Explorative study

Cycle 1

Cycle 2

Knowledge about accuracy, reliability and validity by pre-university students and science teachers **(Chapter 2)**	Feasibility of a set of rubrics as self-evaluation instrument for pre-university students **(Chapter 3)**	Feasibility of revised instrument as self-evaluation instrument for pre-university students **(Chapter 4)**

Teaching-learning process to fulfil functions of revised self-evaluation instrument
(Chapter 5)

Effectiveness of using the revised instrument by pre-university students: learning outcomes
(Chapter 6)

CoE model
82 CoE ➤ 47 relevant for evaluation ARV in inquiries

Conclusion
23 CoE items relevant for evaluation ARV in inquiry by novices

Design
23 CoE items in 19 rubrics

Conclusion
21 CoE items relevant:
13 CoE items in rubrics
all items in holistic ARV card and checklist
(=EQI instrument)

Design
EQI instrument (ch.4)
Teaching-learning process (ch.5)
Effectiveness of EQI (ch.6)

Conclusion
EQI is feasible (ch. 4)
Teaching-learning process partly fulfilled (ch. 5)
Effective to evaluate ARV (ch.6)

CONCLUSION, DISCUSSION AND RECOMMENDATIONS

CHAPTER 7

Chapter 7 gives an overview and synthesis of, and a reflection on, the outcomes of the five studies as presented in Chapter 2 to 6. In section 7.1 the aim and main research question of this thesis about learning to evaluate accuracy, reliability and validity (ARV) in school science inquiries is addressed. In section 7.2, the main results and conclusions of the explorative study, the first research cycle and the three studies of the second research cycle, as described in Chapter 2 to 6, are summarised. In section 7.2 also an overview is given of the four design characteristics of the Evaluation of Quality of Inquiries (EQI) instrument and of the 15 design characteristics of the supporting teaching-learning process. As described in section 7.3, during the second research cycle both sets of design characteristics were shown to be feasible for designing an effective teaching-learning process for evaluating the ARV in inquiries in different school science subjects by pre-university students with the EQI instrument. Section 7.4 provides a reflection on the conclusions of the studies in this thesis and on the methodology used. Section 7.5 elaborates on two important limitations of this research: the limited number of participating students and the special role of the teacher in the studies reported here. In sections 7.6 and 7.7 the implications for science education research and for pre-university science education practice are described and discussed. This chapter ends with recommendations for further research on learning to inquire (section 7.8).

7.1 AIM OF THIS STUDY AND MAIN RESEARCH QUESTION

Pre-university science students in the Netherlands perform inquiries as part of the school subjects of biology, chemistry and physics. As a result of a recent change of the formal curriculum, in all three school science subjects, students are required

to learn how to evaluate the accuracy, reliability and validity (ARV) in inquiries. As described in the introduction of this thesis (see Chapter 1), not much research has as yet been done on how to enhance the coherent procedural understanding of students of evaluating the ARV in inquiries in biology, physics and in chemistry. It is also not clear how students can be supported during a teaching-learning process to transfer ARV evaluation skills flexibly from one subject to another. It was expected that a self-evaluation instrument based on the Concepts of Evidence (CoE) model (Gott, Duggan, Roberts, & Hussain, n.d.) and the levels of performance as described in the Structure of Observed Learning Outcomes (SOLO) taxonomy (Biggs and Tang, 2007) could be useful in guiding a flexible application of ARV evaluation skills to new subjects. A self-evaluation instrument was designed, based on the seven principles of good formative feedback (Nicol & Macfarlane, 2006) serving three formative feedback functions in a teaching-learning process aimed at learning to evaluate the ARV in inquiries in different school science subjects:

    (1) self-evaluation of ARV by students;

    (2) support of self-evaluation by peers and teacher; and

    (3) facilitating transfer of self-evaluation of ARV to inquiries in different school science subjects.

Consequently, the overarching aim of the research in this thesis was to gain insight into how pre-university science students can be supported in the process of learning to evaluate the ARV in inquiries in different school science subjects. It was expected that the use of the same self-evaluation instrument in various school science subjects would support learning to evaluate the ARV in inquiries in these subjects. To study whether the self-evaluation instrument did sufficiently support this learning process, this research focused on:

    (1) the feasibility of an instrument for pre-university science students for self-evaluating the ARV in inquiries in various school science subjects;

    (2) the feasibility of a design of a teaching-learning process in which the instrument was used in a flexible way to self-evaluate the ARV in inquiries in different school science subjects;

    (3) the effectiveness (learning outcomes) of self-evaluating the ARV in inquiries in different school science subjects by using the instrument in the designed teaching-learning process.

Therefore, the main research question of this research was:

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

7.2 RESULTS AND CONCLUSIONS OF STUDIES

Five studies (see Chapters 2 to 6) have been conducted to answer the main research question of this thesis. The figure on the first page of this chapter presents an overview of the design and conclusions that were elaborated in the successive research cycles. Below, the research questions and main results and conclusions of each of the five studies are described.

7.2.1 Explorative study: Knowledge about accuracy, reliability and validity
by pre-university students and science teachers

The explorative study, as described in Chapter 2, focused on the exploration of the knowledge of pre-university science students and science teachers from different science subjects about evaluating the ARV in an inquiry. This study was guided by the research questions:

(1) To what extent do pre-university science students and biology, physics and chemistry teachers recognise concepts of evidence 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 concepts of evidence 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?

The major conclusion of this study was that 23 concepts of evidence (CoE) were relevant and feasible for pre-university science students in learning to perform (self-)evaluations of the ARV in inquiries and in the flexible application of this procedural understanding to inquiries in other school science subjects. These 23 CoE all seem necessary for evaluating the accuracy, reliability or validity in inquiries and can be applied to inquiries in various school science subjects. In more detail, pre-university science students and science teachers did have some basic procedural understanding about the CoE that support the evaluation of the accuracy and reliability in an inquiry. The procedural understanding of students and most teachers about CoE that are useful in the evaluation of the validity of an inquiry was less well developed. An indication was found that biology teachers seemed to have more prior knowledge about the validity of an inquiry than chemistry and physics teachers. The knowledge of accuracy and reliability of an inquiry did not differ between biology, chemistry and physics teachers. So, from this explorative study it became clear which CoE should be elaborated in the design of an instrument that supports students to (self-)evaluate the ARV in inquiries and that can be flexibly applied in various school science subjects.

