UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education

Kruit, P.M.; Oostdam, R.J.; van den Berg, E.; Schuitema, J.A.

DOI
10.1080/09500693.2018.1428777

Publication date
2018

Document Version
Final published version

Published in
International Journal of Science Education

Link to publication

Citation for published version (APA):
Kruit, P. M., Oostdam, R. J., van den Berg, E., & Schuitema, J. A. (2018). Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education. *International Journal of Science Education*, *40*(4), 421-441. https://doi.org/10.1080/09500693.2018.1428777

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:08 Sep 2026

International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: https://www.tandfonline.com/loi/tsed20

Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education

P. M. Kruit, R. J. Oostdam, E. van den Berg & J. A. Schuitema

To cite this article: P. M. Kruit, R. J. Oostdam, E. van den Berg & J. A. Schuitema (2018) Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education, International Journal of Science Education, 40:4, 421-441, DOI: 10.1080/09500693.2018.1428777

To link to this article: https://doi.org/10.1080/09500693.2018.1428777

View supplementary material

Published online: 28 Jan 2018.

Submit your article to this journal

Article views: 996

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tsed20

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2018
VOL. 40, NO. 4, 421–441
https://doi.org/10.1080/09500693.2018.1428777

Routledge
Taylor & Francis Group

Check for updates

Effects of explicit instruction on the acquisition of students' science inquiry skills in grades 5 and 6 of primary education

P. M. Kruit [a], R. J. Oostdam [a,b], E. van den Berg [a] and J. A. Schuitema [b]

[a]Centre for Applied Research in Education, Amsterdam University of Applied Sciences, Amsterdam, Netherlands; [b]Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, Netherlands

ABSTRACT

In most primary science classes, students are taught science inquiry skills by way of learning by doing. Research shows that explicit instruction may be more effective. The aim of this study was to investigate the effects of explicit instruction on the acquisition of inquiry skills. Participants included 705 Dutch fifth and sixth graders. Students in an explicit instruction condition received an eight-week intervention of explicit instruction on inquiry skills. In the lessons of the implicit condition, all aspects of explicit instruction were absent. Students in the baseline condition followed their regular science curriculum. In a quasi-experimental pre-test–post-test design, two paper-and-pencil tests and three performance assessments were used to examine the acquisition and transfer of inquiry skills. Additionally, questionnaires were used to measure metacognitive skills. The results of a multilevel analysis controlling for pre-tests, general cognitive ability, age, gender and grade level indicated that explicit instruction facilitates the acquisition of science inquiry skills. Specifically on the performance assessment with an unfamiliar topic, students in the explicit condition outperformed students of both the implicit and baseline condition. Therefore, this study provides a strong argument for including an explicit teaching method for developing inquiry skills in primary science education.

ARTICLE HISTORY
Received 30 June 2017
Accepted 13 January 2018

KEYWORDS
Science inquiry skills; metacognition; skills acquisition; primary education; explicit instruction

Introduction

Scientific inquiry in primary classrooms generally involves designing, conducting and interpreting results of scientific investigations (National Research Council [NRC], 2012). These investigations are intended to teach skills that are considered important in science, such as formulating a research question and analysing data. One of the intended outcomes is that students develop an understanding of why and how scientific investigations are done (Crawford, 2014). Skills for scientific inquiry are usually – if at all – taught by teaching methods primarily based on learning by doing (Duschl, 2008; Roth, 2014). In the Netherlands for instance, inquiry-based learning is advocated in primary

CONTACT P. M. Kruit p.m.kruit@hva.nl Centrefor Applied Research in Education, Amsterdam University of Applied Sciences, Wibautstraat 2-4, 1091 GM, Amsterdam, Netherlands
Supplemental data for this article can be accessed at https://doi.org/10.1080/09500693.2018.1428777.
© 2018 Informa UK Limited, trading as Taylor & Francis Group

education as the preferred method for acquiring content knowledge, science skills and epistemic knowledge (Graft & Kemmers, 2007). In inquiry-based learning, students are encouraged to discover scientific knowledge and inquiry skills by performing authentic tasks (Kirschner, Sweller, & Clark, 2006). Proponents of inquiry-based learning argue that students learn best by actively constructing knowledge. Also, performing authentic inquiries is generally believed to be motivating for students (Gott & Duggan, 1995). As a result, many educators and curriculum developers have strongly recommended inquiry-based learning for science lessons (Kuhn, Black, Keselman, & Kaplan, 2000; Osborne & Dillon, 2008).

There is a growing body of research showing that explicit instruction is necessary to develop inquiry skills (Klahr & Nigam, 2004; Lazonder & Harmsen, 2016; Toth, Klahr, & Chen, 2000). Several studies have investigated the effects of explicit instruction on skill development in the laboratory setting, in particular, on the strategy of controlling variables (CVS). Only a few studies have compared the effects of inquiry-based learning using explicit instruction with an inquiry-based approach, absent of explicit instruction. The aim of this study is to investigate the effects of explicit instruction in an inquiry-based learning setting on the acquisition of inquiry skills for students in primary education.

Categorisation of skills

In order to understand how students learn and how to support learning, the current literature generally encourages the consideration of skill categories in scientific inquiry (Schraw, Crippen & Hartley, 2006; Duschl, Schweingruber, & Shouse, 2007; Zohar & Dori, 2003).

In this study, the notion 'science-specific inquiry skills' is used to denote a student's ability to design and execute scientific investigations. A distinguishing feature of these inquiry skills is that they are directed to a correct application of rules, principles or conventions about the set-up and execution of an investigation (Millar & Lubben, 1996). Planning and performing an investigation requires explicit knowledge about how to conduct scientific experiments such as identifying and controlling variables, observing and measuring, using simple measurement devices and recording, organising and analysing data (Gott & Murphy, 1987; NRC, 2012).

While engaging in scientific inquiry, students apply more general thinking skills to make sense of the data and connect observations to scientific theories (Osborne, 2015). Thinking skills refer to mental activities such as problem-solving, decision-making, inductive versus deductive reasoning and evaluating solutions (Sternberg, 1986). Thinking skills are employed in scientific inquiry when making appropriate inferences from various sources of data and by drawing adequate samples for making inferences (Pintrich, 2002).

Knowledge and application of metacognitive strategies may regulate students' thinking and learning, thus supporting science inquiry (for overview, see Zimmerman, 2007). Metacognitive skills include planning, monitoring and evaluating task performance. Research shows that students do not apply metacognitive strategies spontaneously. Making students aware of such helpful strategies have shown to improve the performance of inquiry activities (White & Frederiksen, 1998).

For science education, it is relevant to consider the reciprocity of skill development and content knowledge. Content knowledge is most often referred to as a conceptual

understanding of facts, concepts, theories and principles (Sternberg, 1986; OECD, 2017). A number of studies have shown that content knowledge is to a certain extent a prerequisite for skill development (Kuhn, Schauble & Garcia-Mila, 1992; Eberbach & Crowley, 2009). In particular, when students generate hypotheses, make observations, evaluate evidence and draw conclusions, prior content knowledge can have a major effect on skill development (Millar & Driver, 1987; Duschl, Schweingruber & Shouse, 2007).

