
International Journal of Science Education

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

# Impact of an inquiry unit on grade 4 students' science learning

María Florencia Di Mauro & Melina Furman

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
VOL. 38, NO. 14, 2239–2258
http://dx.doi.org/10.1080/09500693.2016.1234085

# Impact of an inquiry unit on grade 4 students' science learning

María Florencia Di Mauro[a] and Melina Furman[b,c]

[a]Science Education Department, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina; [b]CONICET, Buenos Aires, Argentina; [c]School of Education, University of San Andrés, Victoria, Argentina

**ABSTRACT**

This paper concerns the identification of teaching strategies that enhance the development of 4th grade students' experimental design skills at a public primary school in Argentina. Students' performance in the design of relevant experiments was evaluated before and after an eight-week intervention compared to a control group, as well as the persistence of this learning after eight months. The study involved a quasi-experimental longitudinal study with pre-test/post-test/delayed post-test measures, complemented with semi-structured interviews with randomly selected students. Our findings showed improvement in the experimental design skills as well as its sustainability among students working with the inquiry-based sequence. After the intervention, students were able to establish valid comparisons, propose pertinent designs and identify variables that should remain constant. Contrarily, students in the control group showed no improvement and continued to solve the posed problems based on prior beliefs. In summary, this paper shows evidence that implementing inquiry-based units involving problems set in cross-domain everyday situations that combine independent student work with teacher guidance significantly improves the development of scientific skills in real classroom contexts.

**ARTICLE HISTORY**
Received 10 March 2016
Accepted 5 September 2016

**KEYWORDS**
Scientific thinking; guided inquiry; experimental design; primary science education

## Introduction

Scientific thinking involves a series of complex cognitive and metacognitive skills that must be developed and consolidated through sustained practice and exercise over time. Broadly defined, it includes the skills, knowledge and practices involved in inquiry, experimentation, evidence evaluation and inference, which contribute to scientific understanding (Zimmerman, 2007).

Specialists emphasise the role of the primary school years in laying the foundations of scientific thought (Harlen, 2000; Osborne, Simon, & Collins, 2003). As it has been emphasised, part of students' learning success or failure in science will depend on their early education (The Royal Society, 2010). Following this belief, many countries have included the learning of basic scientific thinking skills and practices in their primary school curriculum (NGSS, 2013; UK Department for Education, 2013).

**CONTACT** Melina Furman mfurman@udesa.edu.ar Escuela de Educación, Universidad de San Andrés, Vito Dumas 284, Victoria, Buenos Aires, Argentina

Among these skills and practices, the ability to plan, implement, analyse and evaluate experiments is an important educational goal of science education worldwide, given that experimentation is considered to be one of the most important methods for acquiring insight into scientific thinking and inquiry activities (Carey, Evans, Honda, Jay, & Unger, 1989). Overall, an experiment is defined as a procedure by which one or more hypotheses about a particular phenomenon can be contrasted by manipulating one or more variables (Zimmerman, 2007).

The experimental *design* can be understood as the logical structure of an experiment, as the part of the scientific process that involves planning and analysing it (Hidalgo, 2010). As Klahr and Nigam (2004) have argued, designing experiments is a key skill to develop in science learning because it is the core of a wide range of scientific topics and allows students to think in different ways to respond to questions on the natural world. The experimental design skill involves a set of other sub-skills, listed in Table 1.

However, research has shown that the development of the subset of skills involved in experimental design is not an easy goal to attain. For example, Chen and Klahr (1999) showed that only 35% of 4th grade students were capable of recognising errors in experimental designs to solve simple questions related to force and motion. Khun and Dean (2005) showed, in turn, that only 11% of the 6th grade students tested were able to identify the variables that account for natural hazards such as an earthquake when using a software that presented the effects different factors have on the level of hazard risk, despite having specific lessons to develop that skill.

Moreover, these difficulties prevail in a high percentage of adults (Khun, 2007). For instance, Brownell et al. (2014) argue that even introductory-level college students find designing well-controlled experiments challenging. They showed that between 62% and 80% of these students had some inaccurate or partially inaccurate conceptions about planning experiments, particularly related with sample size and repeating experiments. Even more, advanced students in science-related subjects showed some persistent inaccurate conceptions on experimental design.

Accordingly, we have recently shown alarming results in a study which tested 3900 pupils from 130 primary schools in disadvantaged contexts of Argentina. We found that only 9% of 6th graders were able to design a valid experiment to compare heat conductivity in different materials even after being given an example of a similar experiment that involved heat conductivity testing in other materials (Furman, 2012).