7.2.2 First research cycle: Feasibility of a set of rubrics as self-evaluation instrument for pre-university students

The first research cycle (see Chapter 3) focused on the feasibility of the first design of a student self-evaluation instrument for evaluating the ARV in inquiries in different school science subjects. As described in the general introduction (see Chapter 1), rubrics seemed to be a feasible format, but little was known about the design characteristics of rubrics which are supportive in the flexible application of evaluating the ARV in inquiries in different school science subjects by pre-university science students. From the explorative study (see Chapter 2 and 7.2.1) and the literature, four design characteristics were identified, respectively regarding:

- content to self-evaluate ARV in inquiries (23 CoE);
- extent of complexity of the descriptions of the rubrics (SOLO taxonomy);
- extent of detail in view of the use by novices; and
- extent of general application in different school science subjects.

This study 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 the accuracy, reliability and validity in successive science inquiry units?*

We found that the four design characteristics led to a set of feasible rubrics for the self-evaluation of ARV in an inquiry by pre-university science students in different school science subjects. The first design should be adjusted on four aspects. First, the set of 19 rubrics should be reduced to make the instrument more easily manageable for novices in evaluating the ARV. Students had difficulties with the application of seven rubrics in which five levels of performances were described. For novices, the extended descriptions of these seven rubrics were cumbersome to handle and it was concluded that the corresponding CoE should be elaborated in a checklist that is easier to handle and understand by novices in evaluating the ARV in inquiries. Secondly, the students needed an overview of the contribution of the CoE items to the process of evaluating the ARV in an inquiry, for instance, by means of a ARV card in which all items are described in general terminology. Thirdly, when using the rubrics it appeared that the descriptions in some rubrics were not arranged hierarchically and had to be revised by taking a closer look at the levels of the SOLO taxonomy. Fourthly, the students could not always relate the benchmark samples to their own inquiries, due to the variety of topics that were used to formulate these samples. To improve the general application of the instrument, the benchmark samples in the rubrics should all be related to the same inquiry topic. These outcomes were used to supplement and sharpen the descriptions of the four design characteristics of a self-evaluation instrument that is feasible for evaluating the ARV in inquiries in different school science subjects.

7.2.3 Second research cycle: Feasibility of the revised instrument as self-evaluation instrument for pre-university students

The second research cycle focused on three aspects: (1) Feasibility of the revised instrument as self-evaluation instrument for pre-university students (see Chapter 4), (2) Teaching-learning process to fulfil functions of revised self-evaluation instrument (see Chapter 5), and (3) Effectiveness of using the revised instrument by pre-university students: learning outcome (see Chapter 6).

First (see Chapter 4) the feasibility of the revised self-evaluation instrument was determined. The revised self-evaluation instrument contained a set of twelve rubrics, a checklist and an ARV card as an overview tool. The rubrics, checklist and ARV card together formed the Evaluation of Quality of Inquiries (EQI) instrument (see Appendix C). This first study on the redesigned instrument was guided by the research 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?*

From the outcomes of this study it was concluded that the EQI instrument was sufficiently feasible for the self-evaluation of ARV in various school science subjects by pre-university science students, especially when the checklist is filled out by the students before they (self-) evaluate the ARV in an inquiry with the ARV card as an overview tool to support the use of the checklist and rubrics. It was concluded that the rubrics of the EQI instrument were applied as intended by the students. Also, the EQI instrument was feasible for evaluating ARV in an inquiry in different inquiry phases and in different school science subjects. A completed EQI instrument gave the teacher sufficient information about the support that the individual students needed during the process of evaluating the ARV in an inquiry. The findings showed the students' correct use of some CoE items as was elaborated in the EQI instrument in students' discussions about the ARV in inquiries during three successive inquiry units, which indicates that the students were able to apply the evaluation of the ARV flexibly in inquiries from one inquiry subject to another.

The successive research studies as described in Chapters 2, 3 and 4 yielded four design characteristics for a feasible self-evaluation instrument with which pre-university students can evaluate the ARV in an inquiry in different school science subjects. These design characteristics are:

(1) Twenty-one CoE from the CoE model (see Table 3.1, p. 55) are necessary and sufficient for teaching pre-university science students, as novices, to evaluate the ARV in inquiries in various school science subjects (content).

(2) For novices it is sufficient when 13 of the 21 CoE on evaluating ARV in an inquiry (see Table 4.1, p. 81) are worked out in more detail in the self-evaluation

instrument, for example, as rubrics. Descriptions in the rubrics by means of the hierarchical levels of performance of the SOLO taxonomy are useful to differentiate between the various levels of performance by pre-university science students in evaluating the ARV in inquiries. The other eight CoE (see Table 4.1, p. 81) could be elaborated in another tool, for example, a checklist, with fewer levels of complexity for novices in evaluating the ARV in inquiries (extent of complexity).

(3) For novices, a self-evaluation instrument for evaluating the ARV in inquiries has to consist of different parts with different amounts of detail by which the ARV in an inquiry can be checked (e.g. the checklist in the EQI instrument) before the ARV is evaluated (e.g. the rubrics in the EQI instrument). A holistic tool (e.g. the ARV card in the EQI instrument) is necessary to support students in getting an overall view of the CoE that are relevant in evaluating the ARV in an inquiry (extent of detail).

(4) To facilitate the flexible application of the evaluation of ARV from one inquiry subject to another, all descriptions in the rubrics should be accompanied by benchmark samples, which are all based on the same inquiry topic (extent of general application).

7.2.4 Second research cycle: Teaching-learning process to fulfil functions of revised self-evaluation instrument

The second study of the second research cycle focused on the design characteristics of a teaching-learning process in which three intended functions of the EQI instrument should be fulfilled (see Chapter 5). The designed teaching-learning process contained three successive inquiry units – general science, biology and physics. Each inquiry unit contained an inquiry task, orientation tasks, peer feedback tasks and transfer tasks (see Table 5.2, p. 113). During the successive inquiry units, the students worked in groups, so that they could discuss the various tasks involved in the units. In all inquiry units, self-evaluation of the ARV took place at similar points and the students were provided with the complete EQI instrument. The teacher explained the use of the instrument to the students before they used the instrument for the first time. At the start of each unit, the terminology used in the instrument was explained by the teacher. The student workbooks contained instructions for the students to use the EQI instrument at the intended points during the teaching-learning process.