Explicit skill instruction

Although there is evidence pointing to the acquisition of skills through learning by doing (Dean & Kuhn, 2007), a growing number of studies indicate that more effective learning occurs when inquiry-based learning is accompanied by explicit skill instruction (Kirschner et al., 2006; Duschl, Schweingruber, & Shouse, 2007; Lazonder & Harmsen, 2016). Due to limited experience, most students in primary education lack sufficient mastery of strategies and knowledge to effectively conduct a scientific inquiry. Without explicit skills instruction, this leads to the ineffective performance of scientific inquiry (Klahr & Nigam, 2004; Kirschner et al., 2006).

Engaging in a complex task of scientific inquiry without explicit skills instruction is particularly challenging for inexperienced students since their cognitive information processing capacity is still limited (Flavell, 1992). According to the Cognitive Load Theory (CLT), working memory is limited in its capacity to process new information that contains multiple elements. Elements have to be organised in more complex units and stored in long-term memory before the information can be used effectively (Van Merriënboer & Sweller, 2005). Once this is achieved, information stored in long-term memory is accessible when needed, aiding the acquisition of inquiry skills (Kirschner et al., 2006).

Much of what is known about the effects of explicit instruction comes from studies on CVS (Lazonder & Egberink, 2014; Matlen & Klahr, 2013). For example, Chen and Klahr (1999) found in an intervention study with third and fourth graders that explicit instruction combined with probing questions (i.e. why they designed the investigations the way they did and what they had learned) was an effective way of learning how to apply CVS. This is in line with CLT because explicit forms of instruction put less of a burden on working memory when learning new information (Kirschner, et al., 2006). Dean and Kuhn (2007) showed that in particular explicit instruction (in which students were asked to compare and identify different features of catalogues) improved students' CVS more when combined with practice. The positive effects of explicit instruction may also apply to the acquisition of metacognitive skills. Explicit attention to awareness of the task and the metacognitive strategies may facilitate skill development (Pintrich, 2002; Zohar & Dori, 2003; Tanner, 2012).

Explicit skills instruction seems to be particularly important for fostering the transfer of learning. More robust learning of skills has only been achieved when students are able to apply the skills in contexts other than the one in which the skills are learned. Although there is not a clear-cut definition of what different transfer distances entail (Chen & Klahr, 1999), near-transfer can generally be defined as the application of skills in tasks within a particular knowledge domain or with a common structure. Far-transfer is defined as the application of skills in tasks in different domains or tasks with an unfamiliar structure (Strand-Cary & Klahr, 2008).

While achieving transfer across knowledge domains (i.e. topics) is generally difficult (Kuhn et al., 1995; Lazonder & Egberink, 2014), some CVS studies have shown that explicit instruction can facilitate transfer of skills to other tasks with different topics (Kuhn et al., 1995; Klahr & Li, 2005). Likewise, research indicates that young students tend to fail in using the same strategies for performing tasks with different topics (cf. Chen & Klahr, 1999). 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 (Adey & Shayer, 1993; Chen & Klahr, 1999; Georghiades, 2000).

Four-component instructional design

When designing science lessons aimed at the development of science skills, the four-component instructional design (4C/ID) model can be applied. The model may be in particular suitable because the model focuses on the integration and performance of skills, as opposed to content knowledge (Van Merriënboer, Jelsma, & Paas, 1992). Furthermore, the model recommends a combination of tasks in which skills are first practiced separately and then applied in more complex tasks in an integrated manner (Van Merriënboer, Clark, & de Croock, 2002). Most of the difficulty in learning complex skills such as science skills is in applying them simultaneously. In existing design models, it is often assumed that complex skills acquired in simple tasks will be applied spontaneously to new and more complex tasks despite considerable evidence to the contrary (Van Merriënboer, et al., 2002, p. 40). According to the 4C/ID model, therefore, the following components are essential for developing skills: (1) whole learning tasks, (2) part-task practice, (3) supportive information and (4) just-in-time information.

Whole-tasks represent authentic and meaningful scientific inquiries. Students receive explicit instruction and additional skills practice in *part-tasks*. The acquired skills can then be directly applied to more complex whole-tasks. This enables students to see the interrelationships between part-tasks and the task as a whole, stimulating the integration of skills (Van Merriënboer, et al., 2002). By segmenting the complex scientific inquiry activity into manageable smaller and structured part-tasks in which students can learn and practice skills, performance can be enhanced (Lazonder & Egberink, 2014; Lazonder & Kamp, 2012). Whole-tasks and part-tasks are preferably combined and sequenced from relatively simple to more complex (Van Merriënboer & Sweller, 2005; Wu & Krajcik, 2006).

The structuring of the whole-tasks and part-tasks can be effectuated in science lessons by using the different steps of the empirical cycle as a design principle: (1) formulating a research question, (2) formulating hypotheses, (3) designing an experiment, (4) measuring and recording data, (5) analysing and interpreting data and (6) evaluating the outcomes in relation to the research question and hypotheses. Although it is generally acknowledged that scientific inquiry is not a linear process (NRC, 2012), the subsequent activities of the empirical cycle provide a structure that is recognisable for students. It also gives an understanding of how the inquiry process may be organised, which is particularly important for students in primary education who have little experience with scientific inquiry (Donovan, Bransford & Pellegrino, 1999; White & Frederiksen, 1998).

Another aspect of the 4C/ID model is *supportive information*, which comprises the information students do not possess needed for performing a particular scientific inquiry. Finally, *just-in-time information* refers to essential clues, knowledge or feedback

Figure 1. TASC framework.

students have access to the moment they need it during task performance. Just-in-time information can consist of prompts which are gradually withdrawn as students gain proficiency (McNeill et al., 2006). Prompts can take the form of generic 'probing questions' or hints to encourage reflection which helps students to set goals and monitor their understanding (Sahin & Kulm, 2008). Eventually, prompts can gradually be withdrawn until students can apply the acquired skills independently (White & Frederiksen, 2000).

An example of implementing prompts for explicit instruction of metacognitive skills is the application of the TASC framework. TASC stands for 'Thinking Actively in a Social Context' and aims to give young students structure to support their thinking (Wallace, Bernardelli, Molyneux, & Farrell, 2012). Students can be instructed on how to move systematically through the stages of the TASC framework while performing a task. In each stage, several questions can be raised to stimulate students to monitor and evaluate their performance (Figure 1). For instance, the students are asked to think about what they already know about the topic of an experiment, how much information they already have and what information they need (Wallace et al., 2012). These questions are introduced and eventually withdrawn gradually until students are familiar with the questions and apply the metacognitive skills in each following experiment by themselves (White & Frederiksen, 2000).

The present study

The present study adds to the current discussion by investigating the effects of explicit instruction on the acquisition of inquiry skills in classroom settings in grades 5 and 6. Furthermore, it examines to which extent the acquired skills are utilised in tasks with content different from the tasks used in the science lessons. Contrary to most studies (Shavelson, Baxter, & Pine, 1991), skills were not only evaluated by a paper-and-pencil test (PPT), but also by more authentic performance assessments. Additionally, questionnaires were used for evaluating improvement in metacognitive skills.

