M., Kerkstra, A. & Pilot, A. (2002c). Water, verrassend gewoon. Handleiding voor leerlingen. Utrecht: Universiteit Utrecht, sectie Chemiedidactiek. 37 p. (document #02-04).

Jansen, K., Meijer, M., Slob, M., Kerkstra, A., & Pilot, A. (2002d). Water, verrassend gewoon. Docentenhandleiding. Utrecht: Universiteit Utrecht, sectie Chemiedidactiek. 39 p. (document #02-06).

Joling, E. (1993). *Chemie in gesprek: een didactisch onderzoek naar een derde wijze waarop het zakelijk gesprek in chemiebeoefening en onderwijs voortgezet kan worden.* Amsterdam: Universiteit van Amsterdam.

Joling, E., Van Lierop, A., Van Soest, W., Kaper, W., Ten Voorde, H.H., Vos, W., Mellink, E., Snel, B., & Timmer, J. (1988). *Chemie MAVO. Onderzoek naar het functioneren van een leergang scheikunde.* Amsterdam: Stichting Centrum Onderwijsonderzoek van de Universiteit van Amsterdam.

Joyce, B., & Showers, B. (1980). Improving inservice training: The messages of research. *Educational Leadership 37*(5), 379-385.

Kaper, W. (1997). *Thermodynamica leren onderwijzen.* Amsterdam: Thesis publishers Amsterdam.

Kaper, W., & Ten Voorde, H.H. (1991). Problemen in de begripsontwikkeling in relatie tot de aanpak van docent en studie-boek schrijver. *Tijdschrift voor Didactiek der wetenschappen, 9*(1), 3-26.

Kjersdam, F., & Enemark, S. (1994). *The Aalborg Experiment.* Aalborg, Denmark Aalborg University Press.

Klaassen, C.W.J.M. (1995). *A problem -posing approach to teaching the topic of radioactivity.* Utrecht: CD- ß Press.

Knippels, M.C.P.J. (2002). *Coping with the abstract and complex nature of genetics in biology education. The Yo-yo learning and teaching strategy.* Utrecht: CD- ß Press.

Kortland, J. (2001). *A problem –posing Approach to Teaching Decision Making about the Waste Issue.* Utrecht: CD- ß Press.

Kortland, J. (2002, October). PLON: dead or alive? The Dutch physics curriculum development project in 1972-1986 and beyond. Paper presented at the Second international IPN-YSEG Symposium: Context-based Curricula, Kiel, Germany

Kortland, J. (2004). Personal communication.

Kortland, J., & Van der Loo, F.A. (1986). Natuurkunde als gereedschap bij besluitvorming. In H.M.C. Eijkelhof, E. Holl, B. Plupessy, A.E. Van der Valk, P.A.J. Verhagen, & R.F.A. Wiertstra (Eds.), *Op weg naar vernieuwing in het natuurkunde onderwijs*. 's-Gravenhage. Stichting voor Onderzoek van het Onderwijs S.V.O.

Kraijcik, J., Blumenfeld, P., Marx, R., & Soloway, E. (1999). Instructional, curricular, and technological supports for inquiry in science classrooms. In J.Minstell, & E.V.Zee (Eds.), *Inquiry into Inquiry: Science Learning and Teaching.* Washington, D.C.: American Association for the Advancement of Science Press.

Kresner, M., Hofstein, A., & Ben-Zvi, R. (1997). The development and implementation of two industrial chemistry case studies for the Israeli high school chemistry curriculum. *International Journal of science Education, 19*(5), 565-576.

Laugksch, R.C. (2000). Scientific Literacy: A Conceptual Overview. *Science education*, *84,* 71-94

Leach, J., & Scott, P. (2002). Designing and evaluating science teaching sequences: an approach drawing upon the concept of learning demand and a social constructivist perspective on learning. *Studies in Science education, 38,*115-142.

Leach, J., & Scott, P. (2003). Individual and sociocultural views of learning in science education. *Science & Education, 12*, 91-113.

Leach, J., Hind, A., Lewis, J., & Scott, P. (2002). EPSE Project 2: Designing and evaluating short teaching sequences, informed by research evidence. *School Science Review, 84,* 25-28.

Lemke, J.L. (1990). *Talking Science.* Norwood, NJ: Ablex.

Leonte'ev, A.N. (1978). *Activity, Consiousness and Personality*. Prentice-Hall, Inc., Englewood Cliffs, New Yersey.

Lijnse, P.L. (1986). Onderzoekt alle dingen en behoudt het goede. In H.M.C. Eijkelhof, E. Holl, B. Plupessy, A.E. van der Valk, P.A.J. Verhagen, & R.F.A.

Wiertstra (Eds.). *Op weg naar vernieuwing in het natuurkunde onderwijs.* 's-Gravenhage. Stichting voor Onderzoek van het Onderwijs S.V.O.

Lijnse, P. L. (1995). 'Developmental Research' as a way to an empirically based didactical structure of science. *Science Education, 79*(2), 189-199.

Lijnse, P.L. (2001). Didactics of Science: the forgotten Dimension in Science Education?. In R.Millar, J. Leach, & J.Osborne (Eds.), *Improving Science Education, the Contribution of Research*. Open University Press.

Lijnse, P.L. (2003). Developmental research: its aims, methods and outcomes. In D. Kernl, (Ed.), Proceedings of the 6th ESERA PhD summerschool. University of Ljubljana.

Lijnse, P.L., & Boersma, K. Th. (2004). Naar een samenhangend, op handelingspraktijken gebaseerd curriculum voor de natuurwetenschappen. Voorstel voor de ProgrammaRaad Onderzoek van Onderwijs. Utrecht: Universiteit Utrecht, Centrum voor Didactiek van Wiskunde en Natuurwetenschappen.

Lijnse, P.L., & Klaassen, C.W.J.M. (2004). Didactical structures as an outcome of research on teaching-learning sequences?. *International Journal of Science Education, 26*(5)*,* 537-554.

Lijnse, P.L., Kortland, J., Eijkelhof, H.M.C., Van Genderen, D., & Hooymayers, H.P. (1990). A thematic physics curriculum: a balance between contradictory curriculum forces. *Science Education, 74,* 95-103.

Lubeck, H., Speelman, J., & Westerbeek, C. (1996). *Grondig bekeken. Het verhaal*. Instituut voor Leerplanontwikkeling (SLO), Enschede.

Marbach-Ad, G., & Sokolove, P.G. (2000). Can undergraduate biology students learn to ask higher level questions? *Journal of Research in Science Teaching, 37,* 854-870.

Méheut, M., & Chomat, A. (1997). The bounds of children atomism: an attempt to make children built up a particulate model of matter. In P.L.Lijnse, P.Licht, W. de Vos, & A.J. Waarlo (Eds.), *Relating Macroscopic Phenomena to Microscopic Particles.* Utrecht: CD- ß Press.

Méheut, M., & Psillos, D. (2004). Editorial- Teaching-learning sequences: aims and tools for science education research. *International Journal of Science Education, 26*(5), 515-537

Millar, R., & Osborne, J. (1999). *Beyond 2000: science education for the future.* London: King's College.

Nentwig, P., Parchmann, I., Demuth, R., Gräsel, C., & Ralle, B. (2002, October). Chemie im Kontext- From situated learning in relevant contexts to a systematic development of basic chemical concepts. Paper presented at the Second international IPN-YSEG Symposium: Context-based Curricula, Kiel, Germany.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84,* 287-312.

Parchmann, I., Demuth, R., Ralle, B., Huntemann, H., & Paschmann, A. (2001). Chemie im Kontext-Begrundung und realisierung eines lernes in sinnstiftenden Kontexten. PdN-Ch. 50/1:2ff.

Patronis, T. (1999). Students' argumentation in decision-making on a socio-scientific issue: implications for teaching. *International Journal for Science Education, 21*(7), 745-754.

Pieren, L.O.F., Scheffers-Sap, M., Scholte, H., Vroemen, E., & Davids, W. (1995). *Chemie 3 havo/vwo.* Wolters-Noordhoff, Groningen.

Pilot, A., & Bulte, A.M.W. (2005a). De context-concept benadering bestaat niet…, over structuurkenmerken van verschillende benaderingen. Intern Rapport. Utrecht: Universiteit Utrecht, Centrum voor Didactiek van Wiskunde en Natuurwetenschappen.

Pilot, A., & Bulte A.M.W. (2005b, August). Context based chemistry education: Theme & overview of the symposium. Barcelona, ESERA-conference.

Powel, P.C., & Weenk, G.W.H. (2003). Project-led enginering education. Utrecht: Lemma.

Ramsden, J.M. (1994). Context- and activity based science: some teachers' views of the effect on pupils. *School Science Review, 75*(2), 7-14.

Ramsden, J.M., (1997). How does a context-based approach influence understanding of key chemical ideas at 16+?. *International Journal of Science Education, 19*(6), 697-710.

Rennie, L.J., Goodrum, D., & Hackling, M. (2001). Science teaching and learning in Australian school: results of a national study. *Research in Science Education, 31,* 455-498.

Rivet, A. (2003). Contextualizing Instruction and Student Learning in Middle School Project Based Science Classrooms. UmiFilms, nr: 3096184.

Rivet, A., Singer, J., Schneider, R., Kraijick, J., and Marx, R. (2000). The Evolution of Water: Designing and Developing Effective Curricula. Paper presented at the annual Meeting of the National Association for Research in Science Teaching, New Orleans, USA.

Roberts, D.A. (1982). Developing the concept of 'Curriculum Emphases' in science education. *Science education, 66*(2), 243-260.

Roberts, D.A. (1988). What counts as science education? In P.J. Fensham (Ed.), *Development and Dilemma's in Science Education* (pp.27-54) London: Palmer Press.

Roberts, D.A. (1995). Junior high school science transformed: analysing a science curriculum policy change. *International Journal of Science Education, 17*(4), 493-504.

Rop, J. (2003). Spontaneous inquiry questions in high school chemistry classrooms: perceptions of a group of motivated learners. *International Journal of Science Education, 25*(1), 13-33.

Roth, W.M. (1993). The development of science process skills in authentic contexts. *Journal of Research in Science Teaching, 30*(2), 127-152.

Roth, W.M. (1998). *Designing communities*. Kluwer Academic Publishers, Dordrecht.

Roth, W.M. (2003a). Video as Tool for Reflecting on Practice: theoretical Perspectives. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia, USA.

Roth, W.M. (2003b, August). Contradictions in "Learning Communities". Paper presented at the annual ESERA conference, Noordwijk, The Netherlands.

Shamos, A.H. (1995). *The myth of scientific literacy.* Rutgers University Press, New Brunswick, N.J.

Schwab, J.J. (1962). The teaching of science as inquiry. In J.J. Schwab, & P.F. Brandwein (Eds.), *The teaching of science.* Cambridge MA: Harvard University Press.

Schwartz, A.T. (1999). Creating a context for chemistry. *Science and Education, 8,* 605-618.

Shavelson, R.J., Phillips, D.C., Towne, L., & Feuer, M.J. (2003). On the science of education design studies. *Educational Researcher, 32*(1), 25-28.

Sutman, F., & Bruce, M. (1992). Chemistry in the community- ChemCom: a five-year evaluation. *Journal of Chemical Education, 69*(7), 564-567.

Sutton, C. (1996). The scientific model as a form of speech. In G.Welford, J.Osborne & P.Scott (Eds.), *Research is Science and Education in Europe* (pp.143-152). The Falmer Press, London.

Tiberghien, A. (1997). Learning and teaching: Differentiation and relation. *Research in Science Education, 27*(3), 359-382.

Tiberghien, A. (2000). Designing teaching situations in the secondary school. In R. Millar, J. Leach, & J. Osborne (Eds.), *Improving Science Education-The Contribution of Research* (pp.27-47). Buckingham: Open University Press.

Van Aalsvoort, J.M. (2000). *Chemistry in products.* Utrecht: CD- ß Press.

Van Aalsvoort, J.M. (2005). Personal communication.

Van Berkel, B., De Vos, W., Verdonk, A.H., & Pilot, A. (2000). Normal science education and its dangers: The case of school chemistry. *Science and education, 9,* 123-159.

Van Driel, J.H. (1990). *Betrokken bij evenwicht.* Utrecht: CD- ß Press.

Van Hoeve-Brouwer, G.M. (1996). *Teaching structures in chemistry.* An educational structure for chemical bonding. Utrecht: CD- ß Press.

Van Keulen, H. (1995). *Making sense*. Simulation-of-research in Organic Chemistry Education. Utrecht: CD- ß Press.