As introduced in Chapters 1 and 5, the three intended functions of the EQI instrument in the teaching-learning process were:

(1) self-evaluation of ARV in an inquiry by pre-university science students;

(2) support of this self-evaluation by peers and teachers;

(3) facilitating the transfer of self-evaluation of ARV from one inquiry to another.

The design of the teaching-learning process was based on eleven hypothetical design characteristics (hDCs) (see Table 5.1, p. 105). 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?*

We found that the designed teaching-learning process supported fulfilling the first function about self-evaluation and the second function about support of the self-evaluation by peers and teachers. The third function about facilitating transfer was fulfilled to a lesser extent in the enactment of the designed teaching-learning process. Regarding this third function, it was concluded that the students could to some extent identify analogies between the evaluation of ARV in the three successive inquiry sub-jects and improved their abilities in applying knowledge about the evaluation of the ARV in scientific research to their own inquiries. The effort to explain differences in ARV evaluations across various school science subjects proved insufficient to raise the students' ability to evaluate the ARV in these different inquiries. This could probably be changed by supplementing the teaching-learning process with more (successive) inquiry units and tasks. Based on the outcomes of this study, the eleven hDCs were supplemented and sharpened, which resulted in a list of fifteen design characteristics for a teaching-learning process wherein the three functions of the EQI instrument can be fulfilled.

The design characteristics that were changed or the parts that were added after the research study are written in italics.

Function 1: Self-evaluation
The teaching-learning process should:
- contain various assignments aimed at understanding the meaning of the CoE which are embedded in the EQI instrument;
- contain explanations about the relevance of self-evaluations: *first guided then supported by teacher;*
- be supportive in change of ARV during preparing, *conducting, drawing conclusions* and reflecting on inquiry;
- give students a central role at all points of an evaluation;
- *contain learning tasks about effective strategies for evaluating the ARV in inquiries.*

Function 2: Support by peers and teacher
The teaching-learning process should:
- contain peer feedback tasks for students before teacher feedback is given;
- *provide novices with peer feedback discussions but no more than two-by-two;*

- ■ stimulate the exchange of information between peers preceding discussion on changes. *The teacher should ask scaffolding questions during initial peer discussions to support construction of valid, high-level arguments;*
- ■ provide the teacher with the opportunity to give written feedback *on students' written products* linked to the criteria were the students can strive for (e.g. as described in the EQI instrument);
- ■ *provide the teacher with the opportunity to give oral feedback 'along the way', e.g. during conducting an inquiry;*
- ■ allocate sufficient time for discussion between teacher and students about improving ARV in their inquiries.

Function 3: Facilitation of transfer

The teaching-learning process should:

- ■ contain analogies in learning materials of inquiries in different school science subjects. *These analogies should be derived from previously performed inquiry tasks or inquiry units;*
- ■ contain scientific research examples to explain abstract terminology on ARV. *For novices: the chosen scientific research should be close to students' inquiries;*
- ■ contain explanations about differences in evaluating ARV in different inquiries that are relevant for pre-university students. *This part of the teaching-learning process needs more than three inquiry units or more extended inquiry units;*
- ■ *contain smaller inquiry units to demonstrate one aspect at a time in evaluating ARV in different school science subjects.*

These changed and added parts are yet to be elaborated in the teaching-learning process and need to be tested again in class in further research.

7.2.5 Second research cycle: Effectiveness of using the revised instrument by pre-university students: learning outcome

The study described in Chapter 6 focused on the student learning outcomes (effectiveness) of the use of the EQI instrument in a series of four inquiry units. This study was conducted during the second research cycle, as described in Chapter 5, with the EQI instrument described in Chapter 4. For assessment purposes two individually administered chemistry inquiry assessments were added to the teaching-learning process. This study on the learning outcomes was guided by the research question: *What is the effectiveness of the EQI instrument in supporting the transfer of ARV evaluation skills in various subjects performed by pre-university science students?*

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

- To what extent do students think that the EQI is relevant for evaluating the ARV in an inquiry?
- To what extent do students have sufficient knowledge to evaluate ARV in an inquiry?
- 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?
- 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?

It was concluded that students perceived the EQI instrument as relevant in evaluating the ARV in inquiries, gained sufficient knowledge about CoE and applied, with the support of peers and teacher, these concepts in inquiries in other school science subjects. Transfer of knowledge without guidance of peers and teachers was not achieved to the requisite level in the final assessment inquiries, probably because four inquiry units were insufficient.

7.3 CONCLUSIONS ABOUT FEASIBILITY AND EFFECTIVENESS OF THE EQI INSTRUMENT

The main research question of this research was: *What are the design characteristics of a feasible self-evaluation instrument and a supportive teaching-learning process (wherein the instrument is 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?*

To answer this question, the conclusions about the learning outcomes (see 7.2.5) are related to the above described design characteristics of the EQI instrument (see 7.2.3) and the design characteristics of the teaching-learning process (see 7.2.4). Based on the five studies it can be concluded that the design characteristics led to a self-evaluation instrument that can be used by pre-university science students for the purpose of evaluating the ARV in inquiries in different school science subjects. About the teaching-learning process it can be concluded that the design characteristics led to a supportive process for learning to evaluate the ARV in the separate inquiry units. For both the design of the EQI instrument and the designed teaching-learning process, some findings suggested that the designs fostered the transfer of ARV evaluation skills from one inquiry to another. From the empirical study on the learning outcomes (see Chapter 6), it appeared that students were able to transfer the evaluation of ARV to a new inquiry subject when they were supported by peers and teacher during the EQI instrument based evaluation process. Improving the learning outcomes in evaluating the ARV in an inquiry with 'only' the EQI instrument at hand

175

still seemed to be too difficult for some of the students. These difficulties could be caused, for example, by the students' limited familiarity with evaluating inquiries by using the EQI instrument or by their limited knowledge of the use of the CoE items as embedded in the EQI instrument. Another cause might be that the assessment task was more complex than the other inquiry tasks.