The following research questions concerning explicit skill instruction in grades 5 and 6 were addressed: (1) What are the effects on students' skills in scientific inquiry? and (2) What are the effects on transfer of students' skills across science tasks with unfamiliar content?

In line with current knowledge on enhancing skill development, it was hypothesised that receiving explicit skill instruction would positively affect students' science skills and that the merits of explicit instruction would also extend to transfer of students' skills.

Method

Participants

This study was conducted at 12 schools for primary education in the Netherlands. It involved schools that were part of the school network with which the teacher training education of the Amsterdam University of Applied Sciences cooperates. As a result, the schools were located in the urbanised part of the Netherlands. Schools were willing to participate on the basis of several pragmatic factors including permission of school authorities, interest of teachers, willingness to 'do something with science' and available time.

To ensure that the reading and writing abilities of students would not be a limiting factor in relation to the acquisition of science skills, only students in upper grades were included. In total, 705 students participated in the study (51.3% boys and 48.7% girls) with a mean age of 11.5 (SD = .69) from 31 grades 5 and 6 classes (53.3% in grade 5 and 46.7% in grade 6).

Research design

The research constituted a quasi-experimental study with pre-test–post-test design, implemented in grades 5 and 6 of primary schools, designed to investigate the effects of explicit instruction on students' acquisition of inquiry skills (explicit condition). A control condition was included with lessons in which skills were taught in an inquiry-based approach without explicit instruction on inquiry skills so that information about the added value of explicit instruction could be obtained. To contrast both controlled conditions with regular science lessons at schools, a baseline condition was added. Randomisation was carried out within schools at class level.

A total of 705 students participated in the pre-test sessions. Pre-testing included three different measures: a paper-and-pencil test (PPT), a performance assessment (PA) and a metacognitive self-report test (Junior Metacognitive Awareness Inventory [Jr. MAI]). After an 8- to 10-week period students of all three conditions were tested again with a PPT, two PAs (PA-*related-content* with a topic discussed in lessons of both interventions

and PA-*transfer* with an unfamiliar topic designed to measure transfer of skills), Jr. MAI and an extra metacognitive self-report test (SMT). Additional measures for the explicit and implicit conditions included an implementation checklist on class level and a questionnaire for evaluating how enjoyable the lessons were perceived by the students.

PAs were administered in small groups of students outside of the regular classroom, which required additional time and effort for the schools and students. Therefore, to reduce the burden, a subsample ($n = 467$) was randomly selected to partake in the post-test PAs. The subsample was created by random selection of students in each class. To enable comparison between tests, only students who had completed all tests were included for final analysis. Of the subsample, 62 students were excluded due to having been absent from one of the test sessions. Finally, one class which consisted of only two students was dropped from the analysis to prevent estimation bias in multilevel analysis (Maas & Hox, 2005). The final sample consisted of 403 students (Figure 2).

Intervention

The lessons for the intervention were developed by the first author and a primary school teacher. The aim was to enhance development of skills associated with scientific inquiry. Each intervention consisted of eight lessons of 90 minutes each. In general, one lesson was given per week. During this time, students in the baseline group followed their regular curriculum and did not receive any formal instruction on scientific inquiry.

Explicit condition

The lessons for the explicit condition were designed based on the 4C/ID model of Van Merriënboer, et al. (1992). The whole-tasks and part-tasks were structured according to

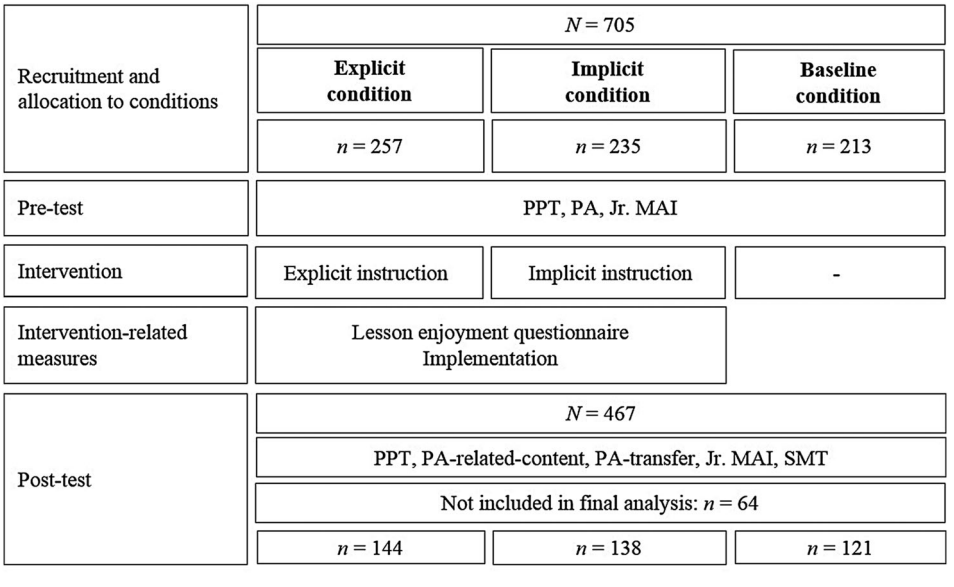

Figure 2. Overview of the study design.

the different steps of the empirical cycle. The topic was heat and temperature. The content knowledge was minimised in the sense that it was addressed at such a low level such that it could be assumed that it would pose no obstacle in skill application. For instance, students at grade levels 5 and 6 are aware of the fact that hot water cools down and that temperature can be measured by using a thermometer.

Explicit instruction concerned clarifying the rationale of inquiry skills by the teacher followed by examples and classroom discussions about how to apply the skills. In part-tasks, the newly learned skills were then practiced with the support of written prompts. For example, in learning how to formulate a research question, the teacher gave explicit instruction about the criteria for formulating research questions with the help of a flow chart (Figure 3). In subsequent part-tasks where students were asked, for instance, to distinguish between properly and poorly formulated research questions, students were reminded of the flow chart. Next, students performed a whole-task: an authentic, structured inquiry in which students had to apply all skills in an integrated manner. In these whole-tasks, prompts were explicitly incorporated as well. Development of metacognitive skills was supported by the use of probing questions from the TASC framework prior to and during each whole-task (Wallace et al., 2012).

During the course of the intervention, the whole-task inquiries gradually increased in difficulty and complexity while at the same time, the prompts were withdrawn. In the final lesson, students performed a scientific inquiry independently.

