

# GENERAL INTRODUCTION

## CHAPTER 1

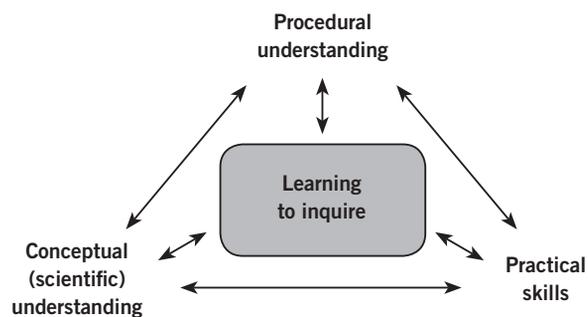
This chapter introduces the five empirical studies reported on in this thesis. The studies are on improving students' evaluation of the accuracy, reliability and validity (ARV) of inquiries in different school science subjects. This project was undertaken because in the Netherlands pre-university science students and many science teachers are novices at evaluating ARV. In section 1.1 the problem and the aim of the research are defined. In section 1.2 the main research question of this study is addressed. In section 1.3 the theoretical background and the main concepts are described. These concepts are: learning to inquire; the meaning of accuracy, reliability and validity in school science inquiries; the transfer of meaning of accuracy, reliability and validity; formative feedback with self-evaluation instruments as rubrics; the SOLO taxonomy as a suitable hierarchy to describe the content of a self-evaluation instrument; and the implementation of a self-evaluation instrument in a teaching-learning process. In section 1.4 the structure of this thesis is described and the research questions of the successive empirical studies are addressed. An outline of the methodology in each study is given. In section 1.5 the context and relevance of the thesis to the field of scientific knowledge and to the educational practices concerned are discussed.

### 1.1 PROBLEM AND AIM OF THE RESEARCH

Pre-university science students in the Netherlands learn to perform inquiries in the school science subjects of biology, chemistry and physics. 'Learning to inquire' in these school science subjects can have various aims for the students. The emphasis can lie on an increase of conceptual (scientific) understanding of the natural world,

on practical skills, or on procedural understanding<sup>1</sup> (Millar, Lubben, Gott, & Duggan, 1994). This study, which is situated in the field of science education, uses definitions of conceptual understanding, practical skills and procedural understanding drawn from previous studies about learning to inquire in school science subjects (e.g. Millar et al., 1994; Gott & Duggan, 1995). The students' practical skills can be developed by providing them with tasks that focus on, for example, the practical use of measuring instruments, use of a thermometer or the construction of tables and graphs (Millar, 2010). Increasing students' conceptual understanding by having them perform inquiries requires tasks in the different school science subjects that focus on domain-specific ideas, which are based on scientific facts, concepts, laws and principles. For example, by conducting an inquiry on mechanics (physics), a student can increase his or her understanding about the relation between gravitational force and the constant acceleration of an object and learn about Newton's laws of motion. Procedural understanding in learning to inquire focuses on the understanding and application of concepts of evidence (CoE). CoE are supportive concepts in the evaluation and improvement of accuracy, reliability and validity (ARV) in an inquiry (Gott & Duggan, 1995; Millar, 2010). Understanding the evaluation of ARV in inquiries is a prerequisite to gain insight into the nature of science as well as the way scientists work and think about their research. Increasing students' procedural understanding helps them to apprehend the criteria that scientists use and to understand their increasingly technological and scientific environment (Aarsen & Van der Valk, 2008; Abd-El-Khalick et al., 2004; Chinn & Malhotra, 2002; Harlen, 2012; Van Rens, Van Muijlwijk, Beishuizen, & Van der Schee, 2013).

Figure 1.1  
Relation between  
various aims for students  
in learning to inquire;  
based on Millar et al.  
(1994) and Gott and  
Duggan (1995, p. 30)



<sup>1</sup> Our use of the terms 'conceptual understanding' and 'procedural understanding' could be confusing for cognitive psychologists. Cognitive psychologists use the terms 'declarative knowledge' and 'procedural knowledge' to distinguish between knowledge about content, facts and propositions (declarative knowledge) and knowledge about skills and strategies that one applies in performing a task (procedural knowledge). Declarative knowledge can be obtained by learning, while procedural knowledge is often unconsciously acquired by the learner (Gagné, Yekovich, & Yekovich, 1993).

Figure 1.1 shows the relation between these three aims of learning to inquire for pre-university science students. It should be noticed that each inquiry always requires conceptual (scientific) understanding as well as practical skills and procedural understanding. It is the choice of the teacher or educational designer which of the three aims to pursue in an inquiry activity (Gott & Duggan, 1995; Millar, 2010). In this study I chose to focus on inquiries that aim at increasing students' procedural understanding, and more specifically, at students' self-evaluation of the ARV of such an inquiry in various school science subjects. I chose this subject area because pre-university science students in the Netherlands and many of their science teachers are novices in evaluating ARV in inquiries in the school science subjects. Students also lack the procedural understanding to apply the evaluation of ARV flexibly from one school science inquiry context to another. Dutch pre-university science students are mostly used to 'hands-on' practical work, in which they use a stepwise approach without much ownership and often with little reflection on their actions, as is described in more detail in Chapter 2. These students can hardly give meaning to ARV let alone evaluate ARV in an inquiry in their school science subjects (Lunetta, Hofstein, & Clough, 2007; Schalk, Van der Schee, & Boersma, 2013). Moreover, pre-university science students seldom recognise that there are many similarities in planning and conducting inquiries, as well as evaluating ARV in inquiries they perform, in the different science subjects. As a consequence, these students do not know how to transfer the evaluation of ARV from an inquiry in physics to another inquiry, for instance, in chemistry (Roberts & Gott, 2002).

In the last decade, research about learning to inquire in chemistry (Van Rens, 2005) and biology (Schalk, 2006) has shown that the CoE model, as described by Gott, Duggan, Roberts, and Hussain (n.d.), can be useful in increasing the procedural understanding of students. Van Rens (2005) and Schalk (2006) focused on procedural understanding in one school science subject but the evaluation of ARV in inquiries also contains many interdisciplinary aspects for the different school science subjects. It is meaningful for pre-university students to learn to transfer their procedural understanding of the evaluation of the ARV of an inquiry from one school science context (e.g. biology) to another (e.g. physics or chemistry) in order to increase their procedural understanding. The transfer of procedural understanding of the evaluation of ARV of inquiries could probably be supported by using a teaching-learning sequence of different inquiry contexts in which students learn to evaluate the ARV of inquiries in a similar way. Therefore, the main aim of this research was to gain insight into how pre-university science students can increase their abilities to apply their procedural understanding on the evaluation of ARV in inquiries flexibly in different school science subjects. To reach this aim, an instrument with which pre-university science students can be supported in learning to evaluate the ARV of inquiries in different school science subjects had to be developed.

