



Experimenting Matters. Learning and Assessing Science Skills in Primary Education

P.M. Kruit

Summary

Operationalization of science skills

In chapter 2 we addressed the first research question by discussing the definition and operationalization of science skills. Despite their prominence, science skills remain a rather “ill-defined domain” (Gobert & Koedinger, 2011). In most curriculum frameworks science skills are defined by the activities which are intended to reflect the work of actual scientists (Lederman & Lederman, 2014). The rationale is that knowing and experiencing how scientists work will enable students to develop an idea of the methodological toolbox of a scientist. The underlying assumption in many teaching programs is that by carrying out a scientific investigation, students will learn science skills naturally and at the same time develop an understanding of how scientists work. Within that context, Osborne (2014) argues that engaging in scientific practice can improve the quality of students’ learning. However, the way scientists work is different from the way students learn. Students in primary education are still novices with regard to scientific inquiry. Science skills need to be acquired in a systematic manner (Metz, 2011). When designing effective instructional methods with the purpose of stimulating the development of science skills, it is essential to acknowledge the cognitive demands underlying these science skills. We argued that three different skills underlie the general concept of science skills. These are defined as science-specific skills, thinking skills and metacognitive skills.

In the context of performing a scientific inquiry, science-specific skills refer to the ability to apply procedural and declarative knowledge which is needed for properly setting up and conducting an inquiry (Gott & Murphy, 1987). Examples of these skills include taking measurements, organizing data into tables, making graphs, or using measurement devices. Science-specific skills are classified as lower order thinking (Newmann, 1990) or reproductive thinking (Maier cited in Lewis & Smith, 1993), and are characterized by knowledge recall, comprehension, the routine employment of rules, and simple application (Goodson, 2000). Science-specific skills defined as such include the practical skills as discussed by Abrahams and Reiss (2015), but they pertain to cognitive processes as well.

Thinking skills include the higher order skills, also frequently referred to as critical thinking (Moseley et al., 2005). Thinking skills involve manipulating information that is in nature complex because it consists of more than one element and has a high level of abstraction (Bloom, 1956; Flavell, Miller, & Miller, 1993). In a scientific inquiry, thinking skills are applied to make sense of the data and to connect the observations to scientific theories (Osborne, 2015) such as formulating hypotheses, making inferences from different sources of data, identifying features and patterns in data, or drawing a conclusion (Millar & Driver, 1987; Pintrich, 2002; Zohar & Dori, 2003).

Metacognitive skills include planning, monitoring and evaluating task performance (Flavell et al., 1993). These skills influence the quality of the scientific inquiry process which in particular demands

self-regulation and use of metacognitive strategies (Schraw, Crippen, & Hartley, 2006). What distinguishes metacognitive skills from thinking skills is that they involve active executive control of the mental processes (Goodson, 2000) or the “thinking about thinking” (Kuhn, 1999; Kuhn & Dean, 2004, p. 270).

Finally, we discussed the influence of content knowledge on skill development and performance. Content knowledge is most often referred to as conceptual understanding of facts, concepts, theories and principles (OECD, 2017). Previous studies have shown that content knowledge is, to a certain extent, a prerequisite for skill development (e.g., Eberbach & Crowley, 2009; Kuhn, Schauble, & Garcia-Mila, 1992). Different levels of content knowledge can result in significant differences in skill performance (French & Buchner, 1999). Even when the level of cognitive abilities is supposed to be a limiting factor for a given individual such as a young student, it is still possible to become an expert in a specific subject area (for example “dinosaurs”) and subsequently perform better in problem solving tasks compared to adults who know less about the subject (Glaser, 1984).

Components of an instructional framework

To answer the second research question of identifying the crucial components of an instructional design for teaching science skills, we developed an instructional framework for teaching science skills based on the operationalization of science skills into three different underlying skills (or subskills) (chapter 2). Grounded on the design principles of this framework, a series of eight lessons was constructed, piloted and used for the intervention of the quasi-experimental pretest-posttest study (chapter 4).

Skills for scientific inquiry are usually taught by instructional methods primarily based on learning by doing (Duschl, 2008; Roth, 2014) despite evidence suggesting that more explicit teaching methods and strategies may be more effective (Klahr & Nigam, 2004; Lazonder & Harmsen, 2016). Most teachers focus on only the practical aspects of scientific inquiry which results in the disregard of the wide variety of cognitive abilities called upon in scientific investigations. Due to limited experience and instruction, students in primary education often lack sufficient mastery of strategies and knowledge to effectively use and apply the skills to a scientific inquiry in an integrated way. In addition, the working memory capacity of students who are novices at performing a scientific inquiry may limit their ability to conduct a complex task. For this reason, we argued that an instructional framework based upon a cognitive approach aimed at acquiring science skills by means of an explicit instruction method may support students’ learning process more adequately than an approach based on learning by doing, and offers educators and teachers guidance in designing and conducting lessons.