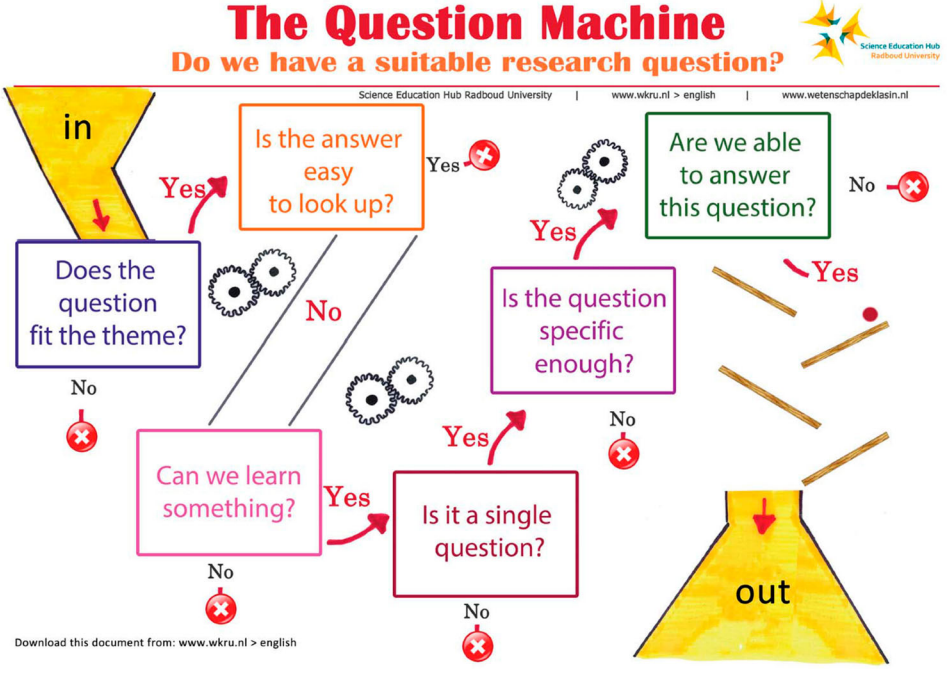

Figure 3. Flow chart with criteria for formulating a research question (Science Education Hub Radboud University, 2016).

Implicit condition

The lessons of the implicit condition were structured based on the 4C/ID model as well. Students in the implicit condition also performed inquiry tasks that were structured according to the different steps of the empirical cycle. All aspects of explicit instruction were absent which entailed that the teacher refrained from explaining the rationale behind the skills and that in part-tasks and whole-tasks, skills were practiced without explicit probing questions or prompts. Rather than giving explicit attention to separate skills first and then part-task practice on skill level, all skills were applied and practiced simultaneously. To control for time on task, the implicit intervention contained part-tasks about content and additional topics were included (see Appendix 1 for both explicit and implicit example tasks). Two lessons concerned growth and development of plants, two lessons on heart and lungs followed by the main topic of heat and temperature (four lessons). Ultimately, the implicit intervention comprised similar numbers of scientific investigations and assignments (Table 1).

Both interventions were piloted in several grades 5 and 6 classrooms which did not participate in the main study. Adjustments were made where necessary, which in general concerned length of assignments or difficulty level of scientific inquiry tasks.

Baseline condition

In Dutch primary schools, science lessons are primarily textbook-based. Developing inquiry skills has low priority in science primary education, indicated by the lack of implementation of instruments to measure progress of skills (Inspectorate of Education, 2015). Inquiry tasks are included at most once a month or less and only by 50% of the teachers (Kneepkens, Van Der Schoot & Hemker, 2011). Students in the baseline condition did not receive formal and structured instruction on inquiry skills during the intervention period.

Applying the hypothesis to the experimental design, it was predicted that students in the explicit condition would improve their ability to perform a scientific inquiry more than students in the baseline condition. Furthermore, it was expected that students in the explicit condition would also outperform students in the implicit condition since the benefits of explicit instruction, including prompts, task-structuring and direct

Table 1. Similarities and differences between the explicit and implicit condition.

	Aspect	Explicit condition	Implicit condition
Similarities	Duration	Eight lessons of 90 minutes	Eight lessons of 90 minutes
	Whole-task	Authentic investigations	Authentic investigations
Differences	Structure of investigations	Investigations explicitly structured for the students according to steps of empirical cycle	Investigations implicitly structured for the students according to steps of empirical cycle
	Part-task	Part-tasks for skill training	Part-tasks for acquisition of content knowledge
	Just-in-time information	Prompts	No prompts
	Metacognition	Explicit attention to planning, monitoring and evaluation	No attention to planning, monitoring and evaluation
	Instruction	Direct instruction on skills	No direct instruction on skills
	Content of lessons	Heat and temperature	Heat and temperature, growth and development of plants, heart and lungs

instruction have been well-established (Lazonder & Egberink, 2014; Van Merriënboer & Sweller, 2005; Zohar & Dori, 2003). Furthermore, it was predicted that the explicit instruction combined with part-task practice aimed at skill acquisition would be specifically beneficial for skill performance in the transfer task (Chen & Klahr, 1999). However, since practicing and applying skills in whole scientific inquiry tasks alone can also enhance skill development to some extent (Dean & Kuhn, 2007), it was expected that students in the implicit conditions would also improve science skills more in comparison to students in the baseline group.

Procedure

The lessons of both interventions were taught by research assistants who had been recruited and trained specifically for this purpose. All assistants had either graduated or were in the final year of an elementary teacher education programme. The students' regular teacher was asked to assist in each lesson. The presence of a familiar teacher would help maintain a good working atmosphere while teaching. All lessons were taught in the students' regular classroom and with their regular classmates.

The research assistants were trained in a 4-hour training session to teach either one of the interventions. Training included instruction, discussion and practice of parts of lessons. In addition, assistants were trained in how and when to provide feedback during tasks. The assistants were provided with a practical guide containing a description of the rationale behind the intervention, detailed lesson plans and information on the skills and content to be covered in the lessons. In addition, to ensure that all assistants would stay on the same track (and no parts would be skipped), PowerPoint presentations were developed for each lesson. In that way, it was assured that all assistants would implement the lessons in a similar way.

To check the faithfulness of the implementation, the regular teacher was asked to monitor the degree of implementation of the activities of each lesson. In addition, the research assistants were monitored by two of the authors by observing a randomly chosen lesson. Throughout the intervention, research assistants and the main author kept in close contact by discussing lessons regularly. Furthermore, each research assistant was asked to film lesson 7 of which random parts were watched by the author for a fidelity check. Additionally, information on student level was collected documenting lesson attendance and to what degree the students found the lessons enjoyable, because both may influence students' participation and scores.

Teachers in the baseline group did not receive any of the materials used in the interventions. However, to be assured of cooperation, they could choose to complete one of the two interventions with materials and trained assistants after the experiment.

Measurement instruments

Paper-and-pencil test

Students' inquiry skills were measured by means of a PPT. The test comprised a total of 46 items (36 multiple choice and 10 open-ended), which were selected from large-scale assessments commonly used in the Netherlands (Cito.com) and other sources (e.g.

SOLpass.org). Content varied from item to item and was considered to be familiar for grades 5/6 students. The 10 open-ended questions were scored by trained raters.

For pre-test–post-test purposes, the items were divided over two test booklets to give two optimal split half tests. A total score for each test was calculated with a maximum possible score of 30 points. The Cronbach's alpha coefficients of the pre-test and post-test were .63 and .70 respectively ($N = 403$).