## 1.2 RESEARCH QUESTION

The main research question of the research was:

*What are the design characteristics of a feasible self-evaluation instrument and a supportive teaching-learning process (in which the instrument has to be used) with which pre-university science students can effectively learn to self-evaluate the accuracy, reliability and validity of inquiries in different school science subjects?*

In order to determine the feasibility of the self-evaluation instrument it was decided to investigate:

- 1) the feasibility of the self-evaluation instrument in evaluating the ARV in inquiries in different school science subjects by pre-university science students;
- 2) the feasibility of a design of a teaching-learning process in which the self-evaluation instrument could be used in a flexible way to learn to evaluate the ARV in different school science subjects; and
- 3) the effectiveness (regarding the learning outcomes) of evaluating the ARV in inquiries in different school science subjects by using the self-evaluation instrument in the designed teaching-learning process.

It was expected that the use of the same self-evaluation instrument in different inquiries would play a supportive role for the students in learning to evaluate the ARV of inquiries in the different school science subjects. Five empirical studies have been conducted in order to answer the above question. Section 1.4 shows the research questions in each of the five studies followed by an outline of the methodology applied. The five studies are described in detail in Chapters 2 to 6.

## 1.3 THEORETICAL BACKGROUND

Six important aspects of students' self-evaluation of ARV in inquiries are further elaborated and defined in this section: *learning to inquire (1.3.1), the meaning of accuracy, reliability and validity in the quality of an inquiry (1.3.2), the transfer of evaluation of accuracy, reliability and validity (1.3.3), formative feedback with self-evaluation instruments as rubrics (1.3.4), the content of a self-evaluation instrument (1.3.5), and the use of a self-evaluation instrument in a teaching-learning process (1.3.6).*

### 1.3.1 Learning to inquire

When defining the problem and aim of this research, the terms 'learning to inquire' and 'inquiry' were introduced. In this study, I chose to use the word 'inquiry' because the teaching-learning process as described in the relevant empirical studies includes practical investigations that focus on learning about the evaluation of ARV as well as thinking activities that enable students to apply the evaluation of ARV flexibly in inquiries in various school science subjects. Based on a constructivist view on learning, 'inquiry' can be seen as a multifaceted activity, while 'investigation' is more focused

on conducting an experiment (Harlen, 2012). Anderson (2007, p. 809) defined four elements of inquiries that accord with a constructivist view of learning:

- 1) 'Learning is an active process of individuals constructing meaning for themselves.
- 2) The meanings of each individual construct are dependent upon the prior conceptions the individual already has. In the process, these prior conceptions may be modified.
- 3) The understandings each individual develops are dependent upon the contexts in which these meanings are engaged.
- 4) Meanings are socially constructed; understanding is enriched by the engagement of ideas in concert with other people.'

In the National Science Education Standards of the USA the definition of learning to inquire focuses on the construction of evidence in science (NRC, 1996): 'Scientific inquiry refers to diverse ways in which scientists study the natural world and propose explanations based on evidence derived from their work' (p. 23). From this definition it can be deduced that learning to inquire requires students to undertake learning tasks about the construction of evidence and the development of scientific knowledge that is based on evidence from inquiries. The phases of an inquiry, such as posing inquiry questions, planning experiments or observations, performing experiments and observations, interpreting and analysing data, drawing conclusions and reflecting on the inquiry as a whole, can be considered as examples of these student tasks. Minner, Levy and Century (2010) concluded that students' active participation in all phases of an inquiry contributes to their understanding of science concepts. In addition, Harlen (2012) concluded that pre-university students make more progress in learning to inquire when they understand their actual levels of performance on different aspects of an inquiry and know how to improve their levels of performance when they are engaged in science that is related to situations in their daily lives.

These definitions of learning to inquire and previous research studies show that learning to inquire needs a complex set of learning tasks. When these definitions are applied to inquiries in a school science classroom setting that aims at increasing the students' flexibility in evaluating the ARV of inquiries in different school science subjects, it can be deduced that the tasks undertaken by students in the teaching-learning process should, at least:

- give students an active role in evaluating ARV;
- incorporate as many different phases of an inquiry as possible;
- be classified at various levels of understanding which clarify where students 'can head to';
- build on students' prior knowledge and attitudes about the evaluation of ARV in inquiries;
- make explicit how evaluating ARV in inquiries differs from one school science subject to the other;

- include sufficient opportunities for students to discuss with peers and teacher the ARV in inquiries; and
- include examples from scientific research studies which demonstrate to the students how scientists evaluate the ARV of their scientific research.

These aspects of learning to inquire are elaborated in more detail and further explored in the empirical studies (see Chapters 2 to 6).

### 1.3.2 Meaning of accuracy, reliability and validity in quality of inquiries

In the previous sections, accuracy, reliability and validity have been introduced as important concepts for evaluating the quality of inquiries. In everyday language, accuracy means ‘how exact something is’. When someone speaks about reliability it is mostly meant as, ‘something or someone can be trusted’. Validity, in daily situations, means: ‘something can be justified’ or (in Dutch) ‘someone is healthy’. These definitions of ARV can be expected as prior knowledge of pre-university students who have no experience in evaluating the ARV of inquiries.

Millar et al. (1994, p. 220) described accuracy as: ‘Understanding the appropriate degree of accuracy that is required to provide reliable data which will allow a meaningful interpretation’. For the purpose of the empirical studies in this thesis, which focus on three different school science subjects at pre-university level, evaluating the accuracy of an inquiry was defined as *evaluating the extent of the accuracy in the measurements and observations in an inquiry*. To improve the accuracy of an inquiry, it is, for example, necessary to determine whether the measuring instruments are adequate for measuring the maximum and minimum expected values, and whether the scale on the measuring instrument is accurate enough for the planned measurements. Accuracy can, for instance, be improved when two or more researchers, who work independently, conduct the same measurements or observations and when all different measurements or observations in the same experiment are done in the same systematic way (Boohan et al., 2010; Gott & Duggan, 2003; Gott et al., n.d.; Squires, 1985).

Gott et al. (n.d.) described reliability in their CoE model as: ‘Understanding the implications of the measurement strategy for the reliability of the resulting data; can the data be believed?’. For the purpose of the empirical studies in this thesis, evaluating the reliability of an inquiry was defined as *evaluating the extent of the reliability related to repeating the inquiry (repeatability) and reproducing the inquiry by others (reproducibility)*. ‘Results’ can be considered as the set of all relevant measuring values and observations of an inquiry. To evaluate the reliability of an inquiry, it is necessary to determine whether measurements or observations are repeated until the average measurement value or observation outcomes remain stationary. By performing a control experiment in which the research variables do not change it should be ascertained that variables other than the inquiry variables have no influence on the

results of the inquiry. Repetition of the inquiry by another, independent, researcher or the use of another inquiry method (triangulation) should lead to similar results. In addition, the sample of test subjects has to be sufficiently varied to be representative for the whole population and the sample should be large enough to lead to accurate measurement values or observations (Baarda & De Goede, 2001; Gott & Duggan, 2003; Gott et al., n.d.; Watson & Wood-Robinson, 2002).