Van Koten, G., De Kruijff, B., Driessen, H.P.W., Kerkstra, A., & Meinema, H.A. (2002). *Bouwen aan scheikunde, blauwdruk voor een aanzet tot vernieuwing van het vak scheikunde in de Tweede Fase van HAVO en VWO* [Constructing Chemistry, a blueprint for reconstructing the chemistry curriculum for upper secondary education]. Enschede, The Netherelands: SLO.

Van Oers, B. (1998). From context to contextualizing. *Learning and Instruction,* 8, 473-488.

Van Rens, .E.M.M. (2005). *Effectief scheikundeonderwijs voor 'leren onderzoeken' in de tweede fase van het vwo.* Een chemie van willen, weten en kunnen. Onderwijscentrum VU, Vrije Universiteit Amsterdam.

Verhoeff, R.P. (2003). *Towards systems thinking in cell biology education.* CD-ß Press.

Viennot, L., & Rainson, S. (1999). Design and evaluation of a research based teaching sequence: the superposition of electric fields. *International Journal of Science Education, 21*(1), 1-16.

Vollebregt, M.J. (1998). *A problem-posing approach to teaching an initial particle model*. CD-ß Press.

Vygotsky, L.S. (1981). The concept of activity in soviet psychology. In J.V. Wertsch (Ed.), The Genesis of Higher Mental functions. M.E. Sharpe, Inc., Publisher, Armonk, New York.

Wandersee, J.H., Mintzes, J.J., & Novak, J.D. (1994). Research on alternative conceptions in science. In D.L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp.177-210). New York: Macmillan.

Watts, M., Alsop, S., Gould, G., & Walsh, A. (1997). Prompting teachers' constructive reflection: pupils' questions as critical incidents. *International Journal of Science Education, 19*(9), 1025-1037.

Wiersma, R.F.A. (1984). A study on classroom environment and on cognitive and affective outcomes of the PLON curriculum. *Studies in Educational Evaluation.* p.273-282.

Zee, E.H., Iwasyk, M., Kurose, A., Simpson, D., & Wild, J. (2001). Student and teacher questioning during conversations about science. *Journal of Research in Science Teaching, 38*(2), 159-190.

Appendices

Appendix 1 Interaction structures

In this appendix a more detailed description is given of the interaction structures that were used in the design of the scenario (see Chapter 5) and the evaluation of the lessons (Chapter 7). The idea of interaction structures was inspired by the work of Jay Lemke (1990), who analysed interaction between students and teachers in traditional science lessons (see section 4.6).
I will first present a general description of the interaction structures and their common features. Next, I will present a detailed description of each of the five types of interaction structures that were designed for this specific study.

General features of interaction structures

The different interaction structures all have a comparable basic structure of the following parts:

Preparation of a context -part
Question-part
Evaluation/elaboration-part

Each of these parts consists of one or more teaching-learning activities, depending on the complexity of the part.
The interaction structures differ in how the different steps in the basic structure are interpreted in more detail. The *way* a context for a question is prepared, and the *way* the answers are elaborated upon and are evaluated is very much, if not completely directed by the teacher. The teacher thus raises expectations with the students about what they can expect of the further progress of the episode, but also about the type of output, and about who needs to answer. In the description of different possible interaction structures (below) the expectations these structures should raise among students are interpreted as 'messages to the students'.

Every interaction structure starts with the preparation of a context for the central question of the episode. This stage setting can vary from a short introduction to a series of 'preparing questions'. But: when setting the stage for a question the teacher *always* makes sure that the students bridge the content-related outcomes of the previous learning activity with this successive learning activity. By doing this, the teacher makes sure that the question will logically emerge from the previous learning activity. The important thing is that students must experience this logic. The teacher has to verify al the time if this is the case. S/he can do that by letting students formulate 'the bridge' as well as the question themselves first, in their own words. Or s/he can pose the question and ask the students to explicit the logic of the question (the bridge). Of course, the learning activities should be designed in such a way that explicit attention has been paid to the logic of the successive questions.

Furthermore, the type of evaluation activities depends on which interaction structure is used. In a 'cross discussion' or a 'collect input' interaction structure the teacher is not supposed to contribute to the output of the activity with respect to the *content*. In other words: students can do that themselves. The evaluation merely has the form of formulating a summary of uttered opinions or answers. However, the evaluation part in the 'thinking-sharing-exchanging' interaction structure does need a content-related input of the teacher. When this interaction structure is used it is expected that students can't come up with a certain level of output by themselves (most of the time of this involves a reflective, more abstract kind of output). The teacher has the job on one hand to help the students to reach through their output to a 'higher level', on the other hand to let students experience that their output *really* contributed to this higher level.
Next, the five interaction structures that were designed for this study are described. In 'Messages to the students' it is described from the students' perspectives how the respective interaction structure is expected to match with the intended content-related progression.

Description of the interaction structures

1. Collect input

Messages to the students

There is not 'one right answer', but it is a matter of opinion (so you don't all need to agree). You have to express your opinion with a certain clarity and depth (depending on the structure of the desired content-related outcome). Not *all* of you need to express your view, but in the end you will have a few striking examples of the intended answers. These few striking examples form the basis of the next question (so your input *does* matter)

Description

Preparing context of a question: The teacher always makes sure that the students bridge the content-related outcomes of the previous learning activity with this successive learning activity. By doing this, the teacher makes sure that the question will logically emerge from the previous learning activity. The important thing is that students must experience this logic. The teacher has to verify al the time if this is the case. S/he can do that by letting students formulate 'the bridge' as well as the question themselves first, in their own words. Or s/he can pose the question and ask the students to explicit the logic of the question (the bridge). Of course, the learning activities should be designed in such a way that explicit attention has been paid to the logic of the successive questions.

Question	The teacher asks for student's opinions, not specifically addressing one student. Not *all* the students in class need to speak out. If no one answers s/he can try to provoke students to enter the discussion by expressing disputable statements, or to add more 'interesting' (for students) aspects of the theme.
Answer	Students answer
Evaluation	The teacher (having the intended type of answer in mind) evaluates the answer by addressing questions like: Does the answer meet with the characteristics of the intended answer? Was it indeed an opinion the student expressed? Was it indeed the student's own opinion? Was the complexity of the answer sufficient?
Elaboration	The teacher can elaborate as a way to evaluate: 1. Asking the student for example: 'Is this your opinion/what you think?" 2. Let the student deepen his/ her argumentation /express his/her opinion more clearly by saying things like: 'Do you mean this…., Can you explain why?, Could you be more specific?' 3. Involve more students in the discussion by saying things like: (to the class) 'do you feel the same?','does someone feel differently?'
Completion/ bridge to prepare the context of the next question	The teacher summarises the content relate outcomes some short and proceeds by linking these outcomes with remarks to the preparation of the context of the next question.

Variations

1A Summarise input

The only difference with 'collect input' is that the teacher not only collects student input, but he also summarises the input in such a way that students feel that their input contributed to the summary of the teacher (see for example episode 2, section 5.3.2).

1 B Categorise input

The only difference with 'collect input' is that the teacher not only collects student input, but he also categorizes the students input into different, for students obvious, categories (see for example episode 4, section 5.3.4).

2. Thinking together-exchanging

Messages to the students

There is a best answer/strategy, which isn't easy to deduct. Or, there are different well-founded opinions possible about a certain issue, which are not easy to formulate. It is necessary that all of you think about it and try to formulate a best answer/strategy or a well-founded opinion. Not all of you need to express his or her individual answer/strategy/opinion in the classroom in the exchange part, but everybody can be asked to do so. This individual answer part and 'whole class' exchange part will be sufficient. The aim is to produce a best answer/ strategy and reasons why it is not perfect, or a few striking examples of well-founded opinions with a certain clarity and depth (depending on the structure of the desired content-related outcome), which are constructed of the outcomes of the exchange of the students answers (so your input *does* matter). The best answer/strategy or the list of striking examples forms the basis of the next question, so everybody's input *does* matter for the progression of the teaching learning strategy.

Description

Preparing context of a question:	The teacher always makes sure that the students bridge the content-related outcomes of the previous learning activity with this successive learning activity. By doing this, the teacher makes sure that the question will logically emerge from the previous learning activity. The important thing is that students must experience this logic. The teacher has to verify al the time if this is the case. S/he can do that by letting students formulate 'the bridge' as well as the question themselves first, in their own words. Or s/he can pose the question and ask the students to explicit the logic of the question (the bridge). Of course, the learning activities should be designed in such a way that explicit attention has been paid to the logic of the successive questions
Question-individual	Teacher poses/repeats the question and stresses that several answers / possible strategies / different opinions are possible. Furthermore, he /she indicates that it isn't easy to get to the 'best answer/strategy[17]' or to a 'well founded opinion' or that it is important that everybody thinks about this and that, therefore, 'everyone has to think about it first'. *All* the students in class need to speak out by writing down their answer/strategy/opinion, so that the class can make use of everybody's specific input. The teacher explains to the students that they will use their individual answers/strategies to produce the best 'class answer/strategy' they can come up with in the 'whole class exchanging part' or the best 'striking examples of a well founded opinion'. To underline the complexity of task, the teacher explicitly asks each student to elaborate also on what they found difficult about the question and are uncertain about.

[17] Of course in some situations several 'good strategies' are possible.

Answer	Students produce individual written answers/ strategies/opinions
Exchanging	The teacher asks for student's answers/strategies/opinions and uncertainties, specifically addressing one student at the time. Not *all* the students in class need to speak out, but the teacher can address everyone. Students are also invited to give additional comments/input.
Evaluation	The teacher evaluates the answers/strategies/opinions and uncertainties by (having the intended answer in mind) letting students comparing them to each other and let them respond to the expressed uncertainties. The teacher can structure the discussion with the question: what would be the best answers/strategy, or, what are striking examples of well-founded opinions?
Elaboration	If an answer/strategy/opinion is not clear, the students or teacher can ask the student concerned to further clarify his/her answer/strategy/opinion (For example: 'What do you mean by..? Why do you think..?'). If an answer/strategy is not complex enough, the students / teacher can ask for further information. If an opinion is not well-founded the students/ teacher can ask the student involved for more solid arguments.
Completion/ bridge to prepare the context for the next question	The teacher summarises the content-related outcomes embodied in the list of striking examples of well-founded opinions or one best class answer/strategy. These are, of course, for an important part produced by the students. The teacher proceeds by linking these outcomes with some short remarks to the preparation of the context of the next question

3. Thinking-sharing-exchanging

Messages to the students

The purpose is to formulate 'one best answer/strategy' (which isn't perfect), which isn't easy. It is necessary that all of you are involved in the discussion and contribute to the 'best answer/strategy' with a certain clarity and depth (depending on the structure of the desired content-related outcome). The class-answer is constructed of the outcomes of the exchange of the group answers of your group (to which each of you has contributed in a sense, so everybody's input *does* matter for the formulation of the 'best answer'). The class answer forms the basis of the next question, so everybody's input *does* matter for the progression of the teaching learning strategy.

Description

Preparing context of a question:	The teacher always makes sure that the students bridge the content-related outcomes of the previous learning activity with this successive learning activity. By doing this, the teacher

	makes sure that the question will logically emerge from the previous learning activity. The important thing is that students must experience this logic. The teacher has to verify al the time if this is the case. S/he can do that by letting students formulate 'the bridge' as well as the question themselves first, in their own words. Or s/he can pose the question and ask the students to explicit the logic of the question (the bridge). Of course, the learning activities should be designed in such a way that explicit attention has been paid to the logic of the successive questions
Question-individual	Teacher poses/repeats the question and indicates that several answers / possible strategies are possible. Furthermore, he /she indicates that it isn't easy to get to the 'best answer/strategy' and that, therefore, 'we have to think about it in different stages'. First, *all* the students in class need to speak out by writing down their own list of key features of the 'best answer/strategy' and their personal uncertainties, so that the group can make use of everybody's specific input. The teacher explains to the students that they will use their individual lists to produce the best 'group list of key features' they can come up with in the 'sharing part'. He /she explains that these group-answers/strategies will be used in a 'whole class exchange part' to finally produce a best answer/strategy of the class. To underline the complexity of task, the teacher explicitly asks each group to also elaborate on what they found difficult about the question and are uncertain about.
Answer	Students produce individual written lists of key features and uncertainties.
Sharing	The students are asked to compare their individual lists in groups and express their uncertainties. The purpose is to use each input to construct the best possible list of key features and diminish as much as possible uncertainties.
Evaluation	The group members elaborate on each individual list, thereby addressing differences and similarities. And by doing that, the group tries to identify key features of the problem.
Elaboration	If a student's input is not clear, group members can ask for clarification. If a student's input is superficial, group members can ask for additional information.
Negotiation	Group members negotiate about possible key features of the best answer/strategy. Every group member must agree.
Conclusion	The group writes down a most complete list of key features they can come up with, and the aspects they are uncertain about.