7.4 REFLECTIONS ON THIS STUDY

7.4.1 Reflection on the conclusion

From the empirical studies it cannot be determined with certainty what caused the relative lack of progress with transfer, especially in the final assessment task, when students individually used the EQI instrument to evaluate the ARV in an inquiry. The short length of the intervention could be the cause: three inquiry units could be too few to give all individual students enough procedural knowledge to evaluate the ARV in a new domain. Inquiry in chemistry was new for these students. In just three exercises students may not be able to gain sufficient insights into how to evaluate the quality in another inquiry subject. In discussions with peers and the teacher, students can develop their previously gained insights to a higher level of understanding and explain them in a more general way. Thinking individually, without peer and teacher-guided discussions, does not easily lead to new connections between the new situation and previous situations (Hogan, Nastasi, and Pressley, 1999). Another explanation could be that the students were still learning the meaning of the terminology (CoE items) used in the EQI instrument in inquiries in different school science subjects. The lack of transfer by individual students could also be due to them experiencing a gap between the chemistry inquiry topics in the assessment tasks and the general science, biology and physics inquiry topics the students had worked on (Novick, 1988), or because the inquiries varied widely in levels of difficulty. It could also be that the topics in the assessment tasks were more difficult to understand for some students than the topics in the inquiry units. This difficulty with the transfer of concepts could probably be solved if the teaching-learning process were extended to include more practice tasks. Some suggestions for extending and improving the teaching-learning process were mentioned in the revised design characteristics for the teaching-learning process (see 7.2.6). By extending the intervention, the workload of learning the content and procedure of the EQI instrument and also more activities for transfer of the content and procedure to a new subject could be spread. Such an intervention could also involve specific student tasks on how to attain higher SOLO taxonomy levels of performance, for instance, by improving the designed student inquiries.

7.4.2 Reflection on methodology

In the course of the empirical studies it was decided to focus on the performance of the students and not on the role of the teachers. The explorative study had indicated that teachers had limited knowledge about ARV in inquiries. Although the participa-

ting teachers in the first test cycle in class received some training, these teachers also had difficulties in supporting the students' self-evaluations of ARV in inquiries with the self-evaluation instrument. This first test cycle yielded a lot of information about which difficulties the teachers had in guiding the self-evaluation of students, but these difficulties of the teacher at the same time also influenced the learning outcomes of the students (effectiveness of the instrument). To circumvent the problems caused by the novice status of the teachers in evaluating ARV in inquiries they themselves performed, it was decided for the second test cycle to create a more controlled class-room setting with one of the designers as the teacher and two independent observers in the classroom (see 7.5 Limitations of this research). So, after the explorative study, the role of teachers in evaluating the ARV in inquiries was not (further) investigated, although we realised that this implied a compromise with regard to the naturalistic classroom setting (Barab, 2004) in which design-based research usually takes place. However, this shift to a controlled classroom setting allowed us to gain insight into how the functions of the EQI instrument can be supported by the teaching-learning process and yielded knowledge on the abilities of the students to apply the evaluation of ARV in inquiries flexibly in different school science subjects. Unfortunately, in this research there was no time left for a third test cycle with a focus on the professional development of teachers in evaluating the ARV in inquiries.

As Kelly (2006) stated, the field of design-based research still has to mature. Based on our research, choosing a more controlled classroom setting before investigating a naturalistic setting can be a useful phase in testing educational interventions during design-based research. Brown, as long ago as 1992, expressed concern that too many variables confuse the findings resulting from research in naturalistic settings. To determine how different variables influence a teaching-learning process it is some-times worthwhile to exclude some variables first and then re-add these variables to the educational intervention. This leads to a long, cumulative design process, wherein it should be defined precisely when and how some variables will be actively excluded during the test phase. For example, a teaching-learning process on the evaluation of ARV in inquiries in different school science subjects could be distributed over the whole pre-university science curriculum.

7.5 LIMITATIONS OF THIS RESEARCH

In previous sections some limitations of the studies in this thesis have already been discussed. Two aspects have not yet been mentioned, however. First, the studies in this thesis were performed with interventions on a local level with a small number of students. This problem occurred in a number of other studies in which rubrics were tested (Panadero & Jonsson, 2013). Designing effect studies into the use of rubrics is not useful, however, until it has been tested whether these rubrics actually can be used to evaluate ARV in inquiries by a group of students.

Second, the role of the teacher in fostering the development of ARV evaluation skills in inquiries has not been addressed. The successive studies reported here provided a proof of principle that such a development can be brought about in an appropriate teaching-learning environment with the help of the EQI instrument. This purposeful restriction in the scope of the research implies that the outcomes of the studies on the feasibility of the EQI instrument, the teaching-learning process and the learning outcomes should be qualified. The role and the professional development of the teacher has, as yet, to be further explored. The chosen procedure with a designer of the teaching-learning process as teacher in the second test cycle (see Chapter 4, 5 and 6), created a situation in which the contribution of the teacher to the teaching-learning process was optimal. Therefore, in the studies of the second test cycle the teacher could realise the learning process of the students as intended in the design. To determine whether the teacher was teaching as planned, all lessons were observed and followed-up by other researchers. These researchers also independently analysed the data sources of the second test cycle. The results of this data analysis were discussed until consensus was reached (see Chapters 4, 5 and 6).

7.6 IMPLICATIONS FOR RESEARCH ON SCIENCE EDUCATION
In the first chapter of this thesis, six main concepts were introduced which had to be further explored in successive studies. In this section we will reflect on what has been learned in these successive studies about *learning to inquire with the CoE model (7.6.1), meaning of accuracy, reliability and validity in inquiries (7.6.2), flexible application (transfer) of procedural understanding (7.6.3), formative feedback with self-evaluation instruments (7.6.4), the content of a self-evaluation instrument (7.6.5),* and *the use of a self-evaluation instrument in a teaching-learning process (7.6.6).*

7.6.1 Learning to inquire with CoE model
As stated previously, learning to inquire requires a complex set of students' learning tasks about the construction of evidence and the development of scientific knowledge that is based on evidence from inquiries (see 1.2.1 – Chapter 1). In previous studies it was for example shown that the more aspects are actively experienced by students during an inquiry, the deeper their understanding of the construction of scientific evidence is. Students make more progress in learning to inquire when they understand their actual level of performance on different aspects of an inquiry, know how they could improve their levels of performance and are engaged in science which is related to situations in their daily lives (Minner, Levy & Century, 2010; Harlen, 2012). So, based on the reported studies, a design of a teaching-learning process was made with a series of inquiry assignments in which different aspects of learning to inquire were elaborated. The focus of the learning tasks was on learning to evaluate ARV in an inquiry and the evaluation of ARV was further specified by concepts of evidence from the CoE model (Gott, Duggan, Roberts, & Hussain, n.d.).