Gott et al. (n.d.) described validity in their CoE model as: 'Understanding the implications of the design for the validity of the resulting data; an overall view of the task to check that it can answer the question'. For the purpose of the empirical studies in this thesis, evaluating the *validity* of an inquiry is defined as *evaluating the extent to which the design of the inquiry and the measurements and observations of variables lead to a valid conclusion with respect to the inquiry question*. An inquiry can be determined as valid when what is measured or observed is what should be measured or observed, and nothing else. To evaluate the validity of an inquiry, it is necessary to determine whether the inquiry is driven by a relevant, specific and concrete inquiry question and/or a testable hypothesis and whether the same independent and dependent variables are investigated throughout the whole inquiry. The chosen inquiry method should lead to results that are relevant for answering the inquiry question and sufficient to draw a conclusion. The conclusion should be fully based on the results of the inquiry (Gott & Duggan, 2003; Gott et al., n.d.; Roberts & Gott, 2002).

Accuracy, reliability and validity can be considered as three interconnected concepts that together lead to consistency in scientific evidence from inquiries. Therefore, *to determine the quality of an inquiry, the accuracy and reliability as well as the validity of the inquiry need to be evaluated*. According to Gott and Duggan (2003), an inquiry that is accurate and reliable as well as valid is a high quality inquiry. They make it a rule that an inquiry can be invalid but still reliable, but an inquiry cannot be valid if it is unreliable. In inquiries in which the variables can be controlled well these rules can be fulfilled: accurate measurement values lead to reliable results and as a consequence to a valid inquiry as a whole. This argumentation does not fit all inquiry contexts in the school science subjects. Especially in inquiry contexts with living test objects it is difficult to control all influencing variables. For example, in conducting a behavioural study on chimpanzees in a zoo, the influence of other animals cannot be controlled and a researcher cannot be sure about the particular state of mind of the chimpanzees during the observations. A sample of the chimpanzees in the zoo can lead to unreliable results, but as long as the researcher answers a research question on this specific group of chimpanzees, the inquiry can be valid (Pantin, 1968). It can be expected that above differences between inquiries with living test objects and with non-living test objects will confuse novices in evaluating the ARV of inquiries, because in the different school science subjects they perform both inquiries with non-living test objects

(e.g. in physics) and inquiries with living test objects (e.g. in biology). Therefore, in the tasks in the teaching-learning process as described in the studies of this thesis it has to be made clear for the pre-university students how the evaluation of ARV should be applied to inquiries with non-living test objects and those with living test objects.

### 1.3.3 Transfer of evaluation of accuracy, reliability and validity

As mentioned in section 1.3.1, evaluating ARV of inquiries in different school science subjects implies that pre-university science students know, or learn, how to transfer the evaluation of ARV from one inquiry to another. Salomon and Perkins (1989) distinguish two ways of transfer that are used in educational settings. The first, *low-road transfer*, focuses on extensive student exercises to foster the use of learned concepts in a context that is to a limited extent different from a previous context. The studies for this thesis, however, particularly focused on the second way of transfer: *high-road transfer*. High-road transfer is about the mindful abstraction of procedural understanding as learned in one inquiry to apply it in a new inquiry that is to a substantial extent different from a previous inquiry. According to Gagné, Yekovich and Yekovich (1993), high-road transfer can be stimulated, for example, when students are explicitly asked to write down and discuss issues with peers. At the same time, students should develop their conceptual understanding to facilitate high-road transfer. High-road transfer, and, more specifically, mindful abstraction, is an important concept in the research literature on learning psychology. For the purpose of this study, I decided to focus on the flexible application of procedural understanding as one aspect of mindful abstraction (Van Oers, 1998). This procedural understanding is more precisely defined in this study as the evaluation of ARV in inquiries in different school science subjects. Whether the achieved procedural understanding is applied to the evaluation of ARV in an inquiry in another school science subject depends on the student's motivation to transfer knowledge (e.g. regarding the concepts of ARV) to another inquiry as well as on the student's capacity to apply his or her procedural understanding (e.g. about evaluation of ARV) in a flexible way (Broad & Newstrom, 1992; Van Oers, 1998).

From previous research studies, four factors relevant to the capacity of students to apply procedural understanding flexibly were identified. First, students should have a relevant basic procedural understanding by which the process of transfer can be initialised. For this purpose, prior knowledge from everyday situations should not be neglected to support transfer, because this prior knowledge can confuse the students. Therefore, they should be guided to build the new (scientific) procedural understanding about ARV of inquiries upon their previous (daily-life) understanding of ARV. The teacher should also be aware that some of the previous knowledge students bring to a new inquiry might impede their learning, for example, because of alternative conceptions (Bransford, 2000; Chen, Yanowitz, & Daehler, 1995; Novick, 1988).

The second factor that supports a flexible application of learned procedural understanding is that procedural understanding is taught in a variety of inquiries. By learning in a variety of inquiries in school science subjects, students can learn when, where and how to apply their procedural understanding in different inquiries. How many different inquiries should be undertaken depends on the degree of complexity of the procedural understanding that students need to apply flexibly in an inquiry. The higher this complexity, the more inquiries (or more extensive inquiries) are needed for high-road transfer to take place. For such transfer to occur, it is crucial that the topics of the inquiries are sufficiently related – from the perspective of the students (Beishuizen & Asscher, 2001; Bransford, 2000; Gilbert, Bulte, & Pilot, 2011).

The third factor is about the recognition of subject-specific patterns by the students. Novices, like pre-university science students, in evaluating ARV in an inquiry should first perform learning tasks in science subjects in which they can recognise subject-specific patterns before they can learn to make explicit what the differences and similarities are in the evaluation of ARV in inquiries in other science subjects (Bransford, 2000; Gilbert, Bulte, & Pilot, 2011).

Fourth, flexible application of procedural understanding requires a (classroom) setting in which students have an active role in learning and transferring their procedural understanding about the ARV of inquiries. In this active process, students apply their procedural understanding to new inquiries and have the opportunity to ask their questions, whereafter the teacher has to decide what should be explained to individual students and what should be elucidated to the whole class (Bransford, 2000; Bransford & Schwartz, 1999).