Exchanging		The teacher assigns a spokesman for each group and asks each spokesman to express the group's list of key features and uncertainties. S/he explains that the focus will be on composing a best answer/strategy in which the uncertainties are properly addressed.
	Collection	The teacher explains that the purpose in this part is to get an overview of the different group input: what useful answers did the different groups come up with and what are the major uncertainties, which have to be addressed. The teacher evaluates each group list of key features by (having the intended best answer/strategy in mind) letting students comparing these with each other. The teacher can structure the discussion with the question: what would be the most complete list of key features our class 'answer/strategy' *must* address (in order to be the best)?
	Elaboration	If a group's input is not clear, the students or teacher can ask a group to further clarify (For example: 'what do you mean by..”?). If a group's input is not complex enough, the students / teacher can ask for further information.
	Transmission	Based on the most complete list of key features and linked general uncertainties (which should be less than the individual expressed uncertainties!) produced in the prior part, the teacher formulates or lets a student formulate the best class answer /strategy. Thereby diminishing remaining uncertainties by putting them into perspective. The teacher does this as much as possible in line with what is put forward by the different groups and asks the class to respond. All prior parts have prepared the students for this step.
Completion/ bridge to prepare the context for the next question		The teacher summarises the content-related outcomes embodied in the 'one best class answer/strategy and why it isn't perfect' and proceeds by linking these outcomes with some short remarks to the preparation of the context of the next question.

Appendix 2 The questions of the final test and the score-criteria

Question 1

Question 1 was formulated as follows:
Marjolein wants to find out if she can make the water form a pond nearby drinkable. She uses a plastic bottle, with which you can produce drinkable water. This bottle is developed for travellers. Marjolein has filtered some water from the pond and has purified the water with the bottle. Afterwards, Marjolein wants to check if the water is drinkable. She tests the water on the following parameters: E-coli-bacteria, nitrite, acidity and chloride.

A. On the next page you can see a part of the official list of water quality criteria[18]. The real list is much longer. Why can Marjolein normally suffice with testing the parameters E-coli bacteria, nitrite, acidity and chloride if she wants to test if the water has drinking water quality?

After Marjolein concludes that the water is drinkable, she drinks it. A few hours later, she feels dizzy and sick and she develops a persistent headache. Marjolein discovers that a plant is dumping its wastewater in the pond. She wants to know why she is ill, but the people of the plant do not respond. In view of the type of plant, the water might have contained:

Formaldehyde, methyl bromide, dioxaan or phenol
You can find information on these substances in the appendix

B. Check of each substance if Marjolein might have swallowed too much of it. Use the information of the appendix. Explain, based on your findings, which substance you think is the most probable.

The answers to question 1A were marked according to the following criteria:
Adequate:
These four parameters give a good general indication of the drinking water quality (or: if one of those four parameters exceeds the norm for drinking water quality, it is a sign that something might be wrong). Besides other parameters are not probable, it is much too expensive and time consuming to test all the parameters on the list.

Incomplete but sufficient:
These four parameters give a good general indication of the drinking water quality (or: if one of those four parameters exceeds the norm for drinking water quality, it is a sign that something might be wrong).

Wrong/insufficient:
All other answers

The answers to question 1B were marked according to the following criteria:
Adequate:
methyl bromide is the most probable based on the Marjoleins symptoms and toxic effects of *each* of the substances.

Incomplete but sufficient:
If a student only explains why methyl bromide is the most probable, and not why all the others are improbable.

Wrong/insufficient:
All other answers

[18] Part of the original list was taken up in the test

Question 2
Question 2 was formulated as follows:
Mei Ling wants to test the water quality of a ditch in a preserved nature reserve. Below you will find apart of the official list of parameters and norms for a ditch in a nature reserve[19].

A. Explain why this list differs from the list of parameters and norms for drinking water

Mei Ling uses Merck kit tests. Like when drinking water is standard tested, Mei Ling does not have to test all the parameters from the list. She tests the four relevant parameters in this case:

Parameter	test result	Scale of the accompanying Merck kit colour cards.
Nitrate	20 mg/litre	0-10-20-25-50-75 mg/litre
Phosphate	13 mg/litre	0-5-10-15-20-25 mg/litre
Acidity	7	2-3-4-5-6-7-8-9
Chloride (in this case also a Merck test)	100 mg/litre	0-25-50-100-150-200-250 mg/litre

B. Explain what conclusion Mei Ling might draw about the water quality of the ditch. Use your own experiences with the Merck kit tests.

Merck kit tests last a limited of time. The phosphate test turns out to be a year over due. Mei ling is afraid that the test is not reliable anymore.

C. How might Mei Ling find out if the phosphate test is still reliable? Describe your strategy in such a way that Mei Ling knows exactly what to do and why.

The answers to question 2A were marked according to the following criteria:
Adequate:
The answers should contain two notions:
1. Ditch water has a different function than drinking water.
2. The water quality criteria therefore differ.

Incomplete but sufficient:
Ditch water and drinking water are two different water functions

Wrong/insufficient:
All other answers

The answers to question 2B were marked according to the following criteria:
Adequate:
1. All parameters seem to answer their norms.
2. The phosphate result however, is very close to the norm. Considering the scale of the colour card, and the inaccuracy of these types of tests (you cannot see the difference between 13 and 17 mg/litre), the test result is doubtful. Mei Ling cannot conclude just like that, that the water quality is good.

Incomplete but sufficient:
The results are within the norm, but you have to take in to account the inaccuracy of these types of tests (in general).

[19] This list includes the norms for the parameters which Mei Ling has tested: Nitrate: 75 mg/ litre; Phosphate: 15 mg/litre; acidity: 6,5< pH 8,5;Chloride: 200 mg/litre.

Wrong/insufficient:
All other answers

The answers to question 2C were marked according to the following criteria:
Adequate:

1. First you should make a phosphate solution, so that you know what the concentration is and what the test result should be.
2. Next you should test the solution with Merck kit test and compare the test result with what you know it should be.

Or:

1. You should take the water sample and test it with a different test method
2. You should compare the test result with the original test result, are they the same?

Incomplete but sufficient:
Answers that only refer to notion 1., such as:
Send the water sample to a lab.
Use a different test method

Wrong/insufficient:
All other answers

Question 3
Question 3 was formulated as follows:
A laboratory frequently monitors the water quality of a swimming pool.

A. Which steps must the lab go through, in order to be able to judge the water quality properly?

The laboratory must report its conclusions. They do this by filling in a standard form, so that everyone can see how they came to their conclusion.

B. Name five types of information, which should be filled in on this form.

The answers to question 3A were marked according to the following criteria:
Adequate:
1. water use – determine water quality criteria; 2. test; 3. compare & consider inaccuracies 4. judge

The answers to question 3B were marked according to the following criteria:
Adequate:
Five of the following possibilities: date, which laboratory, water use, water quality criteria: ‘parameters and norms’ or ‘norms’, sample-method, test method; number of samples, the results, deviation in the results, judgment and argumentation, suggestions of further testing.

Incomplete but sufficient:
When four things are mentioned

Wrong/insufficient:
All other answers

Appendix 3 The post-trial questionnaire

The last few weeks, you have worked on the unit 'water quality'. You have noticed that the lessons were observed and taped. This was because you worked on the theme in a special way. We are therefore very interested in your experiences and ask you to please fill in this questionnaire.
Thank you.

General impression
The topic of the unit was "judging water quality"

1 How interesting did you find the topic "judging water quality"?
(please circle your answer)

1	2	3	4	5
very interesting				not at all interesting

2 How important do you consider learning about this topic?

1	2	3	4	5
very important				not important at all

3. Do you feel that you learned at lot doing this unit?
Yes / no
Namely: …

4. What do you consider the most important thing you learned? What the less important thing?
Most important: …

Less important: …

5. Is there anything you think differently about now? If yes, what?

6. How do you feel about the water quality lessons if you compare them to your regular chemistry lessons?

7. How do you feel about doing the water quality tests if you compare that to regular chemistry experiments?

8. What did you like and what did you not like?
I liked: …

I did not like: …

9. Do you think the unit was difficult or easy generally speaking?

1	2	3	4	5
Very easy				very difficult

10 What do you think was good about the unit and what do you think was not?
Good: …

Not good: …

11. What do you think needs to be changed or improved about the unit?

The tasks.
The unit consisted of 16 tasks. With the following questions we want to see if you saw the purpose of the tasks. You can skip through your workbook to refresh your memory.

12. Was it clear to you after the introductory chapter 1 (up to task 3) what you were going to do the next lessons and how you were going to do this?

13. Were there any tasks that were not logical to you, of which you did not understand why you had to do it?

14. When doing the tests, did you experience any insecurity about the tests and/or the test results?

15. Did the tasks of chapter 4 help you to diminish these doubts (if you had them)?

The following two questions got lost when the questionnaire was photocopied:

16. In tasks 15 you were asked to fill in a final report en to see if you could find the procedure steps in the report.
Did you know which steps were meant?

Yes/no, namely …

17. Suppose you want to judge the water quality of water with a specific use. Do you now know what kind of information you need and how you might proceed?

Please explain

Appendix 4 Function descriptions

In tables A, B and C an overview of the episodic function descriptions of the third version and the proposals for a possible fourth version are presented. The tables serve to clarify further the distinction between didactical functions for students and didactical functions for the teacher by lining up the descriptions of the third versions next to a proposal to formulate a possible fourth version. They are structured as follows. In the first column presents the episode numbers. In the second column the function descriptions of each of the episodes of the third version are listed. In the third column and in fourth column an overview is presented of the proposal for each of the episodes of the didactical functions for the students (column 3) and the didactical functions for the teacher (column 4) respectively. The sole purpose of these tables is to clarify the implications of the distinction between didactical functions for the students and didactical functions for the teacher, without presenting a complete fourth version of these function descriptions. Therefore, for some of the episodes the adjusted function descriptions are incomplete (or even absent). Sometimes I will suffice with proposing some directions or ingredients for a proper didactical function description.

Table A Function descriptions of episodes 0-2

Episode functions-third version		**Episode functions-fourth version**	
Nr	*Function*	**Didactical functions for students**	**Didactical functions for teacher/designer**
0	X	X	*In order to* achieve that students are truly involved in learning about how chemistry functions in practice *I am going to* prepare myself for teaching an exemplary module of meaningful chemistry education: the instructional version of the practice of monitoring water quality.
1	Students are to feel generally motivated to get involved in the instructional version of the practice about judging water quality. They are to get a clear view of the purpose and the unfolding of the lessons, and their role in it.	*In order to* find out how good water quality is maintained, *we are going to* find out how people monitor water quality in order to solve certain cases, because these cases are important: good water quality is an essential need in life.	*In order to achieve* that students are truly involved in learning about how chemistry functions in practice *I am going* to achieve in the stage setting of the instructional version of the practice that students identify as much as possible with the overall purpose and at the same time secure that their relevant pre-knowledge is properly activated by letting them discuss several exemplary cases of the authentic practice and collecting their opinions about these cases to make sure that they feel that their input matters.
2	When applying their intuitive knowledge about the procedure to be followed in the exemplary case, students are to experience that their intuitive knowledge is not sufficient. They know that the water should be tested, but lack specific issue knowledge, which might be formulated as the following knowledge need: 'What does the water sample contain?'	*In order to* find out how people monitor water quality, *we are going* to judge the water quality of an exemplary case. *In order to* be able to judge the water quality of this exemplary case, *we are going* to test the water. *In order to* be able to test the water of this exemplary case *we are going to* find out from the authentic practice what we should test the water on and how to do this.	*In order to* achieve that students see the point of solving an exemplary case as a way to pursue their aim to find out how people solve such cases, *I am going to* invite them to solve an appealing (drinking water) exemplary case together to involve them more, and *I am going* to collect, as a representative of the authentic practice, their questions for more specific information on what to test and how in order to be able to test the water (in such a way that they feel that their input matters).