Performance assessments

Students were additionally assessed by three PAs. All three PAs concerned comparative investigations: students are asked to examine the relationship between two variables (Shavelson, Solano-Flores, & Ruiz-Primo, 1998). Each PA was constructed according to the same template, following the different activities of the empirical cycle and allowing for comparison of students' scores between PAs. PAs differed only in topic of investigation. In the PA-*pre-test Skateboard*, students roll a marble down a ruler to examine the relation between the distance of the marble on the ruler and the distance the marble covers at the end of the ruler while pushing forward a paper wedge (Ahlbrand, Green, Grogg, Gould, & Winnett, 1993). The PAs *Hot chocolate* and *Bungee jump*, based on tasks in TIMSS (Martin et al., 1997), were deployed for post-testing. *Hot chocolate* (PA-*related-content*) concerns the relationship between an amount of hot water and its rate of cooling, which corresponds to the topic of heat and temperature addressed in both intervention groups. In *Bungee jump* (PA-*transfer*), students investigate the changing length of a rubber band as additional weights are added, a topic unfamiliar to both groups. This last test served to assess the transfer of skills to a new task.

Each PA contained 14 quantifiable items with a maximum score of 34 points (see Appendix 2 for PA-*pre-test Skateboard*). Scoring of items was based on students' answers written down in worksheets. The Cronbach's alpha coefficients of PA-*pre-test*, PA-*related-content* and PA-*transfer* were .67, .70 and .67 respectively ($N = 403$).

Metacognitive self-report tests

Metacognitive skills were measured by means of two different self-report tests. For pre- and post-testing purposes, the Jr. MAI, a self-report inventory for grades 3–5 developed by Sperling, Howard, Miller and Murphy (2002) was administered. Jr. MAI consisted of 12 items with a three-choice response (never, sometimes or always). The mean score was calculated to obtain a measure of general metacognitive ability. The pre-test and post-test Cronbach's alpha coefficients were .54 and .64, respectively.

The second metacognitive self-report test – Science Meta Test (SMT) – measures metacognitive self-regulatory skills including orientation/planning, monitoring and evaluation (Schraw & Moshman, 1995). In contrast to the more general Jr. MAI, items were constructed to obtain information about the extent to which metacognitive skills are applied specifically in the PAs in post-testing. A mean score was calculated of the total of 13 items with a three-point scale (not, a little, a lot). Cronbach's alpha coefficient was .77 ($N = 403$).

General cognitive ability

To be able to control for general cognitive ability, students' scores were obtained for Reading comprehension and Arithmetic/mathematics from a semi-annual assessment of The

National Institute for Educational Testing and Assessment. For Reading comprehension and Arithmetic/mathematics, reliability scores indicated by MAcc (Accuracy of Measurement) are >.87 and >.95, respectively, for grades 5 and 6 (Janssen, Verhelst, Engelen, & Scheltens, 2010; Weekers, Groenen, Kleintjes, & Feenstra, 2011). Ability is expressed by different levels which indicate the actual performance level of a student compared to a norm group (A = upper 25% of all children, B = 25% above mean, C = 25% below mean, D = next 15% below C, E = lowest 10%). The scores, provided by the participating schools, were transformed into a five-point scale (A = 5 to E = 1). A student's general cognitive ability score was constructed by means of summing up the scores of both tests.

Intervention-related measures

Integrity of implementation

Degree of implementation was measured by the percentage of instruction and activities of each lesson actually carried out. A class level variable representing the degree of implementation was calculated by averaging the percentages of the number of lessons of each class.

Lesson enjoyment

Students were asked to fill out a short questionnaire of seven items with a five-point Likert scale at the end of the final lesson of the intervention. The original questionnaire was developed for measuring enjoyment of math lessons by Martinot et al. (1988). The items were adapted for primary science lessons and an extra item was added. A final score was calculated by summing the scores (maximum of 28) of each item ($\alpha = .84$).

Administration procedure

The PPT-pre-test, PA-pre-test and Jr. MAI-pre-test were administered in this specific order just before the start of the intervention. Post-testing comprised the PPT-post-test, PA-*related-content*, PA-*transfer* and the two metacognitive self-report tests. To control for sequencing effects, administration of PA-*related-content* and PA-*transfer* was randomly alternated. Post-Jr. MAI and SMT were administered directly after PA-transfer, taking just a few minutes to complete.

All tests were administered by trained test leaders who followed detailed protocols for test administration. The PAs were administered individually in groups of four to eight students. The PPT and PA each took 45 minutes to complete.

Scoring

Trained raters scored the open-ended questions for the PPTs and PAs. Interrater reliability was calculated by determining intra-class correlation after a training session on a random sample of 11% of the tests. Additionally, another average of 19% of the scores per rating session was scored by two raters so that additional interrater reliability could be estimated during the process of individual rating. First round ICC ranged from .80 to .94 (ICC, two-way random, absolute agreement), and second round from .66 to .93. Finally, for each rater, stability of scoring was estimated which ranged from .81 to 1.00.

Analysis

An a priori power analysis indicated that a sample of 68 students would be sufficient to test a main intervention effect of medium size with a statistical power of $\beta = .80$ at the conventional alpha level of .05 with multiple regression analysis. The sample of more than 400 students allows a test of small-to-medium effect sizes, also taking into account intracorrelations (ICC values between .10 and .30).

The dataset included one independent variable (condition: explicit instruction, implicit instruction, baseline) and five dependent variables (Post-PPT, PA-*related-content*, PA-*transfer*, Post-Jr. MAI and SMT). Because of the nested structure of the data, multilevel models were used to investigate the effects of the conditions on the five dependent variables. MLwiN for multilevel analysis (Rasbash, Steele, Browne, Goldstein, 2009) was used to fit five different models (one for each dependent variable) with two levels (students within classrooms). Adding a third level for school did not significantly improve the fit of the models.

The student variables (age, gender and general cognitive ability), grade level and pre-test scores were included as control variables since they might influence students' scores. Variables were examined for accuracy of data entry, distributions and missing values. All underlying assumptions (i.e. normality) were met.

As explained above, 64 students were not present at one or more test sessions. These students were excluded from further analysis (Figure 2). There were no missing values on individual items in the PPTs and PAs. There was a maximum of 6.7% missing values on items of all three metacognitive self-report tests. Little's MCAR test including all variables was not significant, indicating that no identifiable pattern existed in the missing data (χ^2 (5552) = 5470.73, $p = .78$). Of the students who were present for all test sessions, the expectation-maximization imputation method was performed for the missing items of each metacognitive self-report test separately.

Ultimately, the Jr. MAI test was not included in the analysis because of insufficient reliability of the pre-test ($\alpha = .54$) and a serious lack of variance at the post-test.

For the benefit of interpretation of the intercept, all continuous control variables were centred around the grand mean.

Results

Descriptive statistics

Table 2 presents the effects of randomisation on the composition of the three conditions. Conditions were comparable with regards to gender (χ^2 (2) = 2,3, $n = 403$, $p = .319$), but not to grade level, with more students in grade 6 in the implicit condition compared to

Table 2. Comparison of conditions in terms of the control variables.