The four factors described above could facilitate the flexible application of procedural understanding about the evaluation of ARV. These factors were elaborated in the teaching-learning process that is designed for the empirical studies as described in Chapters 3 to 6.

#### **1.3.4 Formative feedback with self-evaluation instruments as rubrics**

An active role of students in the flexible application of procedural understanding (see 1.3.3) can be stimulated by giving students the opportunity for self-evaluation of their performance. Andrade and Valtcheva (2009) showed that the use of a self-evaluation instrument can be helpful in the flexible application of acquired knowledge and skills to another, new situation. They concluded that the use of a self-evaluation instrument forces students to reflect on their knowledge and skills, to revise their performance and to formulate steps to improve their performance. The same self-evaluation instrument should be supportive in different learning tasks. As a consequence, self-evaluation instruments need to give students feedback on the performance itself to enable them

to judge their own level of performance. Such a self-evaluation instrument can also be a useful tool to provide formative feedback by peers and teachers on the performances that were first self-evaluated by the students (Andrade & Valtcheva, 2009).

As Black and William (2009) stated, formative assessment focuses on evaluations made by students (and teachers) that guide the next (improving) steps in their performance, in contrast to summative purposes of assessment, which requires teacher's judgements against standards without next steps. Crucial in formative feedback is the central, active role of the students by which they develop self-regulation in their learning. Formative feedback contains two sequential core activities. First, a student should become aware of the gap between the goal and his or her present practical skills, conceptual and/or procedural understanding. Then, the student must undertake action to close the gap in order to attain the goal. Consequently, by using a self-evaluation instrument to give formative feedback, the future learning of students can be adapted to their individual needs (Bell, 2007; Black & William, 2009; Harrison, 2012).

Furtak and Ruiz-Primo (2008) describe a continuum of formative feedback inspired by previous research studies. At one end is formal formative feedback which is planned feedback activity using a (self-)evaluation instrument. The feedback activity can be performed by the students themselves, by their peers and by their teacher. Teachers also can perform planned formative feedback to inform themselves about the results of their teaching and the next steps that should be taken. At the other end is informal formative feedback in which teachers are continuously noticing, recognising and responding to questions and suggestions of students in so-called 'teacher-student interactions'. Teachers often perform this interactive formative feedback to support the learning of individual students with respect to intended learning. Black and William (2009) conclude that it is difficult for anyone other to discern the thinking behind a student's response other than the student him- or herself, especially when feedback is given on written materials. They suggest that during both formal and informal formative feedback activities in the classroom, peers and the teacher should try to clarify all students' responses by asking students to explain them. As Van de Pol, Volman and Beishuizen (2012) stated in their research on scaffolding, the responsibility for learning can be given to the students under the condition that a teacher supports the students' learning by contingent scaffolding. A (self-)evaluation instrument that is understood and used in the same way by both the students and the teacher seems to be a helpful tool in supporting contingent scaffolding of learning processes. In the model of Van de Pol et al. (2012) about contingent teaching, the teacher first has to use diagnostic strategies (e.g. by using an evaluation instrument) to determine the level of performance of the students, whereafter he or she can choose which support students need to improve their performance. Such an evaluation instrument can also be used by students to evaluate and improve their own performances.

When developing a (self-)evaluation instrument for formative feedback on the ARV of an inquiry, two questions should be kept in mind: (1) What should be evaluated? and (2) How can procedural understanding be evaluated and not just the recall and understanding of the general meanings of accuracy, reliability and validity? (Bell, 2007; Black & William, 2009). Rubrics, as instruments in which different levels of student performance are described, can be used to give formative feedback on the ARV of an inquiry. Use of rubrics increases transparency of performance criteria to students (Panadero & Jonsson, 2013). In the past decades, many different rubrics have been designed by teachers and educational designers and researchers to be used as formative feedback self-evaluation instruments with qualitative descriptions of (levels of) performance. However, many rubrics for secondary and higher education consist of ambiguous descriptions of performance levels of skills and strategies across their scale levels (Tierney & Simon, 2004). The review study of Jonsson and Svingby (2007) shows that most rubrics focus on the assessment of the content of students' products (essays, reports) rather than on the processes or strategies used by students. In particular, it is not known which characteristics of a set of rubrics can help to improve the strategies of students in evaluating the ARV in inquiries in different school science subjects. Therefore, in this study, for the design of the rubrics as a formative feedback self-evaluation instrument I sought design characteristics suitable to describe the levels of performance of the evaluation of ARV. These descriptions of the levels of performance had to be feasible for flexible application of the evaluation of ARV in different school science inquiries (see Chapters 3 and 4).

### 1.3.5 Content of a self-evaluation instrument

The effectiveness of the use of rubrics regarding the learning outcomes of the students in evaluating the ARV of inquiries depends on the quality of the descriptions in the rubrics. To show the successive levels of performance in evaluating the ARV of inquiries, all descriptions in the rubrics should be represented hierarchically (Arter & McTighe, 2001; Moskal, 2000). As mentioned in section 1.3.4, despite the frequent use of rubrics in secondary and higher education, most rubrics contain ambiguous descriptions of levels of performance. The descriptions are based on the limited experiences of the teachers and not on a scientifically tested classification system. This implies that an empirically based taxonomy is needed that describes the levels of performance in a more principled and hierarchical way (Jonsson & Svingby, 2007; Tierney & Simon, 2004).

Chan, Tsui, Chan and Hong (2002) explored the application of three educational taxonomies – the Structure of Observed Learning Outcomes (SOLO) taxonomy, Bloom's taxonomy and a reflective thinking measurement model – in measuring students' cognitive learning outcomes. They found that the SOLO taxonomy is preferable for monitoring different kinds of learning outcomes, for example, in practice-oriented

studies. Changes of domain-specific topics in which the SOLO taxonomy was used did not seem to reduce the effectiveness of the taxonomy in determining the learning outcomes of students. Also, the SOLO taxonomy can be applied to undergraduate students (Chick, 1998). As such, it is assumed that the SOLO taxonomy is suitable for the studies in this thesis, which focused on the flexible application of the evaluation of ARV. The SOLO taxonomy focuses on the levels of intended learning outcomes and is supportive for students in evaluating their performance at particular points in a learning task (Biggs & Tang, 2007; Hodges & Harvey, 2003; Levins & Pegg, 1993).