Table B Function descriptions of episodes 3-5

Episode functions-third version		**Episode functions-hypothetical fourth version**	
Nr	*Function*	**Didactical functions for students**	**Didactical functions for teachers**
3	Students satisfy their induced knowledge need, which will raise the need for taking the next step in this process: 'how much of what is allowed in drinking water?'	*In order to* find out how people monitor water quality *we are going to* judge the water quality of an exemplary case. *In order* to be able to judge the water quality of this exemplary case, *we are going to* test the water on the parameters that are used in the authentic practice and compare the test results to some standard derived from the authentic practice of what counts as good water.	In order to…I am going to Directions: the teacher should for example see the point, beforehand, of providing students information on what to compare their test results to, when *students* ask for this, of letting them write down their doubts when testing in light of being able to functionally introduce for students the concepts of reliability and accuracy of the tests and the argumentation behind the water quality criteria later on.
4	Students are to link their doubts to: a. Trustworthiness of the limited list of parameters and norms. b. Trustworthiness of the test results.	*In order to* find out how people monitor water quality *we are going to* judge the water quality of an exemplary case. In order to be able to judge the water quality of the exemplary case, we are going to find out if these criteria are really sufficient and if we can really trust our test results.	*In order to* achieve that students continue to be involved in the teaching-learning process, seeing the point of every step and feel that their input matters in light of achieving their aim, *I aim going to* collect the students doubts about the tests, their performances and the water quality criteria in such a way that they realise they need to solve these doubts to achieve their aim to solve the case and find out how people solve such cases. *I am going to* collect the students doubts as literally as possible, making sure that students feel that their input matters and categorize them in two types of doubts, as each type will be addressed later on in a separate episode, in a way that is obvious for students.
5	Students address the episode question by learning how the list of four parameters and norms to be standard tested in this case is established.	*In order to* find out how people monitor water quality *we are going to* judge the water quality of an exemplary case. In order to be able to judge the water quality of the exemplary case, we are going to solve our doubts about the water quality criteria.	

Table C Function descriptions of episodes 6-8

Episode functions-third version		**Episode functions-hypothetical fourth version**	
Nr	*Function*	**Didactical functions for students**	**Didactical functions for teachers**
6	Students address the episode question by developing a notion about reliability and accuracy of the used test methods and how this might influence their judgment.	*In order to* find out how people monitor water quality *we are going to* judge the water quality of an exemplary case. In order to be able to judge the water quality of the exemplary case, we are going to solve our doubts about our test results. So that we can finally solve the exemplary case and find out how people monitor water quality.	In order to achieve that students learn as about the procedure which is followed in the exemplary case, I am going to involve students in activities in which they learn about the concepts variability and position of test results and how they should use these concepts when judging the accuracy of their test results. I will do that by using their doubts about the test results as means to functionally introduce these concepts in such a way that they see the point of every step and feel that their input matters
7	Students solve the exemplary case and summarise their procedural knowledge and feel the need to test the usefulness of the procedure in the other exemplary cases (because they expect this to be the case).	In order to find out how people monitor water quality, we are going to verify with the authentic practice if we really achieved our aim.	
8	Students are to reflect on 'what they have learned': a procedure for how existing practices judge water quality in standard situations.	In order to…we are going to	In order to…I am going to

Summary

The subject of this thesis is a developmental research project which was performed by the Centre for Science and Mathematics Education at Utrecht University between 1999 and 2005. The project addressed the question of how to involve students in learning chemistry through a proper implementation of three characteristics of meaningful: a *context*, a *need-to-know* approach and *attention for student input.* The implementation of these three characteristics was expected to result in the students being generally motivated, seeing the point of every activity in the teaching-learning process and having the feeling that their input matters. As one module is object of study, the central, situated question of this research has been:

What is an adequate teaching-learning process in a module about judging water quality for initial chemistry education (students: 14-15) which properly embodies the three characteristics of meaningful: a motivating context, a proper need to know and a proper attention for student input?

The project has resulted in an exemplary design heuristic for other modules that aim for meaningful chemistry education and in the idea of functional embeddedness as a central principle in designing meaningful chemistry education. By functional embeddedness as a design principle I mean that student activities are to be so embedded in an overall purpose that students realise beforehand how the activities are going to contribute to this (their) overall purpose. In this case, the overall purpose was: finding out about how people judge water quality. This conclusion is relevant in two ways. Firstly, the idea of functional embeddedness is a more directive design principle than the three characteristics of meaningful. It provides for more specific constraints on the design. Secondly, the principle of functional embeddedness brings certain didactical issues to the fore that are generally recognised in science education research and also *connects* them. One of these issues is how to functionally embed the top-down established learning goals of a teaching learning process in the interests and goals of the *students* to achieve that the students become motivated and see the point of getting involved in the teaching learning process.

Chapter 1 describes in the form of anecdotes how certain problematic features of chemistry education lead to a lack of involvement of students. The unclear relevance of school chemistry, its high level of abstraction and the lack of students' interest in the subject are widely recognised problems.

Chapter 2 describes three main problematic features of chemistry education that emerge from the literature and explain the students' lack of involvement in learning chemistry. These problematic features are: traditional school chemistry has the character of a rhetoric of conclusions, the sequence of its content is incoherent at both the concept and the curriculum level and it shows a lack of attention for student input. Three widely applied potential solution strategies which can be seen as addressing these problems are identified: *context*, *need-to-know* and *attention for student input*. These three 'characteristics of meaningful chemistry education' are adopted as solution strategies. This has led to the situated research question formulated above. It

was expected that with these strategies the problematic features could be solved and students would become more involved in the learning of chemistry as follows.

Context
A well defined, for students' recognisable context for concepts provides for the use of the concepts with a distinct function, and thereby makes students' use of the concepts meaningful and motivating. The 'rhetoric of conclusions' feature can be avoided, when the emphasis is shifted from 'getting an overview of the conceptual products of chemistry' to the 'functional use of concepts in relation to a certain relevant, recognisable context'. A *coherent development of concepts* can be achieved in this way, as concepts are to be used with a 'distinct function and meaning'.

Need-to-know
Addressing students' questions on a need to know basis, which also implies building properly on their existing knowledge, provides for an increasing involvement of students in the teaching-learning process, as they will see the point of what they learn every step of the way. This characteristic, together with the characteristic of a well-defined context, can provide for the development of a *coherent emphasis.* The first and second characteristics both result in a coherent sequencing of the content.

Attention for student input
The third characteristic is closely related to the second: if the aim is to incorporate the need-to-know approach in the design of a teaching-learning process, then 'real attention for student input' cannot be avoided. In a successful *need-to-know* approach, students have more insight into and experience the functionality of 'what comes next'. As a result, the teacher has more opportunity to pay real attention to their input, which now should become a driving force of the content-related progression. Consequently, students will *feel that their input matters*.

In **chapter 3** it is argued that developmental research according to Lijnse (1995) is the proper strategy for answering the research question. The situated research question has been addressed in three research cycles with three successive, increasingly adequate, versions of the design. Each time the realised teaching-learning process was evaluated in detail using a scenario of argued expectations of every teaching-learning activity. These argued expectations form the argumentation as to why certain design decisions were expected to give content to the characteristics of meaningful chemistry education.

How the first two research cycles led to the third version of the design is discussed in **chapter 4**. Each research cycle led to the formulation of specific didactical problems that emerged from the problem of properly operationalising and implementing the three characteristics of meaningful. Each time this led to new ideas and operationalisations and to a new version of the design and the scenario.

The first research cycle
In the first version of the design, the three characteristics were operationalised as follows:
Context: A relevant and recognisable driving question that is expected to appeal to students. In this case: Is the water in our neighbourhood clean enough?
Need-to-know: A set of sub-questions that students are expected to need for answering the driving question. The sub-questions are derived from the driving question by putting oneself, with the learning goals in mind, in the position of the students.
Because it was a learning goal that students should master the procedure used to judge water quality, some of the sub-questions can in retrospect be seen as somewhat reflecting the steps of this procedure, although at the time this was not an explicit aim.
Attention for student input: This characteristic was operationalised in two ways:
Students experienced certain autonomy of choice, in particular with respect to which water they want to test. Students presented their own findings to their classmates. In a final reflective class discussion, guided by the teacher, the conclusion emerged that they all followed the same procedure. This insight and the understanding that all groups contributed to this insight should have given students the feeling that their input matters.

The evaluation of the first version of the design showed that students were very motivated to answer the driving question. They were also excited by the fact that they could choose their own water and present their own findings and they appreciated this autonomy of choice. However, students did not see the point of many sub-questions. It could therefore be concluded that the need-to-know had not been adequate. In addition, the procedure (the learning goal) was not a well-integrated part of the teaching-learning process Because of this, students did not realise that all groups contributed to the learning goal of 'making the common procedure explicit'.

In order to solve these problems a problem-posing approach was adopted as a refinement of the need-to-know. A problem-posing approach basically means that the teaching-learning activities are designed in such a way that students are put in the position that they feel the need and see the point, in the perspective of some broad motive (e.g. to answer an appropriate, top-down established, driving question) to extend their knowledge in a certain direction. It is primarily a way of emphasising that in the design and in the evaluation of the teaching-learning process special and detailed attention is paid to the creation of such a broad motive and connected 'knowledge needs' or 'content-related motives' among students.
The procedure was expected to become an integrated part of the teaching-learning process as follows. *Firstly*, students were to develop a broad motive for explicitly developing the procedure for judging water quality. *Secondly*, the procedural steps were used to evoke content-related motives by using the intuitive knowledge of students about what would be the logical next step and by using the evaluative question 'Would you now drink the water, do you think it meets the criteria?'. Each next procedural step was explicitly summarised in an extra activity. *Finally,* the procedure (which should be the purpose of the students) was explicitly reflected upon in a concluding activity.

A problem-posing approach had implications for *the attention for student input* characteristic as well. The teacher was to pay attention to the student input, as far as it was an important driving force of the teaching-learning process, e.g. when students, at certain points in the teaching-learning process, express their content-related motives as questions.

The second research cycle
In the second version of the design, students were to test the drinking water quality of water they had produced from surface water. The three characteristics were operationalised as follows:
Context: A relevant and recognisable driving question that was expected to appeal to students. In this case: Is the water clean enough?
Need-to-know: Within an overall broad motive (corresponding to the driving question) a series of connected and nested content-related motives should be triggered among students by using students' intuitive knowledge of what would be the next logical step of the procedure and the quality control question to induce the need for a next necessary procedural step.
Attention for student input: Students had certain autonomy of choice, in particular a choice of the surface water from which they were to produce drinking water. The input of all groups was necessary to achieve the learning goal of making the common procedure for judging water quality explicit. Students were to develop the insight that each group followed the same procedure and realise that all groups contribute to this insight throughout the teaching-learning process. The teacher was to pay attention to the student input, as far as it was an important driving force of the teaching-learning process.

The evaluation of the second version of the design showed that students felt broadly motivated by the driving question and that the intended content-related motives were triggered among students (or could have been) in the teaching-learning process. However, this was not always the case. Students still did not see the point of some activities, such as the summarising activities and the concluding activity in which students were to reflect on the procedure. At some points they lost track of the driving question altogether, for example, when they have to purify surface water and produce drinking water. This activity was included to let students choose their own surface water (which was expected to contribute to the *attention for student input* characteristic) but it interrupted the problem-posing story line. It was concluded that the didactical structure was still not good enough.

The evaluation also showed that students still experienced and appreciated the autonomy to choose their own water and present their own findings, although the activity of producing drinking water quality from surface water interrupted the problem-posing story line and undermined the *need-to-know*. For the *attention for student input* characteristic it was an adequate operationalisation. However, the conclusion was also drawn that in the designed teaching-learning process no structural and explicit attention was paid to the role of the teachers and how they could have made students feel that their input mattered as a driving force in the teaching-learning

process. As a result, the teachers often tended to more or less ignore student input. In line with Lemke's findings (1990), the teachers relied on certain types of interaction patterns (or dialogue structures) that are common in traditional science teaching where the teacher is largely responsible for the content-related progression. This interfered with the problem-posing teaching-learning process in which student input formed a driving force. Consequently, the type of interaction patterns prevented students from sufficiently feeling that their input mattered in the process.
Finally, the procedure still was not a well-integrated part of the teaching-learning process.

The evaluation had the following implications for the interpretation of the three characteristics of meaningful and their interrelatedness. With respect to the *context* and *need-to-know* characteristic, it was concluded that the idea of a driving question and the broad motive to answer such a question (*context)* did not sufficiently direct what content-related motives could be triggered among students at which moments (*need-to-know*). The attempt to strengthen the relationship between *context* and *need-to-know* led to the idea of establishing an instructional version of an authentic practice.