As shown in Table 5.6 (Chapter 5), a design of a teaching-learning process for novices on learning to evaluate ARV in inquiries they themselves perform requires a set of tasks about effective strategies in evaluating the ARV in inquiries. An effective strategy could be learning to make use of the different parts of the EQI instrument as developed in the studies of this thesis at appropriate points in a curriculum spanning several years and various school science subjects. The ARV card, for example, helps students to get a 'helicopter' view over what can be evaluated, the checklist can be supportive in checking whether all important steps of an inquiry were taken and, finally, the rubrics can be helpful in evaluating the ARV quality in an inquiry. Such an instrument, which is composed of components that can be used separately, can also be useful to support students with different learning styles in learning to inquire over a longer period of time (e.g., more years in the curriculum). Testing this hypothesis was beyond the scope of this research.

About the CoE model it was elucidated in this thesis how the general descriptions in the model can be described in more detail and more focused on the performance of novices in the process of learning to inquire. These descriptions, as displayed in the EQI instrument, were necessary to teach students to evaluate ARV in an inquiry. Next to this, although the CoE model was originally developed for learning to inquire in secondary school physics education, in this study most of the CoE were also proven to be supportive in learning to inquire in other school science subjects.

7.6.2 Understanding of accuracy, reliability and validity in inquiries

As stated in the general introduction (see Chapter 1), for novices in evaluating the ARV in inquiries it is necessary first to learn about the meaning of ARV in inquiries before they can evaluate the ARV in inquiries in various school science subjects. As Carlsen (2007) stated, science learning is a process moving from abstract language to concrete application. Therefore, it is necessary that students learn to transfer a concrete application to new situations. This implies that a student first has to learn the meaning of ARV, as in learning a new vocabulary, and thereafter can learn how to apply this knowledge on the meaning of ARV in new subjects.

The studies showed that it is not easy to define the ARV in inquiries in a way that can be understood and applied appropriately in different school science subjects by novices. By designing, performing and evaluating as many different inquiries as possible, the understanding and application of the evaluation of ARV by pre-university science students could be improved. It should be further elaborated how many and what type of inquiries have to be performed by students to become more capable in evaluating ARV in inquiries.

From the first research cycle (see Chapter 3) it was found that students need learning materials with consistently worked-out examples, not only for the explanation of the meaning of ARV in inquiries in general but also for the application of specific

CoE in certain inquiries. From the experiment in this first test cycle it can also be found that the students, as novices, had to apply too many CoE at a too complex level when evaluating the ARV in inquiries. The application of CoE by evaluating the ARV improved when some CoE were worked out at a high level of complexity in the EQI instrument and other CoE at a lower level of complexity (see Chapter 4 and 6). So, like learning a new language, it seems to be important to introduce a new vocabulary for evaluating ARV in inquiries step-by-step: students first have to learn CoE, at the appropriate level of complexity, which are relevant for the first inquiry, followed by expanding the learned CoE and levels of complexity for the next inquiry etcetera. It should be elucidated in more detail how this expansion can be worked out in the EQI instrument and the inquiry tasks.

Next to this, in both test cycles it was shown that the students had difficulties in making a distinction between evaluating ARV in inquiries with non-living test objects and inquiries with living test objects. The differences in evaluating the accuracy and reliability in both situations, especially, were quite confusing to them because of their lack of knowledge on (evolutionary) variation between living test objects. It should be further explored how the differences between evaluating ARV in both situations can be better explained to novice learners.

7.6.3 Flexible transfer of evaluating accuracy, reliability and validity to new subjects

As mentioned in Chapter 1, Salomon and Perkins (1989) distinguished two ways of transfer: a focus on extensive practice to foster the use of acquired skills in a similar setting *(low-road transfer)* and mindful abstraction of strategic understanding as learned in one setting in order to apply it in a new setting *(high-road transfer)*. For the purpose of the studies in this thesis it was decided to focus on one aspect of high-road transfer: the flexible application of procedural understanding (Van Oers, 1998) of ARV evaluation of inquiries to new school science subjects. However, testing the learning materials in the classroom showed that it is almost impossible in the restricted time schedule to arrange flexible application of the ARV in inquiries without exercising the use of ARV in a new inquiry. For novices, learning to evaluate ARV in inquiries is such a complex task that an optimal mix of tasks on low-road and high-road transfer should be incorporated in the teaching-learning process.

As shown in the study about the learning outcomes (see Chapter 6), it was difficult for some pre-university science students to apply the evaluation of ARV in inquiries flexibly from a general science inquiry to a biology inquiry unit and, subsequently, to a physics inquiry, and to the individual assessments in a chemistry inquiry. It could be that some descriptions were, even after the first test cycle, not sufficiently attuned to a biological or chemistry subject. A main difference between the general science and physics inquiry unit, on the one hand, and the biological inquiry and biochemical

assessment tasks, on the other, was the use of living test objects in the biological and biochemical tasks. As Pantin (1968) mentioned, in inquiries with living test objects it is, for example, difficult to control all irrelevant, influencing variables. Although the differences between evaluating the ARV in inquiries with non-living and with living test objects was explained in the biological inquiry unit, for novices it seemed to be too complex to recognise the differences in evaluation of ARV in inquiries in both school science subjects, especially without guidance from the teacher. How much practice novices need evaluating ARV in 'non-living test object inquiries' before they can flexibly apply this knowledge to inquiries with living test objects should be explored. It should also be further explored how the CoE model could be expanded to make it fully applicable to inquiries with living test objects.