The SOLO taxonomy consists of five levels that increase in complexity: prestructural, unistructural, multistructural, relational and extended abstract (Chan et al., 2002). In this study, the prestructural level is defined as using the concepts of accuracy, reliability and validity in everyday language, or in daily life situations. Often, students use tautology to cover their lack of understanding, for instance by saying: 'I have measured accurately during the whole inquiry'. The unistructural level is defined as using only one relevant aspect that is mostly based on quoting or memorisation, whereas in the multistructural level students use various relevant aspects but ignore any inconsistencies or relations between these aspects. At the relational level, the students use the inconsistencies but come to a firm conclusion, whereas in the extended abstract level students recognise that new hypotheses can occur and that their inquiry is an example of a more general case. Based on the theory behind the SOLO taxonomy, the multistructural, relational and extended abstract levels of a rubric need to be built hierarchically on the unistructural level. When the SOLO taxonomy is properly applied to the content of each rubric in the set of rubrics, a student can only reach the relational level when the multistructural level is met completely (Biggs & Tang, 2007; Chan et al., 2002). Hence, in the empirical studies of this thesis the SOLO taxonomy was used to describe the levels of performance in each of the rubrics of a self-evaluation instrument and to design a feasible self-evaluation instrument for a flexible application of the evaluation of ARV in inquiries in the different school science subjects.

### 1.3.6 Use of a self-evaluation instrument in a teaching-learning process

For students to use a self-evaluation instrument effectively for a flexible application of the evaluation of ARV in inquiries, a feasible teaching-learning process is also required. When students are asked to use a self-evaluation instrument to reflect on their own performance in order to improve, it is important that there is consistency between the intended functions of the instrument and the teaching-learning process in which the instrument will be used. If students interpret and enact a teaching-learning process in a completely different way than was intended when it was designed, they will not understand the supportive role of the learning materials used to measure their learning (e.g. a self-evaluation instrument). The learning materials will lose their

function; the teaching-learning process will be ineffective and it will therefore not lead to appropriate student learning outcomes (Ledford & Sleeman, 2000).

With regard to the intended practical use of the self-evaluation instrument in class to learn to apply the evaluation of ARV of inquiries flexibly, students can use the instrument for self-evaluation, and also peers and teachers can do the same evaluation (Andrade & Du, 2005; Nicol & Boyle, 2003). Consequently, in the teaching-learning process that was designed for the empirical studies in this thesis, effective formative feedback on a student's application of the ARV of an inquiry can be yielded when the same self-evaluation instrument is used (i) by the students themselves, (ii) by peers and teacher to give feedback on the students' performance, and (iii) in inquiries in different school science subjects. This leads to three formative feedback functions intended for this instrument in the teaching-learning process:

- 1) Self-evaluation of students' performance in applying ARV
- 2) Support of self-evaluation by peers and teacher
- 3) Facilitating of transfer of self-evaluation of ARV in different school science inquiries.

These three functions are supported by the principles of good formative feedback as described in a review study on formative assessment by Nicol and Macfarlane (2006). They define 'good feedback practice' as 'anything that might strengthen the students' capacity to self-regulate their own performance' (p. 205). They consider self-evaluation as an aspect that contributes to self-regulation of learning processes by students. Self-regulation is a complex process for students to learn, due to the combination of cognitive, motivational and contextual aspects (Pintrich, 2000). Nicol and Macfarlane (2006) gained insight into the various aspects in learning to self-regulate learning processes. In their words (p. 205), good feedback practice:

- 1) helps to clarify what good performance is (goals, criteria, expected standards);
- 2) facilitates the development of self-assessment (reflection) in learning;
- 3) delivers high quality information to students about their learning;
- 4) encourages teacher and peer dialogue around learning;
- 5) encourages positive motivational beliefs and self-esteem;
- 6) provides opportunities to close the gap between current and next performances;
- 7) provides information to teachers that can be used to help shape teaching.

These seven principles seem to support the three functions of a self-evaluation instrument in a teaching-learning process that intends to teach students to self-evaluate their inquiry performance. The principles are further described and elaborated in the empirical studies that are described in Chapter 5 (principles 1, 2, 3, 4, 6, 7) and Chapter 6 (principle 5).

### 1.4 THESIS OUTLINE

This thesis reports on a design-based research study on learning to inquire in different school science subjects by pre-university science students, and in particular on evaluating the ARV of inquiries with a self-evaluation instrument. This study was performed in three parts: an explorative study to gain insight into pre-university students' and science teachers' knowledge about aspects of evaluating the ARV of inquiries and two cycles of designing, testing and evaluating an intervention. In cycle 1, the feasibility of a set of rubrics to evaluate the ARV of inquiries in three school science subjects was studied. In cycle 2, the research focused on the students' use of the revised set of rubrics, the accompanying teaching-learning process with inquiry learning tasks designed for the study, and the students' learning outcomes.

Five empirical studies were performed to investigate the three research parts. Figure 1.2 displays an overview of the research and the corresponding chapters in this thesis.

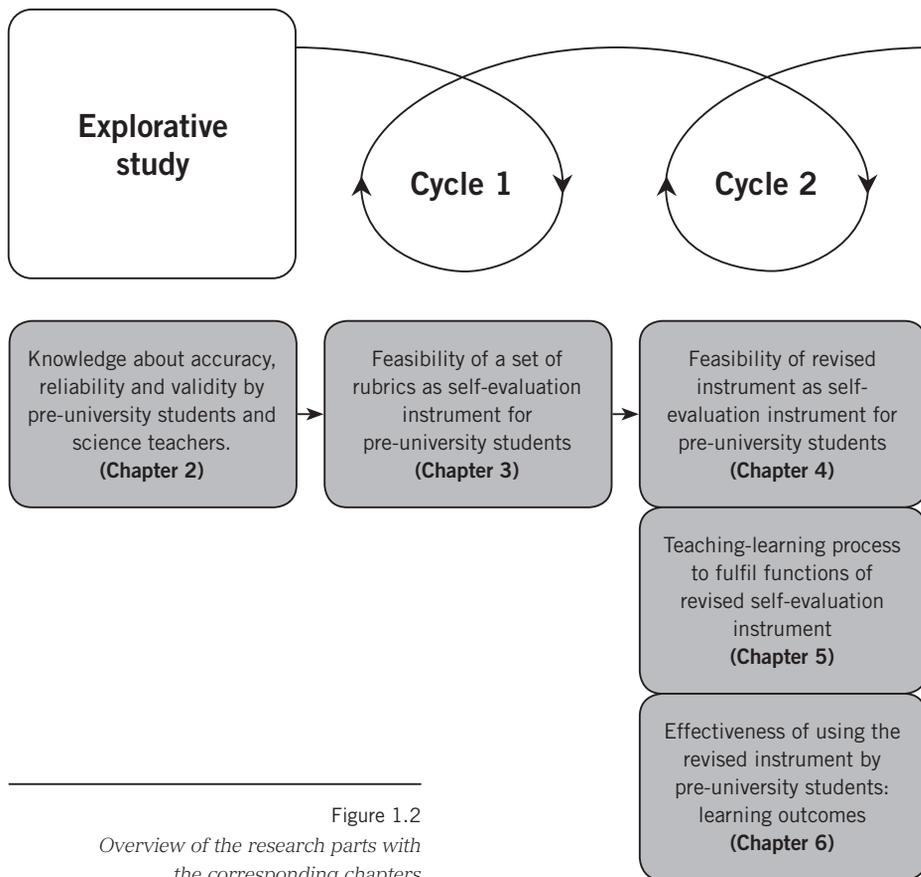


Figure 1.2  
Overview of the research parts with  
the corresponding chapters

The **explorative study**, as described in **Chapter 2**, was focused on the knowledge held by pre-university science students and teachers of biology, physics and chemistry about evaluating the ARV of inquiries. This study was guided by three research questions:

- (1) *To what extent do pre-university science students and biology, physics and chemistry teachers recognise CoE that can be related to accuracy, reliability and validity in a student's inquiry?*
- (2) *What is the accordance between biology, chemistry and physics teachers in the recognition of CoE in relation to accuracy, reliability and validity?*
- (3) *What do pre-university science students know about the meaning of accuracy, reliability and validity in inquiries?*

This qualitative research (Denscombe, 2007) involved a think-aloud task (Bowen, 1994) followed by an interview with science teachers (n=6) and pre-university science students (n=6), and a questionnaire that was filled out by pre-university science students (n=38). The answers from the interviews and questionnaires were analysed on:

- (a) use of correct and relevant CoE;
- (b) use of incorrect, but relevant CoE;
- (c) use of everyday language to describe the concept being asked about;
- (d) nonsense/ambiguous answer; or
- (e) no answer.

In the **first research cycle** (see **Chapter 3**), four design characteristics were identified from the explorative study and the literature to design a set of rubrics that was expected to function as a self-evaluation instrument for pre-university science students to evaluate ARV in the inquiries. To test the feasibility of the set of rubrics, a series of three successive inquiry units was designed and implemented in class. This first test cycle was guided by the research question: *To what extent are the design characteristics essential and sufficient for designing a set of rubrics that is feasible for pre-university science students to self-evaluate accuracy, reliability and validity in successive science inquiry units?*

In this study, 24 pre-university science students, aged 16 or 17, and two science teachers from one upper secondary school participated. The students worked in groups on three successive inquiry units in general science, biology and physics, respectively. Student pairs as well as individual students were taken as the unit of analysis (Cole & Engeström, 1993). A qualitative research method (Cohen & Manion, 1994) with triangulation of data (Yin, 2003) was used to determine the feasibility of the set of rubrics. The latter was based on four criteria (Nieveen, 2009):

- (a) each rubric is used as intended;
- (b) the students and teacher can work with all the rubrics;

- (c) the rubrics support the determination of the level of student understanding; and
- (d) the rubrics lead to a positive change in use of 'scientific' terminology by students and teachers.

The collected data sources were: field reports of classroom observations, video recordings of all lessons, audio recordings of discussions in four student groups, completed rubrics of all student groups and those of the researchers (including notes made on the feasibility), completed student worksheets with inquiry plans, inquiry data and results, conclusions and evaluations of ARV. Immediately after finishing an inquiry unit, all students individually completed a questionnaire about their opinions on the rubrics. Four students were interviewed after completing the questionnaires and the teachers were interviewed after each period in which the rubrics were used. All interviews were audio-recorded and transcribed. All data sources were analysed on the above described criteria (a)–(d). Based on the outcomes of the study, the set of rubrics was revised for use in the second research cycle.

The **second research cycle**, as described in **Chapters 4, 5 and 6**, aimed at determining the feasibility and effectiveness of the revised self-evaluation instrument, now called the Evaluation of Quality of Inquiries (EQI) instrument in a designed teaching-learning process. This revised EQI instrument contained rubrics, a checklist and an ARV card that gives students an overview. In this study, 27 pre-university science students, aged 16 or 17, and a qualified biology teacher participated. The students worked in twelve groups on three successive inquiry units in general science, biology and physics during seven afternoon sessions. In the last session they completed an assessment task on two inquiries in chemistry.

The study in **Chapter 4**, on the feasibility of the revised EQI instrument, was guided by the question: *What is the feasibility of the EQI instrument for the evaluation of the accuracy, reliability and validity in inquiries in different school science subjects by pre-university science students?*

The feasibility of the EQI instrument was determined by a formative evaluation as part of a design-based research (Van den Akker, Gravemeijer, McKenney, & Nieveen, 2006). A qualitative method (Cohen & Manion, 1994) with triangulation of data (Yin, 2003) was used. The collected data sources in this study were: field reports of classroom observations, video recordings of all lessons, reflection reports of the teacher, all completed student and teacher checklists, rubrics and ARV cards, students' inquiry plans and worksheets. Immediately after each session, the students individually completed a questionnaire on the feasibility of the instrument. One week after the last session, all the students were interviewed in pairs on their experiences with

the feasibility of the instrument. All interviews were audio-recorded and transcribed. The data sources were analysed on four criteria (Nieveen, 2009):

- (a) the EQI instrument is used as intended;
- (b) the students and the teacher can work with the EQI instrument;
- (c) the EQI instrument supports the determination of the level of student understanding; and
- (d) the EQI instrument leads to a positive change in use of 'scientific' terminology by students and teachers.

The study as described in **Chapter 5** focused on the design characteristics of a teaching-learning process in which the three functions of the designed EQI instrument could be fulfilled. These three functions were:

- 1) self-evaluation of ARV by students;
- 2) support of self-evaluation by peers and teacher; and
- 3) facilitation of transfer of self-evaluation of ARV to different inquiries.

This study was guided by the research question: *To what extent does the designed teaching-learning process fulfil the three intended functions of the EQI instrument when this teaching-learning process is enacted in class?*

A design-based research approach (Gravemeijer & Cobb, 2006) was used. In two micro design cycles, the design characteristics emerged. Each microcycle contained a design phase and a phase of analysis of and reflection on the educational design. In the data collection, triangulation of data sources was used (Yin, 2003). The collected data sources in this study were: field reports from observations, reflection reports of the teacher, completed rubrics, checklists and the ARV cards of the students, completed students' workbooks, students' questionnaires, video recordings of the teaching-learning process in class, and audio recordings of discussions between students and teacher during the teaching-learning process. Relevant parts of the recordings were transcribed. One week after the last session, all the students were interviewed in pairs about their experiences using the EQI instrument in the teaching-learning process. All interviews were audio-recorded and transcribed. The data analysis focused on eleven indicators (see Table 5.1, p. 104) in the teaching-learning process for the three functions of the EQI instrument.