A practice involves a central, characteristic procedure (or activity) with distinct purposes and aims (in this case: judging water quality in order to decide whether it meets the criteria for its use). The characteristic procedure that is followed in such a practice is functional for achieving these purposes and aims. In this case, that would be the water quality judging procedure that is followed by chemistry analysts in a lab using standard methods.
The basic idea is that such a characteristic procedure can be used to strengthen the relationship between *context* and *need-to-know*, thereby making the procedure an integrated part of the teaching-learning process, if the following conditions are met.

a. Students should appreciate the purposes and aims of the authentic practice. Presenting to them appealing exemplary cases of the practice is expected to contribute to a broad motive to learn about the practice by simulating one of the exemplary cases, in the sense that it relates what is going to happen in chemistry class to society at large.
b. Students should have intuitive knowledge of the characteristic procedure that is followed in the authentic practice and appreciate its functionality in achieving its purposes and aims. The evaluations of the previous versions proved that students do have such intuitive knowledge of the procedural steps. This intuitive knowledge and appreciation of 'what would be a logical next step' can be used to induce content-related motives among students.

With respect to the *attention for student input* characteristic, the interaction between the teacher and the students should be structured at the level of *each* activity and within a format that matches the complexity of the content-related progression of the respective activity.
It was concluded that the teacher should be given explicit guidelines for how to direct interactions in order to achieve this. The teacher should for example give the students the opportunity to express the content-related motives that emerged from an activity

(to be addressed in the next activity, and so on) in such a way that they feel that by doing this they drive the teaching-learning process. These considerations have led to the design and implementation of the interaction structures that provide the teacher with guidelines for directing interaction. The complexity of the student input determines the type of interaction. A typical guideline for the teacher is, for example, that s/he can suffice with collecting students' opinions, because s/he can rely on the fact that students are able to do so and see the point of expressing the intended opinions.

Finally, with respect to the aspect of autonomy of choice of the *attention for student input* characteristic: in an instructional version of the authentic practice, students will have less autonomy of choice than in the second version of the design, because the instructional version of the authentic practice now dictates what cases are relevant to present to the students. It is up to the designer/researcher to decide which selection of cases would be appealing and appropriate to present to the students to motivate them and which specific case would be interesting and appropriate as an example to solve.

The third research cycle
The evaluation of the second version of the design led to a third version. The third research cycle is presented in detail in section 4.6 (third operationalisation of the three characteristics), **chapter 5** (the scenario), **chapter 6** (the final test: methods of evaluation) and **chapter 7** (the evaluation). Section 4.6 and chapter 5 describe how the three characteristics of meaningful were interrelated and given content in the third version of the design.

Context and *Need-to-know*: The reconceptualisation of (the connections between) the characteristics *context* and *need-to-know* led to the idea of designing an instructional version of an authentic practice. Students were to become broadly motivated by the purposes of the authentic practice to adopt their role in its instructional version and find out how people in the authentic practice judge water quality by simulating the authentic practice in a sense. The students' intuitive knowledge of such a procedure is used to design a problem-posing teaching-learning process, thus creating an *instructional* version of the procedure of the authentic practice.
Attention for student input: Student input contributes throughout the complete teaching-learning process to the common purpose of developing the procedure for judging water quality. The teacher directs interaction at the level of setting the stage and of the evaluation of *each* of the episodes of the problem-posing story line, according to the chosen interaction structure.

The evaluation of the third version of the design showed that students were motivated by the purpose of the authentic practice through the selection of appealing exemplary cases. However, the evaluation also showed that students did not explicitly adopt their intended roles and the specific connected purpose of making the common procedure explicit. This was due to the design of the activities and to the fact that the teacher did not explicitly invite the students to do so. However, in comparison to the second version, students' content-related motives were induced more as expected and

intended in light of their aim (to solve the case). As a result, students felt the purpose of the flow of the activities more than in the second version. It was concluded that the evaluation of the third version of the design did not to give cause for adjustments at the level of the didactical structure and that at this level the didactical structure of the design was adequate. The intended story line of the teaching-learning process was consistently designed.
The fact that some of the content-related motives were not induced as intended was due to the design of the activities. The evaluation showed that the intended content-related motives *could* have been triggered, had the activities been properly designed, for example because they emerged at other moments in the teaching-learning process. Especially in the unfolding of the more complex episodes, which constituted of a series of activities, students tended to lose track of the purpose of their activities. As a result, they did not experience the activities of these episodes as functional beforehand, in light of achieving their purpose, but rather afterwards.

With respect to *attention for student input* the following was concluded. Students felt, more than in the second version, throughout the teaching-learning process that their input contributed to a common purpose (only not making the common procedure explicit but solving the exemplary case). The teacher structurally paid more attention to the input of students. As a result, students were more involved and increasingly felt that their input mattered.
However, the more complex interaction structures did not provide the teacher with sufficient guidelines as to how to direct the interaction at the level of each activity. It was concluded that interaction structures do not cover the full scope of *attention for student input*: students should feel that their input matters *for their purpose.*

These findings suggested that the operationalisation of the *attention for student input* characteristic needed rethinking. The design proved to be more adequate in triggering the students' content-related motives in light of a, for them, more clear purpose (compared to the second version of the design). As a result they also felt that what they did (their input) drove the teaching-learning process and therefore *mattered* in achieving their purpose. Therefore, designing a proper instructional version of a practice that meets the criterion that students see the point of every activity beforehand, in light of achieving their purpose, actually *is* paying proper attention to student input in the design of a teaching-learning process.

In **chapter 8,** the situated research question is answered and its contribution to didactical theory is discussed.
With respect to the situated research question, it can be concluded that the didactical structure of the design, figures 4.6A and B, was adequate. The ideas of an instructional version of an authentic practice and the implementation of interaction structures have contributed to the creation of meaningful chemistry education. These ideas (including the problems that emerged from the third evaluation) can be seen as a design heuristic for other designers of modules of meaningful chemistry education. Obviously, this heuristic does not provide for algorithmic decisions, which only have to be 'executed' to end up with a meaningful teaching-learning process. They rather

serve as a specified heuristic scheme of design steps and criteria for decisions. In practice this means that for a new module a detailed, situation specific and intensive study of the teaching-learning process is needed to establish to what extent meaningful chemistry education has been realised in a design. This study is a first exploration of 'an instructional version of an authentic practice' and 'interaction structures' and several aspects need further study, especially with respect to the specific didactical problems that emerged from the third research cycle:

- To achieve that students adopt their role in the instructional version of the practice and the connected purpose of finding out about the procedure (the 'stage setting') is to a certain extent still problematic.
- To achieve that students experience the activities as functional beforehand, in light of this purpose (which should now be their purpose) is to a certain extent still problematic at the level of the design of activities, especially in the more complex episodes of the teaching-learning process.
- The teacher role (including interaction structures) in all this (e.g. setting the stage and directing interaction) is to a certain extent still problematic.

The second conclusion that can be drawn is that the operationalisation of the three characteristics of meaningful can, in retrospect, be seen as governed by one single principle, which I have called the principle of *functional embeddedness*. In retrospect, this idea can be seen as directing the operationalisation of all three characteristics of meaningful throughout this study, in the first cycles rather intuitively but in the third cycle more explicitly. The three characteristics can finally be rephrased as follows.

Context
The purpose of the instructional version of the authentic practice should be as much as possible *functionally embedded* in the values, goals or interests of students or in those of the society at large.

Context, *need-to-know* and *attention for student input*
Activities should be as much as possible *functionally embedded* in the purpose of the instructional version of the practice (which now should be the students' purpose).

The teacher role
a. The teacher role should contribute to inviting students to play their part in the instructional version of the practice by reducing the tension between top-down set learning goals and students' basic interests as much as possible.
b. The interaction structures should provide the teacher with guidelines for directing the interaction between teacher and students at the level of each activity in such a way that students are given their part of the responsibility of the progression of the process as intended.

Functional embeddedness is a more directive design principle than 'meaningful' or 'need-to-know'. Also, the principle of functional embeddedness not only brings certain didactical issues to the fore that are more generally recognised in science education research and (not surprisingly) partly coincide with the problematic issues that emerged from this study, but it also *connects* them. For example, to solve problem 1 (below) some projects let students choose their own learning goals to a certain extent (giving them autonomy of choice). However, the more student are left to decide on the learning goals, the more difficult it will be to solve problem 2.

1. External functional embeddedness of learning goals in basic interests and goals**.**
2. Functional embeddedness of activities in the learning goals of the teaching-learning process.
3. Functional embeddedness of the teaching-learning process in the curriculum.
4. A functional fine-tuning of the teacher's activities to all issues 1-3.

Samenvatting

Dit proefschrift beschrijft een onderzoek dat tussen 1999 en 2005 is uitgevoerd aan het Centrum voor Didactiek van Wiskunde en Natuurwetenschappen aan de Universiteit Utrecht. Aanleiding voor het onderzoek is de observatie dat leerlingen scheikunde vaak een abstract en weinig relevant vak vinden. De vraag hoe scheikundeonderwijs zodanig vormgegeven kan worden, dat leerlingen werkelijk betrokken raken in het onderwijsleerproces stond centraal. Op grond van een literatuurstudie werden drie strategieën voor, of 'karakteristieken van betekenisvol scheikunde onderwijs' geïdentificeerd: *context*, *need-to-know* en *aandacht voor de eigen inbreng van leerlingen*. De verwachting was dat een adequate implementatie van deze karakteristieken in een onderwijsontwerp zou leiden tot gemotiveerde leerlingen die het nut van elke onderwijstaak inzien en die het gevoel hebben dat hun eigen inbreng ertoe doet. Als onderwerp voor het experimentele ontwerp werd gekozen voor 'het beoordelen van waterkwaliteit'. De onderzoeksvraag was:

Wat is een effectief onderwijsleerproces voor het beoordelen van waterkwaliteit voor chemieonderwijs in de basisvorming waarin de drie karakteristieken van betekenisvol scheikundeonderwijs adequaat geïmplementeerd zijn?

Het project heeft geresulteerd in een empirisch onderbouwde ontwerpheuristiek voor vergelijkbare modulen die als voorbeelden van 'betekenisvol scheikundeonderwijs' functioneren. Ook heeft het geresulteerd in de conclusie dat het idee van 'functionele inbedding' het richtinggevende ontwerpprincipe is geweest in de drie onderzoeksrondes. Met functionele inbedding als een ontwerpprincipe bedoel ik dat de handelingen van leerlingen zodanig ingebed moeten worden in een overkoepelend doel dat het voor leerlingen bij voorbaat duidelijk is hoe deze handelingen gaan bijdragen aan het bereiken van dat (hun) doel. In het onderhavige geval is dit overkoepelend doel: uitzoeken hoe waterkwaliteit beoordeeld wordt. Dit is om twee redenen relevant. Ten eerste is het principe van functionele inbedding een meer richtinggevend ontwerpprincipe dan de drie eerdergenoemde karakteristieken. Ten tweede verheldert het principe van functionele inbedding bepaalde didactische dilemma's die ook breder in onderzoek naar natuurwetenschappelijk onderwijs een rol spelen.

In **hoofdstuk 1** worden aan de hand van anekdotes problemen van het scheikundeonderwijs beschreven: de onduidelijke relevantie van het vak en het hoge abstractieniveau. De anekdotes illustreren dat leerlingen hierdoor het vak als moeilijk en weinig relevant ervaren. Deze problemen worden zowel nationaal als internationaal breed onderkend.

Hoofdstuk 2 gaat in op de oorzaken van het hoge abstractieniveau en de onduidelijkheid over de relevantie van het schoolvak. Het hoofdstuk beschrijft drie oorzaken van deze problemen van scheikundeonderwijs die uit onderzoek daarover naar voren komen. Deze oorzaken zijn:

- Schoolscheikunde is vooral een succesverhaal, een 'rhetoric of conclusions', dat buiten beschouwing laat hoe en waarom wetenschappers tot bepaalde theorieën kwamen.

- De vakinhouden zijn incoherent geordend. Dit geldt zowel op het niveau van concepten als op het niveau van invalshoeken (‘curriculum emphases’).
- Er is amper aandacht voor de eigen inbreng van leerlingen in het onderwijsleerproces.