		Gender %		Grade %		General ability	Age	PPT-pre-test	PA-pre-test
Condition	*N*	Boy	Girl	5	6	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Explicit	144	46.5	53.5	54.9	45.1	7.15 (2.35)	11.44 (.65)	15.04 (4.21)	9.69 (4.29)
Implicit	138	54.3	45.7	42.0	58.0	6.80 (2.46)	11.61 (.72)	15.52 (4.51)	11.24 (5.07)
Baseline	121	46.3	53.7	58.7	41.3	6.92 (2.24)	11.34 (.62)	15.88 (4.51)	10.14 (5.64)

Note: Maximum possible score of PPT-pre-test is 30; and of PA-pre-test is 34 points.

the other conditions (χ^2 (2) = 8,10, n = 403, p = .017). Analyses of variance indicated that students between conditions did not differ significantly on general cognitive ability ($F_{(2, 400)} = 0.77$, $p = .465$). There was a significant difference in age ($F_{(2, 400)} = 5.13$, $p = .006$), with students in the baseline condition being somewhat younger than students in the implicit condition (mean difference = −0.26, CI = −0.46 to −0.07). Of the pre-tests, conditions did not differ in scores for the PPT-pre-test ($F_{(2, 400)} = 1.20$, $p = .302$). However, analysis showed significant differences between scores on the PA-pre-test ($F_{(2, 400)} = 3.54$, $p = .030$) with a lower score for the explicit condition.

Descriptive statistics of intervention-related measures

Of the total of 168 lessons in the explicit and implicit condition taught in 21 different classrooms, an implementation score was missing from four lessons. For the classes with missing scores, an average score was calculated on the basis of the lessons scored. Multilevel regression analysis showed that the degree of implementation for the implicit condition ($n = 144$) was not significantly different to that of the explicit condition ($n = 138$), ($B = 2.38$, $SE = 0.176$, $p = .174$).

An independent-samples t-test showed no significant difference in enjoyment ($t(237) = 1.56$, $p = .121$, equal variances not assumed) between students of the implicit condition ($M = 18.11$, SD = 5.40) and those of the explicit condition ($M = 16.88$, SD = 7.20),

Finally, an independent-samples t-test indicated that students' number of attended lessons did not differ significantly between the explicit condition ($M = 7.60$, SD = .73) and the implicit condition ($M = 7.70$, SD = .64), $t(280) = -1.12$, $p = .264$.

Descriptive statistics of post-tests and correlations

Table 3 presents the means and standard deviations for all post-tests. Because scales of the scores are different for each test, the means are not comparable one-on-one.

Table 4 shows the correlations (Pearson's r) between tests. No significant correlations were found between the SMT and other tests, other than PA-*transfer*. Results show that general cognitive ability is considerably correlated with pre-tests as well as the PPT-post-test, whereas correlations with post-PAs are smaller.

Effects of interventions on skills in scientific inquiry

Table 5 presents the results of the multilevel analysis on the effects of condition on skills in scientific inquiry, controlling for pre-tests, general cognitive ability, age, gender and grade

Table 3. Means and SD for post-tests.

		Explicit ($n = 144$)		Implicit ($n = 138$)		Baseline ($n = 121$)	
	Max score	*M*	SD	*M*	SD	*M*	SD
PPT	30	16.56	4.59	16.52	4.94	15.40	5.02
PA-*transfer*	34	13.72	4.69	12.53	5.59	11.30	5.07
PA-*related-content*	34	13.67	4.74	13.49	5.50	10.39	4.88
SMT	3	2.12	.41	2.11	.35	2.16	.31

Note: Max score = maximum score possible of test; Mean and SD of SMT are indicated by average item scores.

Table 4. Correlations between all tests ($N = 403$).

		1	2	3	4	5	6
1	PPT-pre-test						
2	PPT-post-test	.60*					
3	PA-*pre-test*	.54*	.52*				
4	PA-*transfer*	.39*	.47*	.52*			
5	PA-*related-content*	.36*	.45*	.43*	.63*		
6	SMT	.02	−.03	.04	.15*	.01	
7	General cognitive ability	.58*	.57*	.49*	.35*	.34*	−.04

Note: *, correlations are significant at $p < .01$ (two-tailed).

level. Results show that the effects of age and grade level were not significant. The two pre-tests and general cognitive ability positively contributed to the scores of all three post-tests. There was also an effect of gender: girls scored significantly higher on the tests than did boys. Adding the condition variable did not significantly increase model fit for the PPT (Δ *IGLS* = 3.64, df = 2, $p = .162$). Effect sizes in terms of Cohen's d (Cohen, 1992) were small for both the explicit condition ($d = 0.27$) and implicit condition ($d = 0.20$). These results indicate that there was no significant effect due to the interventions on the ability of students to apply skills in the PPT. In contrast, condition did have a positive significant effect on scores of the dependent variable PA-*related-content* (Δ *IGLS* = 15.84, df = 2, $p = <.001$). Students of both intervention conditions did significantly better than students in the baseline condition with medium effect sizes $d = 0.66$ for the explicit condition and $d = 0.58$ for the implicit condition. The interventions also contributed to a significant better model fit for PA-*transfer* (Δ *IGLS* = 8.09, df = 2, $p = .018$). However, in contrast to PA-*related-content*, only the estimate for the explicit condition was significantly higher than the baseline condition ($B = 2.48$, $SE = 0.82$, $p = .003$) with medium effect size $d = 0.48$. The effect size for the implicit condition was small ($d = 0.18$). In other words, only explicit instruction had a positive effect on the ability to perform an investigation with a new and unfamiliar topic.

Table 5. Results of the multilevel analyses for each dependent variable ($N = 403$).

	PPT-post-test			PA-*content-related* post-test			PA-*transfer* post-test			SMT		
	Coeff.	SE	*p*	Coeff.	SE	*p*	Coeff.	SE	*p*	Coeff.	SE	*p*
Intercept	14.83	0.61	<.001	8.90	0.73	<.001	9.87	0.73	<.001	2.18	0.06	<.001
PPT-pre-test	0.33	0.05	<.001	0.14	0.07	.028	0.13	0.06	.042			
PA-pre-test	0.19	0.05	<.001	0.21	0.06	<.001	0.35	0.05	<.001			
General ability	0.61	0.10	<.001	0.45	0.12	<.001	0.35	0.12	.003	0.00	0.01	n.s.
Age	-0.08	0.39	n.s.	-0.08	0.48	n.s.	0.55	0.45	n.s.	0.02	0.04	n.s.
Gender[a]	0.79	0.35	.021	2.65	0.42	< .001	2.73	0.39	<.001	−0.02	0.04	n.s.
Grade[b]	0.35	0.61	n.s.	0.15	0.74	n.s.	0.07	0.71	n.s.	−0.03	0.06	n.s.
Condition												
Explicit condition[c]	1.31	0.68	.055	3.48	0.81	<.001	2.48	0.82	.003	−0.04	0.07	n.s.
Implicit condition[c]	0.98	0.70	n.s.	3.05	0.83	<.001	0.91	0.84	n.s.	−0.05	0.07	n.s.
Explained variance %												
Group level	61.07			65.86			60.50			0.00		
Student level	45.39			26.94			34.22			0.00		
ICC	0.12			0.11			0.14			0.09		

Note: n.s., not significant.
[a]Reference category: boys.
[b]Reference category: grade 5.
[c]Reference category: baseline condition.