Explicit instruction aims to make students explicitly aware of the skills they should learn and apply. Making students explicitly aware of the strategies and skills that they are applying to a particular task leads to enhanced mastery which in turn may facilitate transfer, which is considered an indication of more robust learning. Near-transfer can generally be defined as the application of skills

to tasks within a particular knowledge domain or with a common structure. In this study, explicit instruction included explanations from the teacher and classroom discussions on how and when to apply the skills. In each lesson, attention was paid to one particular step within the empirical cycle. Explicit guidance was included in the form of probing questions and prompts provided during task performance.

Explicit instruction concerning metacognitive skills was addressed by introducing the TASC model throughout the lessons. TASC stands for “Thinking Actively in a Social Context” and aims at giving students structure to support their thinking (Wallace, Bernardelli, Molyneux, & Farrell, 2012). The TASC model consists of a series of questions which can be used to make students aware of the need to monitor and evaluate task execution which is intended to improve metacognitive skill application. In each lesson, students were instructed to think about and discuss the TASC questions with the teacher and with each other.

An instructional framework for explicit instruction of skills was developed by applying two structuring principles. The first involved using the principles of the four-component instructional design (4C/ID) model as a starting point (van Merriënboer, Jelsma, & Paas, 1992; van Merriënboer, Clark, & de Croock, 2002). This involves the implementation of whole learning tasks, together with part-task practice, and includes scaffolding and feedback opportunities. The part-tasks were aimed at strengthening the subskills which were then simultaneously applied to a whole task scientific inquiry. This stimulated the integration of the skills that were practiced separately. A series of lessons structured according to the 4C/ID model involves a careful sequence of part-tasks and whole tasks which gradually increase in difficulty and complexity. The second structuring principle of the instructional framework concerned incorporating the steps of the empirical cycle. For primary science education, a generally accepted guiding principle is structuring investigations by following the main steps of the empirical cycle which include: (1) formulate a research question, (2) formulate a hypothesis, (3) design an experiment, (4) measure and record data (5) analyze results, and (6) formulate a conclusion. The empirical cycle reflects all aspects of a scientific inquiry that are included in most curricula as learning objectives. Most science tasks in primary education are more or less structured accordingly. In the instructional framework, the principle of structuring via the steps of the empirical cycle was applied in two different ways. First, in each lesson, students performed a scientific inquiry (whole task) structured identically into the six steps that represent the empirical cycle. Second, in each lesson, the main focus was on one of the steps only.

In this study, the feasibility and usefulness of the instructional framework including the above described crucial components was demonstrated on the basis of a fully developed lesson. This detailed example lesson illustrated an approach for systematically designing science lessons for primary education according to the principles of the 4C/ID model combined with aspects of explicit instruction.

The implementation of the lessons in a variety of classrooms showed that teachers found the lessons to be feasible to accomplish within their daily practice and the school curriculum for science education. Students found the lessons interesting and stimulating. It was therefore concluded that the instructional framework and the operationalization of science skills into science specific, thinking and metacognitive was not only feasible but provided ample opportunities to construct science lessons which could be aligned with the regular science curriculum in schools.

Measuring students' ability to perform a scientific inquiry

In the third chapter, we explored the construction, the validity and the reliability of different instruments for measuring science skills. The instruments included a paper-and-pencil test, three performance assessments, and two metacognitive self-report tests. Previous research showed that it is generally difficult to attain convergence between tests with different test formats. The problems have been mainly attributed to differences in students' level of content knowledge, inconsistencies in rating and occasion sampling variability (students perform the same task differently on different occasions). To examine whether convergence may be improved as well as to assure that all aspects of scientific inquiry were included, the paper-and-pencil test and the performance assessments were systematically constructed based on the three subskills (science-specific, thinking, metacognition) as well as on the different steps of the empirical cycle.

The paper-and-pencil test consisted of a total of 46 items that were subdivided into 10 open-ended and 36 multiple choice questions. For administration purposes, the paper-and-pencil test was split into two optimal split-halves. In the performance assessments, students were asked to conduct a small investigation and formulate their findings and answers on a worksheet. All three performance assessments (Skateboard, Bungee Jump and Hot Chocolate) differed in topic but were each constructed according to the same template which consisted of a total of 14 items. One of the metacognitive questionnaires was based on the Junior Metacognitive Awareness Inventory (Jr. MAI) (Sperling, Howard, Miller, & Murphy, 2002) and consisted of 12 items with a three-choice response. The second metacognitive self-report test - Science Meta Test (SMT) - was designed to measure self-regulatory skills and was specifically aimed at obtaining information regarding the application of metacognitive skills in the performance assessments. Because previous research indicated that general cognitive ability is often related to students' ability in performing scientific inquiry (Pine et al., 2006; Roberts & Gott, 2006), the results of a standardized test were collected as well. This test is conducted every year in Dutch primary schools within the context of a student monitoring system and provides an indication of students' general cognitive ability.