Op grond van een analyse van een selectie van vernieuwingsprojecten zijn drie strategieën geïdentificeerd die beschouwd kunnen worden als potentiële oplossingsstrategieën voor de geïdentificeerde problemen: *context*, *need-to-know* en *aandacht voor de eigen inbreng van leerlingen.* De verwachting is dat de implementatie van deze drie ‘karakteristieken van betekenisvol scheikundeonderwijs’ in een onderwijsleerproces als volgt bijdraagt aan een oplossing voor de problemen:

Context
Een heldere, voor leerlingen herkenbare context waarin bepaalde begrippen worden toegepast voorziet het gebruik van deze begrippen van een duidelijke functie en daardoor van een duidelijke betekenis. Hierdoor wordt de toepassing van deze begrippen voor leerlingen betekenisvol.
De invalshoek van het scheikundeonderwijs verschuift op deze manier van ‘een overzicht krijgen van de conceptuele producten van scheikunde’ naar ‘het functionele gebruik van begrippen binnen een relevante, herkenbare context’. Door de begrippen een heldere functie en betekenis binnen de context te geven kan een coherente ordening van begrippen bereikt worden.

Need-to-know
Wanneer aandacht wordt besteed aan de vragen van leerlingen op een *need-to-know* basis, zien leerlingen het nut van wat ze leren in. Als gevolg daarvan raken ze meer betrokken in het onderwijsleerproces. Dit impliceert dat er in het ontworpen onderwijsleerproces op een goede manier moet worden voortgebouwd op de kennis die leerlingen al hebben. Deze karakteristiek voorziet (samen met de vorige karakteristiek) het onderwijsleerproces van een duidelijke invalshoek. De eerste en tweede karakteristiek zorgen samen voor een coherente ordening van de vakinhouden (vanuit het perspectief van de leerling).

Aandacht voor de eigen inbreng van leerlingen
De derde karakteristiek is nauw gerelateerd aan de tweede: als het doel is het ontwerp uit te werken op basis van de *need-to-know* karakteristiek, dan is werkelijke aandacht voor de eigen inbreng van leerlingen onvermijdelijk. Een succesvol uitgewerkte *need-to-know* geeft leerlingen meer zich op ‘wat er komen gaat’. De docent heeft meer gelegenheid om werkelijk aandacht te geven aan de eigen inbreng van leerlingen, omdat die inbreng nu een drijvende kracht van de inhoudelijke voortgang is. Leerlingen zullen meer het gevoel krijgen dat hun eigen inbreng er toe doet. Deze karakteristiek biedt een oplossing voor het gebrek aan aandacht voor de eigen inbreng van leerlingen.

In **hoofdstuk 3** wordt de keuze voor de onderzoeksstrategie toegelicht. Beargumenteerd wordt waarom ontwikkelingsonderzoek (Lijnse, 1995) de geschikte onderzoeksstrategie is om de onderzoeksvraag beantwoorden. Elke activiteit van het ontworpen onderwijsleerproces wordt in deze strategie gemotiveerd in het scenario en verwachtingen worden met behulp van dit scenario aan de praktijk getoetst. Er hebben drie onderzoeksrondes plaatsgevonden steeds met een verbeterde versie van het ontwerp. De verwachtingen in het scenario komen in iedere ronde in steeds grotere overeen met de empirisch vastgestelde resultaten. Uiteindelijk vormen deze beargumenteerde verwachtingen de de basis voor de redenering waarom bepaalde ontwerpbeslissingen adequaat inhoud geven aan de drie karakteristieken van betekenisvol scheikundeonderwijs.

Hoe de eerste twee onderzoeksrondes hebben geleid tot de derde versie van het ontwerp wordt beschreven in **hoofdstuk 4**. Naast een verantwoording van de keuze voor de vakinhoud en leerdoelen, wordt in dit hoofdstuk beschreven hoe het streven naar een adequate operationalisatie van de drie karakteristieken heeft geleid tot de steeds specifiekere formulering van didactische problemen. Dit heeft voor elke onderzoeksronde geleid tot nieuwe ideeën, operationaliseringen en nieuwe versies van het scenario.

De eerste onderzoeksronde
In de eerste versie van het ontwerp werden de drie karakteristieken als volgt geoperationaliseerd:
Context. De keuze voor een relevante en herkenbare centrale vraag, waarvan verwacht werd dat deze vraag de leerlingen aanspreekt. In dit geval: is het water in onze buurt schoon genoeg?
Need-to-know. Een serie subvragen waarvan de verwachting is dat leerlingen de antwoorden nodig hebben om de centrale vraag te kunnen beantwoorden. De subvragen worden afgeleid van de centrale vraag, door zich in te leven in de leerlingen met de leerdoelen in het achterhoofd.
Omdat het een van de leerdoelen is dat leerlingen de procedure voor de beoordeling van waterkwaliteit leren, kan achteraf worden vastgesteld dat sommige subvragen in zekere zin procedurestappen reflecteren (hoewel dit niet een expliciet doel was).
Aandacht voor de eigen inbreng van leerlingen. Deze karakteristiek werd op twee manieren geoperationaliseerd. Ten eerste ervaren leerlingen een zekere autonomie in hun keuze, in het bijzonder in de keuze van het water dat ze willen testen. Ten tweede presenteren leerlingen hun eigen bevindingen aan hun klasgenoten. In een afsluitende reflectieve klassendiscussie komt, onder leiding van de docent, naar voren dat iedereen min of meer dezelfde procedure heeft gevolgd. Dit inzicht en de realisatie dat alle groepen bij hebben gedragen aan deze conclusie zal naar verwachting bijdragen aan het gevoel bij leerlingen dat hun eigen inbreng ertoe doet.

De evaluatie van de eerste versie van het ontwerp liet zien dat leerlingen erg gemotiveerd waren om de centrale vraag te beantwoorden. Ze waardeerden ook dat ze zelf konden kiezen welk water ze wilden testen en hun bevindingen konden presenteren. Met andere woorden: ze waardeerden de autonomie in keuze. Leerlingen zagen echter niet het nut in van veel van de subvragen. Er kan daarom geconcludeerd worden dat de *need-to-know* karakteristiek niet adequaat was vormgegeven. Ook was de procedure (het leerdoel) niet een voldoende geïntegreerd onderdeel van het onderwijsleerproces. Hierdoor realiseerden leerlingen zich niet dat alle groepen hadden bijgedragen aan een gemeenschappelijk leerdoel (namelijk: het expliciet maken van de procedure).

Om het probleem van de inadequate vormgeving van de *need-to-know* karakteristiek op te lossen werd voor de tweede versie van het ontwerp gekozen voor een probleemstellende benadering. Dit betekent dat de activiteiten zo ontworpen worden dat de behoefte bij leerlingen gecreëerd wordt om hun kennis in een specifieke richting uit te breiden in het licht van een overkoepelend motief (bijvoorbeeld: het antwoorden van een centrale vraag). De probleemstellende benadering is er nadrukkelijk op gericht om in het ontwerp en in de evaluatie expliciete en gedetailleerde aandacht te besteden aan het opwekken van inhoudelijke motieven bij leerlingen, zowel een breed overkoepelend motief als meer specifieke deelmotieven.

De procedure voor het beoordelen van waterkwaliteit moest beter geïntegreerd worden in het onderwijsleerproces. De verwachting was dat dit als volgt bereikt kon worden. *Ten eerste* moesten leerlingen gemotiveerd worden voor het doel 'de procedure voor het beoordelen van waterkwaliteit expliciet maken' (hoe doet men dat?). *Ten tweede* werden nu de procedurestappen gebruikt om inhoudelijke motieven bij leerlingen op te wekken. Dit gebeurde door de intuïtieve kennis van de leerlingen over wat de volgende logische stap zal zijn te gebruiken. Elke volgende stap in de procedure werd bovendien expliciet samengevat in een extra toegevoegde activiteit. *Tenslotte* werd er gereflecteerd op de procedure door leerlingen in een afsluitende activiteit.

Een probleemstellende benadering heeft ook gevolgen voor de *aandacht voor de eigen inbreng van leerlingen*. Deze eigen inbreng is in een probleemstellende benadering een sturende kracht in het onderwijsleerproces. De docent moet aandacht besteden aan de eigen inbreng van leerlingen, bijvoorbeeld wanneer ze hun inhoudelijke motieven verwoorden als vragen.

De tweede onderzoeksronde

In de tweede versie van het ontwerp moesten leerlingen drinkwater produceren uit oppervlaktewater en dit water testen. De drie karakteristieken waren als volgt geoperationaliseerd:

Context: Een relevante en herkenbare centrale vraag waarvan de verwachting is dat deze leerlingen aanspreekt. In dit geval: is het water schoon genoeg?

Need-to-know: Binnen een overkoepelend algemeen motief (het beantwoorden van de centrale vraag) moet er bij de leerlingen een serie geneste inhoudelijke motieven opgeroepen worden (in dat overkoepelende motief). Hiervoor kan aan de ene kant gebruik worden gemaakt van de intuïtieve kennis van leerlingen van de te volgen procedure (de volgende logische stap). Aan de andere kant kan gebruik worden gemaakt van de evaluatieve vraag 'Denk je dat het water aan de eisen voor drinkwater voldoet?' om de behoefte aan een volgende specifieke procedurestap op te roepen.

Aandacht voor de eigen inbreng van leerlingen. Deze karakteristiek wordt op drie manieren geoperationaliseerd.

a. Leerlingen hebben een zekere autonomie van keuze. Ze mogen zelf kiezen van welk oppervlaktewater ze drinkwater willen produceren.
b. De eigen inbreng van alle leerlingen is nodig om het gemeenschappelijke leerdoel te bereiken: het expliciet maken van de procedure voor het beoordelen van waterkwaliteit. Leerlingen ontwikkelen het inzicht dat alle groepen in wezen dezelfde procedure volgen. Ze realiseren zich dat de eigen inbreng van alle groepen heeft geleid tot dit inzicht.
c. De docent moet aandacht besteden aan de eigen inbreng van leerlingen in het probleemstellende onderwijsleerproces.

De evaluatie liet zien dat leerlingen in het algemeen gemotiveerd werden door de centrale vraag. Ook konden in het onderwijsleerproces de bedoelde inhoudelijke motieven worden opgewekt (of hadden kunnen worden opgewekt) bij leerlingen. Dit was echter niet altijd het geval. Leerlingen zagen nog steeds niet het nut in van sommige activiteiten, zoals de samenvattende activiteiten en de afsluitende activiteit waarin leerlingen moesten reflecteren op de procedure. Soms verloren de leerlingen de centrale vraag uit het oog. Bijvoorbeeld toen ze drinkwater moesten produceren uit oppervlaktewater. Deze activiteit was bedoeld om leerlingen de gelegenheid te geven hun eigen water te kiezen. Het was de bedoeling dat deze autonomie in keuze zou bijdragen aan de karakteristiek *aandacht voor de eigen inbreng aan leerlingen*. Het verstoorde echter de probleemstellende verhaallijn. Mede op grond hiervan werd geconcludeerd dat de didactische structuur van het ontwerp nog niet adequaat was. Aan de andere kant liet de evaluatie ook zien dat leerlingen de autonomie die ze kregen om zelf oppervlaktewater te kiezen waardeerden. Voor de karakteristiek *aandacht voor de eigen inbreng van leerlingen* was het een adequate operationalisering, ook al verstoorde deze activiteit de verhaallijn en de *need-to-know* karakteristiek.

Ook moet er geconcludeerd worden dat in het ontwerp de rol van de docenten geen structurele aandacht had gekregen. Docenten bleken onvoldoende voorbereid om een deel van de verantwoordelijkheid voor de inhoudelijke voortgang van het onderwijsleerproces over te dragen aan leerlingen en ze het gevoel te geven dat hun eigen inbreng belangrijk was voor deze voortgang. Hierdoor hadden de docenten de neiging om de eigen inbreng van leerlingen te negeren. In lijn met Lemke's bevindingen (Lemke,1990) vervielen de docenten in bepaalde interactiepatronen (of dialoogstructuren) die gebruikelijk zijn in traditionele scheikunde lessen. Het zijn interactiepatronen die passen in een situatie waarin de docent grotendeels verantwoordelijk is voor de inhoudelijke voortgang van het onderwijsleerproces. Dit interfereerde met een probleemstellende benadering waarin de eigen inbreng van leerlingen een belangrijke drijfveer voor de inhoudelijke voortgang vormt. Hierdoor hadden leerlingen veel minder het gevoel dat hun eigen inbreng ertoe deed dan de bedoeling was. Tenslotte bleek de procedure nog steeds niet goed geïntegreerd te zijn in het onderwijsleerproces.