The research as described in **Chapter 6** aimed at determining the learning outcomes of the students when using the EQI instrument in the teaching-learning process. This study was guided by the research question: *What is the effectiveness of the EQI instrument in the transfer of ARV evaluation skills in various inquiries performed by pre-university science students?*

To determine the effectiveness of the EQI instrument, the following sub questions were set:

- (1) *To what extent do students perceive that the EQI instrument is relevant for evaluating the ARV in an inquiry?*
- (2) *To what extent do students have sufficient knowledge to evaluate ARV in an inquiry?*
- (3) *To what extent can students apply this achieved knowledge on how to evaluate ARV in a, for them, new inquiry with the same or a higher level in the SOLO taxonomy?*
- (4) *To what extent can students, without any guidance, transfer achieved knowledge on the evaluation of ARV to a, for them, new inquiry with a sufficient level in the SOLO taxonomy?*

The effectiveness of the use of the EQI instrument was determined by a mixed methods research approach (Denscombe, 2007) with triangulation of data (Yin, 2003). The collected data sources in this study were: students' questionnaires, completed rubrics, checklists and ARV cards of students and researchers, and students' worksheets with the inquiry question, hypotheses and the inquiry methods of the successive inquiry units including the data sources of the assessment task. After the last afternoon session, all students were interviewed in pairs. All the interviews were audio-recorded and transcribed. To determine the effectiveness of the EQI instrument, the data sources were analysed on four learning aims:

- (1) the students perceive relevance when they learn how to evaluate the ARV in an inquiry and perceive that the EQI instrument is helpful to do so;
- (2) the students learn enough items of the CoE model and are able to evaluate ARV in an inquiry adequately;
- (3) the students use, with support from teacher and peers, items of the CoE model in a new inquiry at the same or a higher SOLO taxonomy level; and
- (4) the students transfer, without guidance from teacher or peers, items of the CoE model to a new inquiry at a sufficient level in the SOLO taxonomy.

In **Chapter 7**, the main findings and conclusions of the five empirical studies are summarised. Thereafter, the main research question about the feasibility and effectiveness of the self-evaluation instrument is answered, followed by a reflection on the outcomes and the methodology used. Finally, the implications of this design-based research study for educational research and educational classroom practice are discussed. Chapter 7 ends with recommendations for further research on how pre-university science students can learn to evaluate ARV in inquiries.

## 1.5 CONTEXT AND RELEVANCE

This PhD study was funded by the Platform Bèta Techniek, as one of the studies in the DUDOC research programme. This programme commenced in 2007 to fund research studies on the feasibility and effectiveness of standards in the new Dutch formal curricula of biology, chemistry, physics and mathematics at pre-university level. These new formal curricula (except for mathematics) prepare students for the pre-university level exams from 2016 on.

This thesis aims at contributing to the scientific knowledge on learning to inquire and on the educational practice of student learning to inquire at pre-university level. The contributions to scientific knowledge lie in different fields. First, this study aimed at gaining insight into the flexible application of the concepts of evidence (CoE) that are related to the evaluation of ARV in inquiries in three different school science subjects. These CoE were originally described as a student's procedural understanding in inquiries in physics and were further studied in other school sciences (e.g. Schalk, 2006; Van Rens, 2005). However, not much is known about the successive use of these CoE in different school science subjects by the same students and the influence on their procedural understanding. Second, this study aimed at gaining more insight into the design characteristics of a self-evaluation instrument, and especially rubrics, with which students are able to apply their knowledge of ARV in evaluations of inquiries flexibly in different school science subjects. Finally, it also aimed at shedding more light on the design characteristics of a feasible teaching-learning process in which a self-evaluation instrument for the evaluation of ARV in inquiries can be enacted in class.

Regarding the educational practice, this study aimed at contributing to the implementation of the new formal curricula of biology, chemistry and physics. Focusing on the evaluation of the ARV in inquiries is one of the standards added in these new formal curricula. Despite these new curricula, most student inquiries at pre-university level in the Netherlands are not yet focused on the evaluation of the ARV of an inquiry. Pre-university science students and many of their science teachers are novices in evaluating the ARV of school science inquiries. Therefore, this study aimed at designing and testing educational materials with which pre-university science students can learn how to evaluate the ARV of inquiries in the school science subjects of biology, chemistry and physics. These materials were also designed and tested to provide science teachers with tools to implement the newly added standards about evaluation of ARV in inquiries in school science curricula.

## REFERENCES

- Aarsen, M., & Van der Valk, T. (2008). Onderzoekende houding, een leerlijn [Investigative approach, a teaching-learning trajectory]. *NVOX*, 33(8), 354–356.
- Abd-El-Khalick, F., BouJaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., Tuan, H. (2004). Inquiry in science education: International perspectives. *Science Education*, 88(3), 397–419.
- Anderson, R. D. (2007). Inquiry as an organizing theme for science curricula. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 807–830). Mahwah, NJ: Lawrence Erlbaum Associates.
- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10(3).
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into Practice*, 48(1), 12–19.
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousand Oaks, CA: Corwin Press.
- Biggs, J., & Tang, C. (Eds.). (2007). *Teaching for quality learning at university* (3rd ed.). Buckingham, UK: Open University Press.
- Baarda, D. B., & De Goede, M. P. M. (2001). *Basisboek methoden en technieken. Handleiding voor het opzetten en uitvoeren van onderzoek. [Handbook of methodology. Manual for design and performance of research studies]*. Groningen, The Netherlands: Stenfert Kroese.
- Beishuizen, J. J., & Asscher, J. J. (2001). *Syllabus Onderwijspsychologie [Educational psychology workbook]* (4th ed.). Leiden, The Netherlands: Department of Psychology, University of Leiden.
- Bell, B. (2007). Classroom assessment of science learning. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on education* (pp. 965–1006). Mahwah, NJ: Lawrence Erlbaum Associates.
- Black, P., & William, D. (2009). Developing a theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Boohan, R., Campbell, P., Charlesworth, J., Galloway, I., Millar, R., Needham, R., & Roberts, R. (2010). *The language of measurement. Terminology used in school science investigation*. Hatfield, UK: The Association for Science Education.
- Bowen, C. W. (1994). Think-aloud methods in chemistry education. *Journal of Chemical Education*, 71(3), 184–190.
- Bransford, J. D. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, DC: National Academy Press.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24(40), 61–100.
- Broad, M. L., & Newstrom, J. W. (1992). *Transfer of training: action-packed strategies to ensure high payoff from training investments*. Reading, MA: Addison-Wesley.
- Chan, C.C., Tsui, M. S., Chan, M.Y.C., & Hong, J.H. (2002). Applying the Structure of the Observed Learning Outcomes (SOLO) taxonomy on student's learning outcomes: An empirical study. *Assessment & Evaluation in Higher Education*, 27(6), 511–527.
- Chen, Z., Yanowitz, K. L., & Daehler, M. W. (1995). Constraints on accessing abstract source information: Instantiation of principles facilitates children's analogical transfer. *Journal of Educational Psychology*, 87(3), 445–454.
- Chick, H. (1998). Cognition in the formal modes: research mathematics and the SOLO taxonomy. *Mathematics Education Research Journal*, 10(2), 4–26.
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218.
- Cohen, L., & Manion, L. (1994). *Research methods in education* (4th ed.). London: Routledge.
- Cole, M., & Engeström, Y. (1993). A cultural-historical approach to distributed cognition. In G. Salomon (Ed.) *Distributed cognitions. Psychological and educational considerations* (pp. 1–46). Cambridge, UK: Cambridge University Press.