A final important issue related to facilitating transfer is the recognition of analogies between different inquiries by pre-university science students. The use of the same EQI instrument in various inquiry units was expected to be an important common framework for the students (see 5.2.3), because coherent use of language and the same patterns in students' mental actions are deemed beneficial for transfer (Gilbert, Bulte, & Pilot, 2011). Although the use of the EQI instrument supported the students in recognising the similarities and differences in evaluating the ARV in those different inquiries, it seems to be necessary to have other identifying marks in the learning ma-terials. These identifying marks could be the use of analogous language in different inquiries, as was done in the learning materials of the studies in this thesis. Which identifying marks in the learning materials are necessary to support the flexible application of the EQI instrument should be further investigated.

7.6.4 Formative feedback with self-evaluation instruments

In both test cycles of this study, rubrics, a checklist and an ARV card were designed and tested as part of a self-evaluation instrument, the EQI instrument. As Panadero and Jonnson (2013) showed in their review study, rubrics might be valuable in supporting student learning and in facilitating improvements in students' perfor-mances, when their use is combined with meta-cognitive activities as self-regulation, self-evaluation or peer feedback. The findings of the studies in this thesis support these conclusions, but before students' learning can be supported fully, it is neces-sary that students learn how to apply the rubrics for evaluating the ARV in an inquiry and how to interpret the descriptions in the rubrics. As Van Dinther, Dochy and Segers (2011) stated, students first need to be confronted with teacher's feedback (i.e., regarding the ARV in their inquiry) to have a realistic opinion about the level of their performances before they are able to use the rubrics for self-evaluation appropriately. This implies a small redesign of the teaching-learning process in which, during the first inquiry units, more time is reserved for receiving and discussing the teacher's feedback on the students' inquiry plans and on the completed inquiries.

181

The EQI instrument seems to have the ability to support students in different ways: the ARV card as an overview instrument, the checklist to check whether all steps have been completed and the rubrics to evaluate the quality of an inquiry. In the studies of this thesis, these three parts of the EQI instrument were all used by the students in a pre-ordered sequence. It may be worthwhile to investigate whether and how these three parts can be supportive in creating a teaching-learning process in which students can choose individually whether the ARV card, checklist or rubric is most useful for them, and at what specific point, during the evaluation of their inquiry.

Reitmeier and Vrchota (2009) compared the use of a rubric with the use of a reflection sheet. No significant differences in learning outcomes of the students were found. Panadero (2011) found that the use of rubrics reduced the anxiety of students for self-regulation. Apart from the use of rubrics to evaluate the quality of an inquiry, therefore, rubrics can provide confidence to the students about their capacity to eva-luate their inquiries adequately. The choice of a domain (of an inquiry) does not seem to influence greatly the effects of using rubrics. The use of a self-assessment script (Panadero, ibid.) enhanced learning how to self-regulate the learning process. The checklist of the EQI instrument can be extended as a self-assessment script, because in the checklist items for evaluating ARV are already elaborated, with an actual focus on checking whether all necessary steps have been completed. By improving the checklist of the EQI instrument, it could be supportive to improving self-regulation skills of the students. Based on the research and review studies reported above, it can be expected that all three components of the EQI instrument can support students in the same process with different functions, once students understand which aspects of their ARV evaluation skills need to be improved.

7.6.5 SOLO taxonomy for levels of self-evaluation instrument

As described in the general introduction (see Chapter 1), a hierarchical taxonomy was needed to describe the successive levels of performance in evaluating the ARV in the rubrics of the self-evaluation instrument. In the studies of this thesis (see Chapter 3 and 4), the SOLO taxonomy has been demonstrated as an appropriate taxonomy for describing the levels of performance in a principled and hierarchical way.

These studies showed that the SOLO taxonomy is useful when describing CoE that can be used by novices at a high level of complexity. Such CoE were described at the five levels of complexity of the SOLO taxonomy, and, more importantly, those five levels of complexity fit the performances of the students. During the first test cycle (see Chapter 3) it was learned that rubrics concerning CoE which were performed by the students at low levels of complexity were not feasible for the students. Students should at least be able to act at the unistructural level of the SOLO taxonomy before it is instructive for students to use rubrics in which higher levels are described. In other

words: students need to display a first hint of scientific thinking about an aspect of an inquiry (for example, the inquiry question) before a rubric based on the levels of the SOLO taxonomy can be appropriately used.

Related to this is the role of the SOLO taxonomy in the process of flexibly applying the rubrics to new subjects. With the support of the teacher and after some practice, the students were able to apply the rubrics appropriately in different inquiry subjects (see Chapter 6), which shows that the SOLO taxonomy is a useful means to foster transfer of ARV evaluation skills. As Panadero and Jonsson (2013) showed in their review study, novices need sufficient time to learn to work with the rubric, and the effect of rubrics appeared to increase in the course of the implementation. Therefore, it should be further elaborated how students can be guided to apply the rubrics in inquiries in different school science subjects with less or no support by their teacher and how many inquiries should be done before students can use the rubrics of the EQI instrument without help.

7.6.6 Use of a self-evaluation instrument in a teaching-learning process

As described in the general introduction, it is important to have consistency between the intended functions of a self-evaluation instrument and the teaching-learning process in which the instrument is to be used (Ledford & Sleeman, 2000). Evaluation instruments can be used for self-evaluation, peer feedback and feedback by the teacher (Andrade & Du, 2007; Nicol & Boyle, 2003). In the studies in which the EQI instrument was tested, the instrument was intended to be used for self-evaluation, peer feedback as well as feedback by the teacher. It was found that the function of self-evaluation could only be appropriately performed if students understood the meaning of descriptions in the EQI instrument. For novices it seems to be important to let them practice first with the evaluation of other inquiries and scientific research with the EQI instrument and also to compare the outcomes of their self-evaluation with the evaluation of their inquiry by their teacher.

We decided to include peer feedback in the teaching-learning process because it has been found that students are often better able than teachers to provide a comprehensible explanation (Nicol & Macfarlane, 2006). We expected that peer feedback on the ARV in inquiries should function in this way. From the studies in this thesis we learned that the function of peer feedback on the ARV evaluation by novices only functioned appropriately when small groups evaluated each other's inquiries (see Chapter 5). Evaluating the inquiries of many peer groups seemed to lead to superficial evaluations of the performances of the other groups. The previously discussed lack of flexible application of the evaluation of ARV by the students (see 7.6.3) already showed that students had difficulty applying the EQI instrument descriptions appropriately to another inquiry subjects without guidance from their teacher. The same

lack of flexible application could arise while evaluating inquiries of different peer groups. It can be expected that two-by-two peer discussions also give more room for the teacher to guide novices in understanding the meaning of descriptions in the EQI instrument than guiding the process of feedback by constantly changing peer groups giving feedback.