De evaluatie had de volgende implicaties voor de operationalisatie van de drie karakteristieken en hun onderlinge relatie. Wat betreft de *context* en *need-to-know* karakteristiek werd geconcludeerd dat het idee van een centrale vraag en het opwekken van een algemeen motief om deze vraag te beantwoorden (*context)* onvoldoende richting gaf aan welke inhoudelijke motieven op welke momenten opgeroepen zouden kunnen (en moeten) worden (*need-to-know*). De relatie tussen deze karakteristieken, *context* en *need-to-know*, moest versterkt worden. Dit leidde tot het idee van het didactiseren van een bestaande, authentieke praktijk. Het idee is als volgt. In een authentieke praktijk wordt een karakteristieke procedure gevolgd om bepaalde duidelijke doelen te bereiken. In dit geval: water kwaliteit wordt getest en beoordeeld om te kunnen beslissen of het aan de criteria voor de waterfunctie voldoet. Met andere woorden: de procedure is functioneel om het doel van de authentieke praktijk te bereiken. De karakteristieke procedure kan gebruikt worden om de relatie tussen *context* en *need-to-know* te versterken en om de procedure als vanzelf te integreren in het onderwijsleerproces. Aan de volgende voorwaarden moet worden voldaan:

c. Leerlingen moeten de doelen van de authentieke praktijk waarderen door ze verschillende voorbeelden uit de betreffende praktijk te presenteren. Dit moet bij leerlingen een algemeen motief opwekken om te leren over de praktijk door een van de voorbeelden te ‘simuleren’.
d. Leerlingen moeten de procedure intuïtief kennen en de logica ervaren van de opeenvolgende procedurestappen in het licht van de te bereiken doelen (een waterkwaliteitsoordeel bijvoorbeeld). Deze intuïtieve kennis van wat de volgende stap zal zijn, kan gebruikt worden om inhoudelijke motieven op te roepen bij leerlingen.

Wat betreft de karakteristiek *aandacht voor de eigen inbreng van leerlingen*, moet de interactie tussen docent en leerlingen adequaat gestructureerd worden op het niveau van elke activiteit. Dit moet afgestemd worden op de complexiteit van de inhoudelijke voortgang. De docent moet expliciete richtlijnen krijgen voor het vormgeven van de communicatie over en weer. De docent moet leerlingen bijvoorbeeld de gelegenheid geven om hun inhoudelijke motieven (waar vervolgens in de volgende activiteit op in wordt gegaan) die door een activiteit opgeroepen zijn zodanig te verwoorden dat leerlingen het gevoel krijgen dat hun eigen inbreng belangrijk is voor de inhoudelijke voortgang van het proces.
Deze overwegingen leidden tot het ontwerpen van ‘interactiestructuren’ die de docent van richtlijnen voorzien voor het dirigeren van de interactie. De complexiteit van de eigen inbreng van leerlingen, bepaalt het type interactiestructuur. Een typische richtlijn voor de docent is bijvoorbeeld dat hij/zij kan volstaan met het inventariseren van de meningen van de leerlingen, omdat hij/zij erop kan vertrouwen dat leerlingen het nut inzien van het uiten van hun mening en hier ook toe in staat zijn. Tenslotte, in een gedidactiseerde authentieke praktijk hebben leerlingen minder ‘autonomie in keuze’ dan in de tweede versie van het ontwerp. De gedidactiseerde praktijk schrijft nu voor welke casussen relevant zijn om aan de leerlingen te presenteren. Het is aan de ontwerper/onderzoeker om te beslissen welke selectie van voorbeelden van de authentieke praktijk geschikt zijn om aan leerlingen te presenteren en welk voorbeeld interessant is voor leerlingen.

De derde onderzoeksronde
De derde onderzoeksronde wordt in detail beschreven in paragraaf 4.6 (de derde operationalisering van de drie karakteristieken), **hoofdstuk 5** (het scenario), **hoofdstuk 6** (evaluatiemethoden) en **hoofdstuk 7** (de evaluatie).

De drie karakteristieken van betekenisvol scheikundeonderwijs zijn in de derde ronde duidelijk met elkaar geassocieerd. Ze zijn als volgt geoperationaliseerd:
Context en *Need-to-know*. Het opnieuw doordenken van de karakteristieken *context* en *need-to-know* en hun onderlinge relatie leidde tot het idee van een gedidactiseerde vorm van een authentieke praktijk. Leerlingen moeten gemotiveerd worden voor de doelen van de authentieke praktijk om hun rol in de gediactiseerde versie te gaan spelen en uit te zoeken hoe mensen in de authentieke praktijk water kwaliteit beoordelen door de authentieke praktijk te simulereren. Leerlingen hebben intuïtieve kennis van de procedure. Die intuïtieve kennis wordt gebruikt in het ontwerp van een probleemstellend onderwijsleerproces. Op deze manier wordt een gedidactiseerde versie van de authentieke praktijk gecreëerd.
Aandacht voor de eigen inbreng van leerlingen. De eigen inbreng van leerlingen draagt gedurende het hele onderwijsleerproces bij aan het gemeenschappelijke doel: het ontwikkelen van een procedure voor de beoordeling van waterkwaliteit. De docent dirigeert de interactie op het niveau van de ‘stage setting’ en tijdens de evaluatie van elke episode van het ondrewijsleerproces. De vorm van de interactie hangt af van de gekozen interactiestructuur.

De evaluatie van de derde versie van het ontwerp liet zien dat de leerlingen inderdaad gemotiveerd werden door het doel van de authentieke praktijk, door de selectie van voorbeelden die werd gepresenteerd. De evaluatie liet echter ook zien dat leerlingen hun rol niet expliciet accepteerden en daardoor ook niet het bij die rol horende specifieke doel van de procedure expliciet maken. De oorzaak hiervan lag in de concrete uitwerking van de activiteiten en in het feit dat de docent de leerlingen niet duidelijk uitnodigde om hun rol op zich te nemen. Aan de andere kant werden, zoals verwacht, de inhoudelijke motieven van de leerlingen veel meer oproepen dan in de tweede versie van het ontwerp. Hierdoor zagen de leerlingen beter het doel van de opeenvolgende activiteiten dan in de tweede versie. Op grond hiervan kan geconcludeerd worden dat de evaluatie van de derde versie van het ontwerp geen aanleiding geeft tot herziening van de didactische structuur van het ontwerp. Op dit niveau is het ontwerp adequaat en is de lijn van activiteiten goed ontworpen.
Het feit dat sommige inhoudelijke motieven niet werden opgeroepen als verwacht was het gevolg van de concrete uitwerking van de activiteiten. De evaluatie liet zien dat de bedoelde inhoudelijke motieven opgeroepen hadden *kunnen* worden als de activiteiten goed waren uitgewerkt. Bijvoorbeeld omdat ze op andere momenten in het onderwijsleerproces door leerlingen werden geuit. Speciaal in de complexere episodes, die uit meerdere activiteiten bestonden, ervoeren de leerlingen de activiteiten niet bij voorbaat als functioneel, gezien in het licht van het bereiken van hun doel (soms pas achteraf).
Wat betreft de karakteristiek *aandacht voor de eigen inbreng van leerlingen* kan geconcludeerd worden dat leerlingen meer dan in het vorige experiment het gevoel

hadden dat hun eigen inbreng gedurende het hele proces ertoe deed. In vergelijking met de vorige versie besteedde de docent structureel meer aandacht aan eigen inbreng van leerlingen. Hierdoor gedroegen leerlingen zich veel meer betrokken.
Op sommige momenten echter voerde de docent de interactiestructuren uit zoals bedoeld, maar konden leerlingen niet hebben ervaren dat hun eigen inbreng belangrijk was voor het bereiken van hun doel. Ze gedroegen zich betrokken, maar de evaluatie liet ook zien dat ze de functie van deze activiteiten voor het bereiken van hun doel niet hadden ervaren. Ook bleken de complexere interactiestructuren de docent onvoldoende richtlijnen te geven voor het sturen van de interactie op het niveau van *elke* activiteit. Op grond hiervan kan geconcludeerd worden dat de ontworpen interactiestructuren niet volledig bewerkstelligen dat leerlingen het gevoel krijgen dat hun eigen inbreng belangrijk is *in het bereiken van hun doel*.

Deze bevindingen suggeren dat de karakteristiek *aandacht voor de eigen inbreng van leerlingen* herzien moet worden. Dat leerlingen het gevoel hebben dat hun eigen inbreng een rol speelt in het bereiken van een gemeenschappelijk doel zou het resultaat moeten zijn van een goed uitgewerkte probleemstellende benadering. Als een probleemstellende benadering goed is uitgewerkt ervaren leerlingen namelijk dat alles wat ze doen in het onderwijsleerproces de drijfveer is in dit proces. Met andere worden, alles wat ze doen draagt bij aan het bereiken van hun doel. 'Het ontwerpen van een gedidactiseerde praktijk die voldoet aan het criterium dat leerlingen het nut van elke activiteit bij voorbaat ervaren in het licht van het bereiken van hun doel' is *hetzelfde* als 'aandacht besteden aan de eigen inbreng van leerlingen'. Interactiestructuren moeten dan beschouwd worden als richtlijnen voor de docent om zijn/haar rol in de gedidactiseerde praktijk adequaat vorm te geven en recht te doen aan de karakteristiek *aandacht voor de eigen inbreng van leerlingen*.

In **hoofdstuk 8** wordt de situatiespecifieke onderzoeksvraag beantwoord en wordt er gereflecteerd op de bijdrage van deze studie aan de didactische theorie.
Wat betreft de onderzoeksvraag kan geconcludeerd worden dat de didactische structuur van het ontwerp (zie figuren 4.6 A en B) adequaat is. De gedidactiseerde versie van een bestaande authentieke praktijk en interactiestructuren hebben bijgedragen aan het creëren van betekenisvol scheikundeonderwijs. Deze ideeën, ondanks de didactische problemen die in derde onderzoeksronde nog naar voren kwamen, kunnen beschouwd worden als een ontwerpheuristiek voor het ontwerpen van andere vergelijkbare modules. Deze heuristiek biedt houvast voor het uitvoeren van ontwerpstappen en voor criteria voor het maken van ontwerpbeslissingen. Om vast te stellen of een ontwerp voldoet aan de verwachtingen zal echter steeds een ontwikkelingsonderzoek nodig zijn. Deze studie is een eerste verkenning van het idee van een gedidactiseerde praktijk en van interactiestructuren. Veel aspecten dienen verder onderzocht te worden, met name de specifieke problemen die in de derde onderzoeksronde naar voren kwamen:

- Het bereiken dat leerlingen hun rol in de gedidactiseerde praktijk aannemen en daarmee het bijbehorende doel (het expliciet maken van de procedure voor het beoordelen van waterkwaliteit) is in zekere zin nog steeds problematisch.

- Het bereiken dat leerlingen elke activiteit bij voorbaat als functioneel ervaren, in het licht van hun doel, is in zekere zin nog steeds problematisch op het niveau van het concrete ontwerp van activiteiten (speciaal die van de meer complexe episodes van het verhaal).
- De docentenrol in dit alles is in zekere zin nog steeds problematisch.

De tweede conclusie is dat, terugkijkend, de operationalisatie van de drie karakteristieken van betekenisvol scheikundeonderwijs vanaf het begin in weze gestuurd werd door één enkel ontwerpprincipe dat ik het principe van functionele inbedding heb genoemd. Dit ontwerpprincipe heeft de operationalisatie van de drie karakteristieken steeds gestuurd, in het begin intuïtief maar steeds meer expliciet. Vanuit dit principe kunnen de drie karakteristieken nu als volgt geherformuleerd worden:

Context
Het doel van de gedidactiseerde praktijk moet zoveel mogelijk functioneel ingebed worden in de belangstelling en doelen die leerlingen hebben.
Het evaluatiecriterium wordt dan:
Ervaren leerlingen het bereiken van het doel van de gedidactiseerde praktijk (en hun rol) als een manier om hun eigen interesses en doelen te verwezenlijken?

Context, *need-to-know* en *aandacht voor de eigen inbreng van leerlingen*
Activiteiten moeten zoveel mogelijk functioneel ingebed worden in het doel van de gedidactiseerde praktijk (nu als het goed is ook het doel van de leerlingen).
Het evaluatiecriterium wordt dan:
Ervaren leerlingen de activiteiten bij voorbaat als functioneel, in het licht van het bereiken van hun doel en hebben ze als gevolg hiervan het gevoel dat hun eigen inbreng ertoe doet?