The results of the tests which were administered to 128 grades 5 and 6 students showed that the tests were sufficiently reliable. Results also indicated that students' ability to perform scientific inquiry was significantly related to general cognitive ability. Positive correlations between the paper-and-pencil test and the three performance assessments showed that the different test formats measured

a common core of similar skills, thus providing evidence for convergent validity. By contrast, we found no relationship between the measure of general metacognitive ability and either the paper-and-pencil test or the three performance assessments. However, the metacognitive self-report test constructed to obtain information about the application of metacognitive abilities in performing scientific inquiry, showed significant - albeit small - correlations with two of the performance assessments.

Additionally, we explored to what extent scores on both subskill and empirical step level can be used to obtain valid and reliable diagnostic information in addition to overall test scores. Each item in both the paper-and-pencil test and the performance assessments was classified by determining the primary skill underlying that particular item. These items were also assigned to one of the steps of the empirical cycle. Then, on subskill and step level, the mean scores, reliabilities and correlations between scores were obtained and discussed. The results showed that scale reliabilities were acceptable on subskill level for the paper-and-pencil test as well as aggregated across performance assessments. In addition, the correlations between the mean scores of each subskill scale indicated that a more precise identification of students' ability in performing scientific inquiry can be realized.

On empirical step level the results showed weak to moderate reliabilities on average and erratic correlations between steps of both the paper-and-pencil test or aggregated across performance assessments. Because of this, we concluded that ability scores on empirical step level should be interpreted with caution, especially when used for summative assessment purposes.

From this study we concluded that with the paper-and-pencil test and the three performance assessments, measurement of science skills can be attained in a reliable and valid manner by systematically constructing items directed to subskills and different empirical steps.

Furthermore, although reliability of skills measured per subskills and step level is limited, we demonstrated that additional diagnostic information for formative purposes can be obtained by examining mean scores on both subskill and step level. By contrast, we concluded that self-report questionnaires were less suitable for measuring metacognitive skills in this group of primary school students and results obtained from these self-report tests need to be interpreted with caution.

Effects of explicit instruction

The fourth research question that we tried to answer in this thesis concerned the effects of explicit skill instruction on students' acquisition of skills in scientific inquiry. In chapter 4, we described the results of a quasi-experimental study with pretest-posttest design on the effects of explicit versus implicit instruction on the acquisition and transfer of science skills. This study was conducted at 12 schools for primary education in the Netherlands. The participants included a total of 705 students (aged 10-12 years) from 31 grades 5 and 6 classes. The study was designed to investigate the effects of an 8-week intervention with explicit instruction on students' acquisition of inquiry skills (explicit condition). The eight lessons of 90 minutes each of the explicit instruction condition were developed

according to the instructional framework described in chapter 2. A control condition was included with lessons in which skills were taught using an inquiry-based approach without explicit instruction on inquiry skills so that information about the added value of explicit instruction could be obtained. All lessons were piloted in a total of three grade 5 and 6 classes in two different schools before using the lessons for the intervention. To contrast both controlled conditions with regular science lessons at schools, a baseline condition was added in which students followed the regular curriculum. Within schools, classes were randomly assigned to conditions (see for overview Figure 4.2).

To obtain information on the acquisition of science skills, all students were tested with the measurement instruments described in chapter 3. In the pre-test session, the measures included the optimal split-half of the paper-and-pencil test, a performance assessment (Skateboard) and the Jr. MAI self-report test. After the intervention period (8 - 10 weeks), a subsample of the students ($n = 467$) were tested again with the other split-half of the paper-and-pencil test, two performance assessments (Bungee Jump and Hot Chocolate) and both metacognitive self-report tests (Jr. MAI and SMT). Specifically, the topic of the performance assessment Hot Chocolate corresponded with the topic *heat and temperature* addressed in both intervention conditions, while the topic of the performance assessment Bungee Jump was unfamiliar to all students. The last test served to assess the near-transfer of skills to a similar task with a new and unfamiliar topic.

In addition, we included measures to obtain information about the integrity of the implementation of the intervention lessons and whether or not students enjoyed the lessons. The results indicated that there were no significant differences between the two intervention conditions on either of these measures. In both conditions the lessons were taught as intended and in general, students enjoyed the lessons.