The last function of the EQI instrument in the teaching-learning process was to support the teacher in giving oral and written feedback on the performances of the students. It is often difficult for teachers to give students insight into their evaluation criteria, while students need this insight before they can adequately self-evaluate their per-formances (Bransford, 2000). It was expected that by using the same instrument for self-evaluation and teacher feedback, teachers and students use the same language about ARV in inquiries (Nicol & Macfarlane, 2006). Speaking the same language as the teacher could help students to get a deeper and more detailed understanding of the descriptions of the instrument (Hogan et al., 1999). However, the first test cycle showed that it was not easy for the teachers to use the self-evaluation instrument appropriately because they were as much novices in the evaluation of ARV in students' inquiries as the students. As a consequence, although Dutch science teachers at pre-university science level are expected to have experience with evaluating and fostering ARV in scientific research from studying for their own science degrees at university, more professional development is needed on how to apply this experience to the evaluation of ARV in students' inquiries.

7.7 IMPLICATIONS FOR PRE-UNIVERSITY SCIENCE EDUCATION
7.7.1 Learning to inquire in school science subjects
Recently, many schools in the Netherlands have started constructing a trajectory on developing inquiry skills in various school science subjects. Such a trajectory can be defined as 'both an interventional research activity and a product […], which includes well-researched teaching-learning activities' (Méheut & Psillos, 2004, p. 516). In the eyes of science teachers, a course on learning to inquire should match all school science subjects as well as all successive school years. In addition, increasing both practical skills and procedural understanding of students should be part of the teaching-learning process, while teachers often do not realise this distinction in 'inquiry skills' (Van der Valk & Van Soest, 2004). Although many schools started with the construction of a trajectory on learning to inquire, little is known about the design characteristics of such a separate programme within the curriculum of the pre-university school science subjects.

This research elucidated the 'procedural understanding' part of a teaching-learning process on learning to inquire, especially one that matches the pre-university school science subjects of biology, physics and chemistry at the pre-exam level. Using the

same instrument in different school science subjects, for example, the EQI instrument, can foster the transfer of ARV evaluation skills in inquiries. Students can benefit from using the same instrument in various inquiry subjects to learn to understand the language used in the instrument, which can accelerate the development of their inquiry skills. In the studies of this thesis, pre-university science students started to use the EQI instrument in their pre-exam school year. They found it difficult to learn, at the same time: (i) a new language, (ii) the procedure of evaluating ARV in inquiries, and (iii) to apply the evaluation of ARV in inquiries flexibly in different school science subjects. In a more extended teaching-learning process the EQI instrument, or parts of it, could be introduced in earlier school years to spread the workload of the students of learning to evaluate the ARV in inquiries. Teachers should keep in mind that the language in the EQI instrument probably has to be simplified for an adequate understanding in earlier school years.

A prerequisite of facilitating transfer – using an instrument from one inquiry unit in another – is that all science teachers use the same instrument and are alert to explaining differences between evaluations of ARV in inquiries in different school science subjects to their students. This implies that the teachers know exactly in which inquiries other science teachers use or have used the instrument, and how their colleagues used the instrument. They also should be sure to give the same meaning to the described items in the instrument and have to be aware of their students' everyday knowledge on ARV. A complex didactical structure is required to teach the 'content' in a teaching-learning process in different science subjects (Lijnse & Klaassen, 2004). A community of practice of teachers and educational researchers, dedicated to learning to inquire, can be helpful in developing such a didactical structure as part of a teaching-learning process on learning to inquire (Van Rens, Pilot, & Van der Schee, 2010; Van Rens, et al., 2014) and to analyse the purposes of the inquiries in which the instrument will be used (Millar, 2010). The designed EQI instrument of this study may be a source of inspiration for teachers.

7.7.2 Teacher training on evaluation of ARV in inquiries

The first and second empirical study in this thesis showed that the procedural understanding of science teachers had to be deepened before they could teach the evaluation of ARV in inquiries to their students. Most teachers are not used to supervising inquiries in which students focus on the evaluation of ARV. Sometimes they know the concepts ARV and the CoE from their own academic science studies, but in teacher training courses little attention is paid to this aspect of learning to inquire.

Due to the new standard on the evaluation of ARV in inquiries in the new formal science curriculum of biology, chemistry and physics in the Netherlands, it would be worthwhile to develop an in-service course for teachers to learn more about this

specific focus on learning to inquire. The desire to have a teaching-learning process on learning to inquire in different school science subjects also provokes teachers to learn about the evaluation of ARV in inquiries. This implies two aspects that should be included in in-service training: apart from learning a 'new language' to speak with the students, teachers should learn how they could provide formative feedback on the ARV in inquiries of their students, for example, with the help of the EQI instrument (Van der Schee & Rijborz, 2003). This last aspect requires a change in perspective on the role of teachers, from an instruction-oriented role to a guidance-oriented role (Van der Valk & De Jong, 2009).

As De Jong, Korthagen and Wubbels (1998) stated in their review study on science teacher education, when new teacher training courses are developed it is important to have a strong connection between, on the one hand, the theory about evaluation of ARV in inquiries by students, and, on the other hand, the teaching practice. A community of practice of teachers and educational researchers can be helpful in bridging the gap between theory and practice. Such communities also can increase the willingness of teachers to make use of the insights from this research. It seems crucial in such courses to let teachers experiment with ideas about evaluation of ARV in inquiries in their own classes (Van Rens et al., 2010). Learning to give formative feedback on the ARV in inquiries can be done by training teachers to ask adequate scaffolding questions during the process of inquiries in class (Van de Pol, Volman, & Beishuizen, 2012; Van der Valk & De Jong, 2009).

7.8 RECOMMENDATIONS FOR FUTURE RESEARCH

This section contains some suggestions for future research on the EQI instrument, evaluation of ARV in inquiries and learning to inquire in general (see also 7.6).