De docentenrol
a. De docentenrol draagt bij aan de functionele inbedding van het doel van de gedidactiseerde praktijk in de interesses van de leerlingen. De docent moet gericht zijn op het zoveel mogelijk overbruggen van de spanning tussen van bovenaf opgelegde leerdoelen en de interesses en doelen van leerlingen.
b. De docentenrol draagt bij aan het tot zijn recht komen van de karakteristiek *aandacht voor eigen inbreng van leerlingen*. De implementatie van interactiestructuren moet de docent van richtlijnen voorzien voor het dirigeren van de interactie tussen docent en leerlingen. Dit moet gerealiseerd worden op het niveau van elke activiteit en zodanig dat leerlingen hun deel van de verantwoordelijkheid voor de inhoudelijke voortgang krijgen.

Het evaluatiecriterium wordt dan:
Vervult de docent zijn/haar rol zoals bedoeld? Heeft dit tot gevolg dat:
a. de hierboven beschreven spanning minder wordt en de leerlingen gemotiveerd raken om hun rol te spelen?
b. de leerlingen zich verantwoordelijk voelen voor de inhoudelijke voortgang van het onderwijsleerproces?

Tenslotte is het principe van functionele inbedding een richtinggevender ontwerpprincipe dan bijvoorbeeld het principe van een '*need-to-know'*. Het verbindt onderstaande vier problemen met elkaar waarbij bepaalde didactische dilemma's naar voren komen die breder onderkend worden in het onderzoek van natuurwetenschappelijk onderwijs.

1. Het probleem van externe functionele inbedding van leerdoelen in de interesses en doelen van leerlingen**.**
2. Het probleem van functionele inbedding van activiteiten en leerdoelen van een onderwijsleerproces.
3. Functionele inbedding van een onderwijsleerproces in een curriculum
4. Een functionele fine-tuning van de activiteiten van de docent wat betreft de drie bovenstaande punten.

Een voorbeeld van zo'n dilemma is de mate van autonomie die leerlingen krijgen in het kiezen van een leerdoel om tegemoet te komen aan punt 1. Echter, hoe meer autonomie leerlingen krijgen, hoe moeilijker het wordt om het probleem onder punt 2 op te lossen.

Dankwoord

Het promotieonderzoek was voor mij een onvergetelijke, intense en heel erg leerzame tijd. Velen hebben in meer of mindere mate en op verschillende manieren bijgedragen aan mijn welzijn in het algemeen en aan de totstandkoming van dit proefschrift in het bijzonder.
Ik wil graag mijn collega's van het Centrum voor Bèta Didactiek bedanken. Allereerst dank ik mijn begeleiders Albert, Kees en Astrid voor hun geduld, voor de vrijheid die ik kreeg in het onderzoek, voor het scheppen van structuur in de chaos en voor alle energie die in mijn begeleiding is gestoken. Ik heb erg veel van jullie geleerd.
Zonder anderen tekort te doen, wil ik nog een aantal mensen graag speciaal bedanken. Om te beginnen denk ik met veel plezier terug aan de gesprekken met Wobbe de Vos. Kees wil ik graag speciaal bedanken voor zijn bereidwilligheid om halverwege de rit mijn begeleiding op zich te nemen toen Wobbe de Vos overleed. Marie-Christine voor haar steun in moeilijke tijden, voor haar gastvrijheid als het weer eens laat werd en voor onze gesprekken over onderwijs en didactiek. Machiel, jarenlang mijn kamergenoot, dat hij het al die jaren met me heeft uitgehouden en ook voor zijn goede raad (en zijn hulp met de computer). Hanneke voor being Hanneke, je humor en je gewillig oor. Daan voor je orginele inzichten. Inge voor de relativerende gezamenlijke anti-stress loopjes. Graag wil ik Dieuwke bedanken voor de inspirerende gesprekken. Ilona voor haar bereidheid om commentaar te leveren op mijn stukken en om mijn interpretatie van protocollen na te kijken. Gjalt voor al zijn hulp die varieerde van het regelen van een laptop tot het leveren van commentaar op stukken. Atie voor de zeer gewaardeerde persoonlijke gesprekken en Sanne voor al die uren die ze gestoken heeft in het redigeren van het Engels. Frits Pater, Riet Leewis, Cees Loffeld, Angelique Roes en Ben Goes, dank voor jullie praktische en technische ondersteuning.
Dank ook aan de docenten en leerlingen zonder wie dit onderzoek niet mogelijk was geweest. Ewout Gijben (Montessori Lyceum Amsterdam), Patrick van de Geijn (Koningin Wilhelmina College, Culemborg), Ton Raaijmakers en Bert Vijfvinkel (College de Heemlanden, Houten), ik heb onze samenwerking als zeer plezierig en inspirerend ervaren. Veel dank voor jullie bereidheid deel te nemen aan dit onderzoek, jullie enthousiasme, gastvrijheid en jullie waardevolle bijdragen aan dit onderzoek.

Tenslotte wil ik mijn familie en vrienden bedanken. Zonder hen had ik dit niet kunnen doen. Ik dank jullie dat jullie er waren. Ik wil speciaal mijn vader en zijn vrouw Judith bedanken, die mij uit enkele diepe dalen hebben getrokken. Mirjam en Paul, Eva en Frank, bedankt voor jullie onwrikbare vertrouwen en jullie niet aflatende steun. En natuurlijk Hans, Emma en Mirjam. Ik kan niet onder woorden brengen hoe belangrijk jullie zijn. Alleen al de gedachte aan jullie maakte alles lichter.

Curriculum Vitae

Hanna Westbroek werd op 5 april 1967 geboren te Den Haag. In 1985 behaalde zij haar gymnasium-ß diploma aan het Fioretti college te Lisse. Hierna begon zij aan een studie scheikunde aan de Universiteit van Amsterdam. In augustus 1991 studeerde zij af in de richting biochemie, met als specialisatie microbiologie.

In 1993 haalde zij haar eerstegraads lesbevoegdheid aan de Universiteit van Amsterdam. Aansluitend was zij scheikunde docente aan het Willem de Zwijger College in Naarden-Bussem. Van 1995 tot 1999 was zij scheikunde, natuurkunde en sinds 1998 ook ANW docente aan het Bernard Nieuwentijt college te Amsterdam-Noord. In mei 1999 begon zij als assisent in opleiding bij het Centrum voor Didactiek van Wiskunde en Natuurwetenschappen (CD-ß) te Utrecht. Binnen deze functie is het onderzoek uitgevoerd waarover in dit proefschrift wordt gerapporteerd.

CD-ß series on research in science education

1. Didactiek in Perspectief – P.L. Lijnse & W. de Vos (Eds.)
2. Radiation and Risk in Physics Education – H.M.C. Eijkelhof
3. Natuurkunde-onderwijs tussen Leefwereld en Vakstructuur – R.F.A. Wierstra
4. Een Onverdeelbare Eenheid – M.J. Vogelezang
5. Betrokken bij Evenwicht – J.H. van Driel
6. Relating Macroscopic Phenomena to Microscopic particles: A Central Problem in Secondary Science Education – P.L. Lijnse, P. Licht, W. de Vos & A.J. Waarlo (Eds.)
7. Kwaliteit van Kwantiteit – H.E. Elzenga
8. Interactieve Video in de Nascholing Reken-wiskunde – F. van Galen, M. Dolk, E. Feijs, V. Jonker, N. Ruesink & W. Uittenbogaard
9. Realistic Mathematics Education in Primary Schools – L. Streefland (Ed.)
10. Ontwikkeling in Energieonderwijs – A.E. van der Valk
11. Methoden in het Reken-wiskundeonderwijs – K. Gravemeijer, M. van den Heuvel-Panhuizen, G. van Donselaar, N. Reusink, L. Streefland, W. Vermeulen, E. te Woerd & D. van de Ploeg
12. De volgende opgave van de computer – J. Zuidema & L. van der Gaag
13. European Research in Science Education – P.L. Lijnse (Ed.)
14. Realistic Mathematics Education – K. Gravemeijer
15. De Grafische Rekenmachine in het Wiskundeonderwijs – L.M. Doorman, P. Drijvers & M. Kindt
16. Making sense – Simulation-of-Research in Organic Chemistry Education – H. van Keulen
17. Perspectives on Research in Chemical Education – O. de Jong, P.H. van Roon & W. de Vos (Eds.)
18. A problem-posing approach to teaching the topic of radioactivity – C.W.J.M. Klaassen
19. Assessment and realistic mathematics education – M. van den Heuvel-Panhuizen
20. Teaching structures in chemistry – An Educational Structure for Chemical Bonding – G.M. van Hoeve-Brouwer
21. Regulatie en homeostase als onderwijsthema: een biologie-didactisch onderzoek – J. Buddingh'

22. Over Natuurkundedidactiek, Curriculumontwikkeling en Lerarenopleiding – P.L. Lijnse & T. Wubbels
23. Integratie en toepassing van biologische kennis – Ontwikkeling en onderzoek van een Curriculum rond het thema 'Lichaamsprocessen en Vergift' – H. Roebertsen
24. Het thema 'reproductie' in het schoolvak biologie – P.C.F. Reygel
25. Teaching electrochemical cells – A study on teachers conceptions and teaching problems in secondary education – J.J.C. Acampo
26. The role of context and models in the development of mathematical strategies and procedures – K. Gravemeijer
27. Thermodynamica leren onderwijzen – W.H. Kaper
28. Interessegeoriënteerd natuur- en scheikundeonderwijs – Een studie naar onderwijsontwikkeling op de Open Schoolgemeenschap Bijlmer – R. Genseberger
29. Flexibilization of mental arithmeticsstrategies on a different knowledge base – The empty number line in a realistic versus gradual program design – A.S. Klein
30. A problem posing approach to teaching an initial particle model – M.J. Vollebregt
31. Met het oog op integratie – Een studie over integratie van leerstof uit de natuurwetenschappelijke vakken in de tweede fase van het voortgezet onderwijs – C. Beeftink
32. Verschillen tussen meisjes en jongens bij het vak rekenen-wiskunde op de basisschool – Eindrapport MOOJ-onderzoek – M. van den Heuvel-Panhuizen & H.J. Vermeer
33. Van vormleer naar realistische meetkunde – Een historisch-didactisch onderzoek van het meetkundeonderwijs aan kinderen van vier tot veertien jaar in Nederland gedurende de negentiende en twintigste eeuw – E.W.A. de Moor
34. Ontwerpend leren in het biologieonderwijs. Uitgewerkt en beproefd voor immunologie in het voortgezet onderwijs – F.J.J.M. Janssen
35. Natuur in pluralistisch perspectief; theoretisch kader en voorbeeldlesmateriaal voor het omgaan met een veelheid aan natuurbeelden – M. Margadant-van Arcken & C.S. van den Berg (Red.)
36. Duurzaamheid als leergebied; conceptuele analyse en educatieve uitwerking – S. Lijmbach, M. Broens, D. Hovinga & M. Margadant-van Arcken
37. A problem-posing approach to teaching decision making about the waste issue – J. Kortland
38. Teaching for scientific literacy: context, competency, and curriculum – O. de Jong, E.R. Savelsbergh & A.H. Alblas
39. Met sprongen vooruit. Een productief oefenprogramma voor zwakke rekenaars in het getallengebied tot 100 – een onderwijsexperiment – J.J.M. Menne
40. A gateway to numeracy. A study of numeracy in adult basic education – M. van Groene-stijn

41. Reinvention of early algebra. Developmental research on the transition from arithmetic to algebra – B.A. van Amerom
42. Education in Israel on collaborative management of shared water resources – M. Dressler
43. Coping with the abstract and complex nature of genetics in biology education. The yo-yo learning and teaching strategy – M.C.P.J. Knippels
44. Werken aan kwaliteitsverbetering van leerlingonderzoek. Een studie naar de ontwikkeling en het resultaat van een scholing voor docenten – Th.J.M. Smits
45. Teaching formal mathematics in primary education. Fraction learning as mathematising process – R. Keijzer
46. Leren kijken door de ontwerpersbril. Het vorm-functieperspectief als leerdoel van natuuronderwijs – D.J. Boerwinkel
47. Als je begrijpt wat ik bedoel. Een zoektocht naar verklaringen voor achterblijvende prestaties van allochtone leerlingen in het wiskundeonderwijs – C. van den Boer
48. Learning algebra in a computer algebra environment. Design research on the understanding of the concept of parameter – P. Drijvers
49. Towards systems thinking in cell biology education – R.P. Verhoeff
50. Design research in statistics education. On symbolizing and computer tools – A. Bakker
51. Modelling motion: from trace graphs to instantaneous change – L.M. Doorman
52. Characteristics of meaningful chemistry education. The case of water quality – H.B. Westbroek

Centre for Science and Mathematics Education (CD-ß)
Universiteit Utrecht
P.O.Box 80.008
3508 TA Utrecht
The Netherlands