- Denscombe, M. (2007). *The good research guide for small-scale social research projects* (3rd ed.). Maidenhead, UK: Open University Press.
- Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education*, 92(5), 799–824.
- Gagné, E. D., Yekovich, C. W., & Yekovich, F. R. (1993). *The cognitive psychology of school learning* (2nd ed.). New York: HarperCollins.
- Gilbert, J. K., Bulte, A. M. W., & Pilot, A. (2011). Concept development and transfer in context-based science education. *International Journal of Science Education*, 33(6), 817–837.
- Gott, R., & Duggan, S. (1995). *Investigative work in the science curriculum*. Buckingham, UK and Philadelphia, PA: Open University Press.
- Gott, R., & Duggan, S. (2003). *Understanding and using scientific evidence*. London: SAGE Publications.
- Gott, R., Duggan, S., Roberts, R., & Hussain, A. (n.d.). Research into understanding scientific evidence. Retrieved 5 November 2013, from <http://www.dur.ac.uk/rosalyn.roberts/Evidence/cofrev.htm>
- Gravemeijer, K., & Cobb, P. (2006). Design research from a learning design perspective. In J. Van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational Design Research* (pp. 17–51). London and New York: Routledge.
- Harlen, W. (2012). Scientific inquiry. In J. Oversby (Ed.), *ASE Guide to research in science education* (pp. 98–105). Hatfield, UK: The Association for Science Education.
- Harrison, C. (2012). Assessment for learning: Classroom practices that engage a formative approach. In J. Oversby (Ed.), *ASE Guide to research in science education* (pp. 142–147). Hatfield, UK: The Association for Science Education.
- Hodges, L. C., & Harvey, L. C. (2003). Evaluation of student learning in organic chemistry using the SOLO taxonomy. *Journal of Chemical Education*, 80(7), 785–787.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
- Ledford, B. R., & Sleeman, P. J. (2000). *Instructional design: A primer*. Greenwich, CT: Information Age Publishing.
- Levins, L., & Pegg, J. (1993). Students' understanding of concepts related to plant growth. *Research in Science Education*, 23, 165–173.
- Lunetta, V. N., Hofstein, A., & Clough, M. P. (2007). Learning and teaching in the school science laboratory: an analysis of research, theory, and practice. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 393–442). Mahwah, NJ: Lawrence Erlbaum Associates.
- Millar, R. (2010). *Analysing practical science activities to assess and improve their effectiveness*. Hatfield, UK: The Association for Science Education.
- Millar, R., Lubben, F., Gott, R., & Duggan, S. (1994). Investigating in the school science laboratory: conceptual and procedural knowledge and their influence on performance. *Research Papers in Education*, 9(2), 207–248.
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction: What is it and does it matter? Results from a research synthesis years 1984–2002. *Journal of Research in Science Teaching*, 47(4), 474–496.
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? [Electronic Version]. *Practical Assessment, Research & Evaluation*, 7(3). Retrieved 5 November 2013 from <http://PARonline.net/getvn.asp?v=7&n=3>.
- Nicol, D.J., & Boyle, J.T. (2003). Peer instruction versus class-wide instruction in large classes: a comparison of two interaction methods in a wired classroom. *Studies in Higher Education*, 28(4), 457–473.
- Nicol, D. J., & Macfarlane, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218.
- Nieveen, N. (2009). Formative evaluation in educational design research. In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research*. Enschede, The Netherlands: SLO.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology*, 14(3), 510–520.

- NRC (1996). National Science Education Standards. Retrieved 10 July 2013, from <http://www.csun.edu/science/ref/curriculum/reforms/nses/nses-complete.pdf>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(1), 129–144
- Pantin, C. F. A. (1968). *The relations between the sciences*. London and New York: Cambridge University Press.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P.R. Pintrich & M. Zeidner (eds.) *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.
- Roberts, R., & Gott, R. (2002). Investigations: Collecting and using evidence. In D. Sang & V. Wood-Robinson (Eds.), *Teaching secondary scientific enquiry*. London: Association for Science Education.
- Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, 24(2), 113–142.
- Schalk, H. H., Van der Schee, J., & Boersma, K. Th. (2013). The development of understanding of evidence in pre-university biology education in the Netherlands. *Research in Science Education*, 43(2), 551–578.
- Schalk, H. H. (2006). Zeker weten? Leren de kwaliteit van biologie-onderzoek te bewaken in 5 vwo. [Are you certain? Learning to ensure the quality of biology research in pre-university education – with summary in English]. Vrije Universiteit, Amsterdam.
- Squires, G. L. (1985). *Practical physics* (3rd ed.). Cambridge, UK: Cambridge University Press.
- Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, 9(2). Retrieved 5 November 2013 from <http://PAREonline.net/getvn.asp?v=9&n=2>.
- Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational design research*. London: Routledge.
- Van Oers, B. (1998). From context to contextualizing. *Learning and Instruction*, 8(6), 473–488.
- Van de Pol, J., Volman, M., & Beishuizen, J. (2012). Promoting teacher scaffolding in small-group work: A contingency perspective. *Teaching and Teacher Education*, 28, 193–205.
- Van Rens, L. (2005). Effectief scheikundeonderwijs voor 'leren onderzoeken' in de tweede fase van het vwo. [Effective chemical education for learning to inquire in upper secondary schools – with summary in English]. Vrije Universiteit, Amsterdam.
- Van Rens, L., Van Muijlwijk, J., Beishuizen, J., & Van der Schee, J. (2013). Upper secondary chemistry students in a pharmacology research community. *International Journal of Science Education*, 35(6), 1012–1036.
- Watson, R., & Wood-Robinson, V. (2002). Investigations: Evaluating evidence. In D. Sang & V. Wood-Robinson (Eds.), *Teaching secondary scientific enquiry* (pp. 89–107). London, UK: Association for Science Education.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.



---

GENERAL INTRODUCTION