To further explore the difference between the two intervention conditions, multilevel analysis was performed on a subsample of students ($N = 282$) who had received either explicit or implicit instruction. In addition to the pre-tests, general cognitive ability, age, gender and grade, the treatment fidelity variable 'implementation' was added as class-level control variable. Only on the performance assessment PA-*transfer*, the model fit improved significantly by adding the condition (Δ *IGLS* = 4.98, df = 1, $p = .026$). Specifically, the students who received explicit instruction performed significantly better ($B = 2.00$, $SE = 0.86$, $p = .016$) than students in the implicit condition. The magnitude of the effect could be considered moderate with $d = 0.40$.

For the metacognitive skills evaluated by the students with the SMT, no additional variance was explained by condition, indicating that condition had no effect on scores.

Discussion

Results of this study confirm that both experimental interventions facilitate the acquisition of skills for performing scientific inquiry. The scores on the PAs show that students in both intervention conditions have significantly higher scores than students in the baseline condition. In particular, when skills are measured by means of a PA with a topic familiar to the content of the lessons, substantial effects are found. This finding is consistent with former research on the influence of content knowledge on developing skills (Duschl, Scheingruber & Shouse, 2007).

Furthermore, comparison of the two experimental conditions reveals that students of the implicit condition were almost just as able to perform the scientific inquiry on a familiar topic as students who received explicit instruction. Although most studies (on CVS) show more effective learning by explicit instruction, this finding suggests that in a carefully structured setting the opportunity to practice skills alone can already improve skill application which concurs with former findings (Dean & Kuhn, 2007).

On a PA with an unfamiliar topic, scores are particularly interesting. Indicated by a medium effect size, students in the explicit condition clearly outperform students in the implicit condition. In both conditions, students had been practicing the same number and variety of inquiries but only students subjected to explicit instruction were able to apply inquiry skills in a PA with unfamiliar topic. This indicates that explicit instruction facilitates the transfer of skills, which seems concordant with findings in CVS studies (i.e. Chen & Klahr, 1999; Klahr & Li, 2005). In the present study, systematic and explicit attention (i.e. by prompts) to skills in new tasks may have promoted more robust acquisition of these skills. Increased awareness of strategies applied in the tasks by means of probing questions may have further strengthened skills acquisition. The decontextualisation of skills has possibly contributed to students' ability to apply the skills they acquired by practicing tasks with different science content, albeit in the same domain (near-transfer). This implies that students were not only able to use these skills, but actually understood how to apply them, which has been shown to be difficult to accomplish in a classroom setting (cf. Klahr & Li, 2005).

The improvement of skills measured by the PPT was much smaller – though almost significant – for students in the explicit condition and as such did not match the outcomes of performance measured by the PAs. This discrepancy may be due to the different test format. Skills measured by the PPT were elicited in a more passive manner in contrast

to applying skills actively in a PA which is considered a more authentic way of assessing skills (Shavelson et al., 1998), thus measuring actual skill proficiency. For instance, formulating a research question for a 'real' inquiry is not the same as identifying a research question among different multiple-choice options. Accordingly, a PPT may not be the most appropriate way to assess skill performance. The structure of the items in the PPT was less similar to the tasks in the intervention lessons and as a result students could depend less on skills learned in the lessons. Instead, students had to rely more on general abilities in the PPT than in the PAs, which is supported by the small/medium correlation found between post-test PAs and general cognitive ability ($r = .35$ and .34) and the medium/large correlation between the PPT and general cognitive ability ($r = .57$).

Results show no significant difference in scores between students in grades 5 and 6, even when age was removed from the model. Upon first glance, this is a surprising outcome as students are likely to have improved their skills just by growing older and acquiring more general skills by the time they are in grade 6 (Duschl, Schweingruber & Shouse, 2007). This may indicate that students participating in this study had very little prior experience with scientific inquiry, thus were novices regardless of what grade they were in. The implication may be that both the design of the instructions as well as the implemented assessment formats are suitable for students of both grade levels, provided that all students share the same (lack of) experience with scientific inquiry.

Despite the improvement of skill proficiency, the SMT failed to elicit development of metacognitive skills. However, it is very unlikely that students – as the high scores above the scale mean suggest – already possess the ability to apply metacognitive skills to scientific inquiry and therefore could not improve their metacognitive skills at all, regardless of any instruction. A more plausible explanation could be that students overestimate their metacognitive skills. This is in line with Veenman, Van Hout-Wolters and Afflerbach (2006) who stated that scores on questionnaires 'hardly correspond to actual behavioural measures during task performance'. It is, therefore, conceivable that many students in grades 5 and 6 do not yet have a mastery of these skills even though they think they do. Nevertheless, it may be that the metacognitive skills did improve and – although not directly measured by the SMT – are indirectly reflected in higher scores on the PAs (Georghiades, 2000).

Limitations and suggestions for future research

An important limitation of this study is that teaching assistants were using the lesson materials for the first time. In a repeat teaching, assistants may be more efficient and effective which could result in greater gains compared to the baseline condition.

A second limitation in the design of this study is that only two post-PAs were included. Assessing skill proficiency with more PAs would be preferable in order to reduce occasion sampling variability, i.e. students performing the same task differently on different occasions (Ruiz-Primo, Baxter and Shavelson, 1993). However, the PAs in this study were made quite elaborate in a bid to resemble authentic research. Deploying more of such extensive assessments in real classroom settings would be laborious for students and require too much testing time.

Furthermore, in this study, the PAs are highly structured and not, as such, representative of scientific inquiry, which is not generally a linear process. However, for students who are novices, open investigations may be too challenging. In further research, students

could be assessed as their proficiency increases, using more open and interesting investigations. Along the same lines, although the PPT was easy to administer and less difficult to score, it may not have been the best choice for measuring progress of skills of the novice learners in this study. The PAs provided a more elaborate and detailed picture of the acquired skills and may have been as such more sensitive to change.

Finally, the post-tests were administered directly following the intervention. This implies that only the short-term effects on inquiry skills of the explicit and implicit teaching method were measured. A retention test was not included in the design because the grade 6 students were no longer available for another round of post-testing. Future research might examine retention and/or transfer by using tasks with a different or more open structure.

Implications for educational practice

The current study has several practical and scientific implications. For students in primary school who have little experience with scientific inquiry, systematic, explicit instruction should be considered and incorporated as a starting course for developing skills. As students gain more proficiency in both skills and content knowledge, more open and interesting inquiry activities may be implemented. Ultimately, the skills can also be deployed for inquiry-based learning of content in addition to learning inquiry skills.

Furthermore, the 4C/ID system of Van Merriënboer, et al. (1992) has proved to be useful in designing structured lessons for teaching inquiry skills. The task-centred approach provided ample opportunity to systematically incorporate important design principles such as explicit instruction and additional skills practice. The whole-tasks were not only pleasurable and motivating for students, but also provided the possibility to apply the inquiry skills in an integrated manner, supporting development of the skills.

To conclude, the present findings have provided additional information to the discussion of effective ways to teach inquiry skills to young students. In this study, the method of explicit instruction was not only compared with a baseline condition, but also with the implicit method of teaching skills. Therefore, this study provides a strong argument for implementing an explicit teaching method to promote and develop inquiry skills in primary science education.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

P. M. Kruit http://orcid.org/0000-0002-3734-1904
R. J. Oostdam http://orcid.org/0000-0003-4701-0153
J. A. Schuitema http://orcid.org/0000-0002-0125-5537

References

Adey, P., & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention program in the high school science curriculum. *Cognition and Instruction*, *11*(1), 1–29.