The students who participated in the second research cycle of this study (see Chapters 4–6) have now already passed their final exams at pre-university level. As part of these exams, they had to complete a Final Assignment with a self-designed inquiry (in Dutch: *profielwerkstuk*). It would be interesting to investigate whether in their Final Assignment these students realised a higher level of ARV in their inquiries and made more advanced use of the CoE to improve the ARV in their inquiries than students who did not participate in this research. The teachers who supervised the Final Assignment could also be interviewed to determine whether they think that the participating students performing better in evaluating the ARV in their inquires than non-participating students.

As mentioned before, the EQI instrument could be introduced in earlier school years to spread the workload of students in learning to evaluate the ARV in inquiries. For this purpose it should be determined how the CoE items have to be described in the

EQI instrument to make them comprehensible for younger pre-university students. Perhaps an extra tool for younger students has to be added to the EQI instrument. In addition, supportive learning materials with which these students learn to evaluate aspects of the ARV in inquiries with the adapted EQI instrument should be designed and tested.

More research could be done on the design of an integral trajectory dedicated to learning to inquire at pre-university level. Previous studies, including those described in this thesis, all focus on parts of such a process. It still remains unclear how students can learn to inquire in a continuous programme from primary school to university. To gain a better understanding of a feasible and effective design of such a trajectory, it might be interesting to start a longitudinal research project in which outcomes of different studies on learning to inquire are combined.

In the Netherlands, as well as learning to inquire at pre-university level (VWO), science teachers are interested in a teaching-learning process for learning to inquire at senior general education level (HAVO). In the formal curriculum of the secondary general education level it is also written that HAVO-students have to learn to evaluate ARV in inquiries. As far as we know, it has not been investigated which aspects are necessary to learn about the evaluation of ARV at secondary general education level to prepare these HAVO-students on their future studies at universities of applied sciences.

Finally, pre-university students learning to inquire need support from their science teachers. Next to research on effective training courses for qualified teachers (see 7.7.2), it should be investigated how courses about learning to inquire and evaluating ARV in inquiries can be included in the pre-service education of science teachers. In particular, novice teachers could benefit from practising the process of providing formative feedback during inquiries, for example, as part of a course about formative feedback as assessment instrument.

REFERENCES

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. Practical Assessment, Research & Evaluation, 10(3).

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. Journal of the Learning Sciences, 13(1), 1–14.

Biggs, J., & Tang, C. (Eds.). (2007). Teaching for quality learning at university (3rd ed.). Buckingham, UK: Open University Press.

Bransford, J. D. (2000). How people learn: Brain, mind, experience, and school (Expanded ed.). Washington, DC: National Academy Press.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of the Learning Sciences, 2(2), 141–178.

Carlsen, W.S. (2007). Language and science learning. In S.K. Abell & N.G. Lederman (Eds), Handbook of research on science education (pp. 57–74). Mahwah, NJ: Lawrence Erlbaum Associates.

De Jong, O., Korthagen, F., & Wubbels, T. (1998). Research on science teacher education in Europe: Teacher thinking and conceptual change. In B. Frazer & K. G. Tobin (Eds.), International handbook of science education (pp. 745–758). Dordrecht, The Netherlands, and Boston, MA: Kluwer.

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., Roberts, R., & Hussain, A. (n.d.). Research into understanding scientific evidence. Retrieved 5 November 2014 from http://www.dur.ac.uk/rosalyn.roberts/Evidence/cofev.htm

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.

Hogan, K., Nastasi, B.K., & Pressley, M. (1999). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. Cognition and Instruction, 17(4), 379–432.

Kelly, A. E. (2006). Quality criteria for design research: Evidence and commitments. In J. Van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), Educational design research. London: Routledge.

Ledford, B. R., & Sleeman, P. J. (2000). Instructional design: A primer. Greenwich, CT: Information Age Publishing.

Lijnse, P., & Klaassen, K. (2004). Didactical structures as an outcome of research on teaching-learning sequences? International Journal of Science Education, 26(5), 537–554.

Méheut, M., & Psillos, D. (2004). Teaching-learning sequences: aims and tools for science education research. International Journal of Science Education, 26(5), 515–535.

Millar, R. (2010). Analysing practical science activities to assess and improve their effectiveness. Hatfield, UK: The Association for Science Education.

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.

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.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. Journal of Experimental Psychology, 14(3), 510–520.

Panadero, E. (2011). Instructional help for self-assessment and self-regulation: Evaluation of the efficacy of self-assessment scripts vs. rubrics. Doctoral dissertation. Universidad Autónoma de Madrid, Madrid, Spain.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. Educational Research Review, 9, 129–144.

Pantin, C. F. A. (1968). The relations between the sciences. London and New York: Cambridge University Press.

Reitmeier, C.A., & Vrchota, D.A. (2009). Self-assessment of oral communication presentations in food science and nutrition. Journal of Food Science Education, 8(4), 88–92.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. Educational Psychologist, 24(2), 113–142.

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 der Schee, J., & Rijborz, J. D. (2003). Coaching students in research skills: A difficult task for teachers. European Journal of Teacher Education, 26(2), 229–237.

Van der Valk, A. E., & Van Soest, M. F. N. (2004). Onderzoek leren doen in de bètavakken, elementen van een leerlijn in de onderbouw van twee scholen. [Learning to inquire in the school science subjects, elements of a teaching-learning sequence in lower secondary education at two schools]. Utrecht: Universiteit Utrecht.

Van der Valk, T., & De Jong, O. (2009). Scaffolding science teachers in open-inquiry teaching. International Journal of Science Education, 31(6), 829–850.

Van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. Educational Research Review, 6(2), 95–108.

Van Oers, B. (1998). From context to contextualizing. Learning and Instruction, 8(6), 473–488.

Van Rens, L., Pilot, A., & Van der Schee, J. (2010). A framework for teaching scientific inquiry in upper secondary school chemistry. Journal of Research in Science Teaching, 47(7), 788–806.

Van Rens, L., Hermarij, P., Pilot, A., Beishuizen, J., Hofman, H., & Wal, M. (2014) Pre-university chemistry students in a mimicked scholarly peer review. International Journal of Science Education, 37(4), 2514–2533.