Multi-level models with two levels were used to analyze the data of a total of 403 students. The scores on the pre-tests (paper-and-pencil test and the performance assessment Skateboard), general cognitive ability, age, gender and grade level were included as control variables. The scores of the Jr. MAI were not included because of insufficient reliability and lack of variance in post-testing. Results of the analysis of the paper-and-pencil test and the performance assessments showed significant effects for the pre-tests, general cognitive ability and gender. Conversely, grade level did not affect the scores on the post-tests. There was no significant effect of condition on the paper-and-pencil test. However, students of both intervention conditions did significantly better than students in the baseline condition on Hot Chocolate, which was the performance assessment with a familiar topic. Only the explicit instruction condition had a significant positive effect on the ability to apply science skills in a performance assessment (Bungee Jump) with an unfamiliar topic, thus providing evidence for transfer of skills. Students who received explicit instruction not only performed better than did students in the baseline condition, but they also outperformed students of the implicit instruction condition.

These findings indicated that science lessons can improve skill application in a carefully structured setting with opportunities for students to practice skills in scientific inquiry tasks. The

results also showed that systematic and explicit instruction on science skills may be necessary for more robust acquisition of these skills. Increased awareness of metacognitive strategies applied in the tasks by means of probing questions from the TASC model may have further strengthened skills acquisition. However, we did not find improvement for the paper-and-pencil test and the SMT and as such did not match the outcomes of performance measured by the performance assessments. For the paper-and-pencil test, this lack of progress may be attributed to the test format: the paper-and-pencil test consisted of primarily multiple choice items. Accordingly, the paper-and-pencil test was less similar to a real-life inquiry than it was to the performance assessments, and the tasks that were carried out in the lessons. For instance, formulating a research question for a “real” inquiry is not the same as identifying a research question from amongst different multiple-choice options.

The SMT failed to elicit development of metacognitive skills, which may be due to limited sensitivity of the test. An alternative plausible explanation could be that students overestimated their metacognitive skills. It is therefore conceivable that many students in grades 5 and 6 did not yet have a mastery of these skills even though they thought that they did.

Performance assessment as a diagnostic tool

In the final study we discussed the research question which focused on whether performance assessments have added value as a diagnostic tool to guide instruction in science classroom practice. In chapter 5, we explored and discussed the use of performance assessments for formative assessment to inform teachers and guide instruction of science skills in primary education.

In general, in daily classroom practice, teachers will spend more time and effort on summative assessments rather than on formatively assessing their students’ progress. However, the information on students’ development of science skills is essential for teachers in order to evaluate and improve upon their own instruction, as well as to provide adequate support and feedback on the learning process of individual students. When developing performance assessments, the structure of these assessments should also provide opportunities for formative evaluation in the classroom. This way, a more fine-grained picture of students’ acquired skills may be obtained and used by the teacher to guide students’ learning. In this study, we explored and discussed the utility of the more specific information that may be obtained by structuring the performance assessments according to the different steps of the empirical cycle.

To this end, we examined the mean scores of the 403 students who had participated in all three performance assessments (Skateboard, Bungee Jump and Hot Chocolate) on “step level”. In addition, we analyzed the response patterns and illustrated these patterns with examples of students’ responses.

In general, mean scores showed that the performance assessments were difficult for most 5 and 6 grade students. However, the difficulty seemed to depend on the step-level of the empirical cycle. For instance, designing an experiment appeared in general to be more difficult than formulating a hypothesis. Differences were also visible between performance assessments. For example, in Bungee

Jump and Hot Chocolate the mean scores for “measure and record” were higher than those in Skateboard. The response patterns revealed that much individual variation existed regarding the different steps of the empirical cycle. That is, scores were spread within almost each step, although most students scored in the lower regions of the points awarded. In particular for the steps of designing an experiment, analyzing results and formulating a conclusion, most students attained low scores. In addition, the examples of student responses provided insight into typical errors that students made. For instance, students’ experiment designs were in general not specific and missed the relevant detailed information needed to understand exactly how the experiment should be executed. Common errors found in the step of analyzing results involved neglecting to base the conclusion on the actual data collected by the students.

Based on these findings, we argued that the approach of implementing performance assessments and analyzing students’ responses can be used by teachers to obtain information on how students perform as a group at the classroom level. This approach also reveals the steps with which students have particular difficulty. Because of this, teachers can use this information to adjust their instruction and activities in the classroom. Students’ responses may enable teachers to provide individual students with specific feedback. We were able to conclude that performance assessments may be useful as a diagnostic tool for monitoring students’ skill performance as well as to support teachers in evaluating and improving their science lessons.