Ahlbrand, W., Green, W., Grogg, J., Gould, O., & Winnett, D. A. (1993). *Science performance assessment handbook*. Edwardsville: Illinois Science Teachers Association.
Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*(5), 1098–1120.
Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.
Crawford, B. (2014). From inquiry to scientific practices in the science classroom. In N. G. Lederman, & S. K. Abell (Eds.), *Handbook of Research on Science Education* (Vol. 2, pp. 515–544). New York: Routledge.
DeanJr, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, *91*(3), 384–397.
Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (1999). *How people learn: Bridging research and practice*. Washington, DC: National Academies Press.
Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*, 268–291.
Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school. Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.
Eberbach, C., & Crowley, K. (2009). From every day to scientific observation: How children learn to observe the biologist's world. *Review of Educational Research*, *79*(1), 39–68.
Flavell, J. H. (1992). Cognitive development: Past, present, and future. *Developmental Psychology*, *28*(6), 998.
Georghiades, P. (2000). Beyond conceptual change learning in science education: Focusing on transfer, durability and metacognition. *Educational Research*, *42*(2), 119–139.
Gott, R., & Duggan, S. (1995). *Investigative work in the science curriculum. Developing science and technology education*. Bristol, PA: Open University Press.
Gott, R., & Murphy, P. (1987). *Assessing investigation at ages 13 and 15. Assessment of performance unit science report for teachers: 9*. London: Department of Education and Science.
Graft, M. van, & Kemmers, P. (2007). *Onderzoekend en Ontwerpend Leren bij Natuur en Techniek. Basisdocument over de didactiek voor onderzoekend en ontwerpend leren in het primair onderwijs*. Enschede: SLO.
Inspectorate of Education. (2015). *Wereldoriëntatie. De stand van zaken in het onderwijs*. Utrecht: Inspectie van het Onderwijs.
Janssen, J., Verhelst, N. D., Engelen, R. J. H., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8*. Arnhem: Cito.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.
Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology*, *14*(2), 217–238.
Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667.
Kneepkens, B., Van der Schoot, F., & Hemker, B. (2011). *Balans van het natuurkunde-en techniekonderwijs aan het einde van de basisschool, 4*. Arnhem: Cito.
Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*(4), 495–523.
Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, *60*(4), i–157.
Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, *9*(4), 285–327.
Lazonder, A. W., & Egberink, A. (2014). Children's acquisition and use of the control-of-variables strategy: Effects of explicit and implicit instructional guidance. *Instructional Science*, *42*(2), 291–304.
Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, *86*(3), 681–718.

Lazonder, A. W., & Kamp, E. (2012). Bit by bit or all at once? Splitting up the inquiry task to promote children's scientific reasoning. *Learning and Instruction*, *22*(6), 458–464.
Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92.
Martin, M. O., Mullis, I. V., Beaton, A. E., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1997). *Science achievement in the primary school years. IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill: TIMSS International Study Center.
Martinot, M. J., Kuhlemeier, H. B., Feenstra, H. J. M. (1988). Het meten van affectieve doelen: de validering en normering van de belevingsschaal voor wiskunde (BSW). *Tijdschrift Voor Onderwijsresearch*, 13(2), 65–76.
Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science*, *41*(3), 621–634.
McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, *15*(2), 153–191.
Millar, R. & Driver, R. (1987). Beyond processes. *Studies in Science Education*, *14*(1), 33–62.
Millar, R., & Lubben, F. (1996). Knowledge and action: Students' understanding of the procedures of scientific enquiry. *Research in Science Education in Europe*, 191–199.
National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
OECD. (2017). *PISA for development assessment and analytical framework*. Paris: OECD.
Osborne, J. (2015). *Practical work in science: Misunderstood and badly used?School Science Review*, 96(357), 16–24.
Osborne, J., & Dillon, J. (2008). *Science education in Europe: Critical reflections (Vol. 13)*. London: The Nuffield Foundation.
Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Into Practice*, *41*(4), 219–225.
Science Education Hub Radboud University. (2016). *The Question Machine*. Retrieved June 28, 2016 from http://www.ru.nl/wetenschapsknooppunt/english/materials/
Rasbash, J., Steele, F., Browne, W., Goldstein, H. (2009). *A user's guide to MLwiN*. London: Institute of Education.
Roth, K. J. (2014). Elementary science teaching. In N. G. Lederman, & S. K. Abell (Eds.), *Handbook of Research on Science Education* (Vol. 2, pp. 361–394). New York: Routledge.
Ruiz-Primo, M. A., Baxter, G. P., & Shavelson, R. J. (1993). On the stability of performance assessments. *Journal of Educational Measurement*, *30*(1), 41–53.
Sahin, A., & Kulm, G. (2008). Sixth grade mathematics teachers' intentions and use of probing, guiding, and factual questions. *Journal of Mathematics Teacher Education*, *11*(3), 221–241.
Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, *36* (1–2), 111–139.
Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, *7*(4), 351–371.
Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. *Applied Measurement in Education*, *4*(4), 347–362.
Shavelson, R. J., Solano-Flores, G., & Ruiz-Primo, M. A. (1998). Toward a science performance assessment technology. *Evaluation and Program Planning*, *21*(2), 171–184.
Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology*, *27*(1), 51–79.
Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement, and improvement*. Washington, DC: National Institute of Education (ED).
Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development*, *23*(4), 488–511.

Tanner, K. D. (2012). Promoting student metacognition. *CBE-Life Sciences Education*, *11*(2), 113–120.

Toth, E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction*, *18*(4), 423–459.

Van Merriënboer, J. J. G., Clark, R. E., de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, *50*(2), 39–64.

Van Merriënboer, J., Jelsma, O., & Paas, F. (1992). Training for reflective expertise: A four-component instructional design model for complex cognitive skills. *Educational Technology Research and Development*, *40*(2), 23–43.

Van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177.

Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3–14.

Wallace, B., Bernardelli, A., Molyneux, C., & Farrell, C. (2012). TASC: Thinking actively in a social context. A universal problem-solving process: A powerful tool to promote differentiated learning experiences. *Gifted Education International*, *28*(1), 58–83.

Weekers, A., Groenen, I., Kleintjes, F. G. M., & Feenstra, H. (2011). *Wetenschappelijke verantwoording papieren toetsen Begrijpend lezen voor groep 7 en 8*. Arnhem: Cito.

White, B. & Frederiksen, J. (2000). Metacognitive facilitation: An approach to making scientific inquiry accessible to all. In J. Minstrell & E. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 331–370). Washington, DC: American Association for the Advancement of Science.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3–118.

Wu, H., & Krajcik, J. S. (2006). Inscriptional practices in two inquiry-based classrooms: A case study of seventh graders' use of data tables and graphs. *Journal of Research in Science Teaching*, *43*(1), 63–95.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223.

Zohar, A., & Dori, Y. J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences*, *12*(2), 145–181.