As it has been pointed out by Metz (2004), the development of scientific skills is a slow process that strongly depends on the pedagogy used. In this respect, various studies have shown that students obtain better results regarding the development of scientific thinking skills when teachers put into practice inquiry-based learning pedagogies (Cañal, 2007; Harlen, 2007; Minner, Levy, & Century, 2009).

**Table 1.** Skills involved in the experimental design skill.

| |
|---|
| a. Recognising a research question from a particular problem |
| b. Proposing a hypothesis as a possible answer to the research question |
| c. Comparing different experimental conditions |
| d. Determining which variable should be changed in order to answer the question |
| e. Defining which variables should remain constant to allow for a valid comparison |
| f. Defining the criteria used to measure, quantify and compare results |
| g. Inferring possible results, that is, predicting what can occur according to the available information |

However, further in-depth studies are needed to explore which pedagogical approaches are most effective to foster the experimental design skill in particular, compared to the different sub-skills already enumerated (see Table 1). As Roesch, Nerb, and Riess (2015) point out, research in the field of promoting components of experimental problem-solving ability needs to clarify the optimal method, school grade and use of appropriate domains and learning contexts with moderate extraneous conceptual load (i.e. the depth of scientific knowledge required to solve the problem) – particularly for average or lower performing learning groups.

For example, the degree of teacher guidance that students need to develop these skills is yet to be determined. More specifically, the most favourable combination between direct instruction and open experimentation to foster enduring learning is currently under discussion. There is no consensus either on the role that conceptual knowledge plays on the development of such skills. That is, if more general and less complex conceptual domains are required for students to learn experimental design, it is more appropriate to work with more specific and complex concepts.

Regarding the relationship between the degree of teacher guidance and the development of the experimental design skill, many studies focus on the promotion of specific associated sub-skills. Along this line, there is a vast history of studies that explores the development of the ability to control variables. For instance, Klahr and Li (2005) evaluated the impact of different types of instruction in this ability among primary school students from grades 3, 4 and 5. The authors named the three types of instruction tested as: direct instruction, Socratic instruction and discovery-based instruction. These types of instruction were defined by the level of teacher guidance provided for students to solve an activity in which they had to determine which of the given factors (type of surface, texture, weight, ramp's length and the type of ball) affected the speed with which a ball rolls down a ramp. In this case, researchers chose a problem in which conceptual knowledge did not play a key role, for their main aim was for students to learn to control variables. Their results showed that direct instruction, that is, teaching how to solve the problem explicitly, precisely and following a particular given structure, helped students solve the activities faster and with better results than the other types of instruction.

However, Dean and Kuhn (2007) found that the immediate positive effects of direct instruction vanished after three months of practice. Consequently, they criticised interventions that directly addressed specific sub-skills and argued that providing certain guidance instead fosters the active construction of more durable knowledge.

Despite the fact that few studies have yet explored the kind of educational interventions that promote the development of experimental skills comprehensively, it seems that a model that combines direct instruction and experimentation with the close guidance of teachers could produce more promising results in terms of student learning and its permanence.

With this concern, a study was conducted at a school in Turkey (Ergül et al., 2011) which showed that the implementation of inquiry-based activities involving scientific skills during a whole year significantly increased the performance of primary school students compared to children who received traditional education. They used a quasi-experimental design (control group and experimental group) involving 144 students from grades 4, 5 and 6 (10–12-year-olds). In the experimental group, students worked in small teams on experimental activities to solve open problems, with teacher interventions that guided group discussions and clarified doubts. Activities were framed in different conceptual domains from physics, chemistry and biology. After the intervention, students from both groups were evaluated through standardised tests adapted to the study population: Basic Science Process Skill Test (BSPST) and Integrated Science Process Skill Test (ISPST). Results indicated that the experimental group systematically obtained better results than the control group on the ability to plan experiments (in general) and to control variables (in particular), among other skills tested.

More recently, Roesch et al. (2015) have studied the impact of inquiry-based teaching to promote the experimental problem-solving ability among 6th grade German students, with successful results. They found that working with students in problem-solving within the conceptual domain of system ecology, which combined phases of direct instruction and open experimentation, fostered specific components of experimental problem-solving ability (generating epistemic questions, planning two-factorial experiments and identifying correct experimental controls). However, the observed effects were smaller than expected, a result they attributed to the high level of complexity of the conceptual domain with which they worked.

This result is related to the other focus of discussion: the role of conceptual domains (i.e. disciplinary content knowledge) when teaching scientific thinking skills. In this regard, there is a line of research that explores the development of experimental skills in a specific conceptual domain to foster both scientific thinking skills and conceptual knowledge (Roesch et al., 2015), whereas others focus on the inquiry process and abilities, leaving science content to play a subordinate role.

This last group of studies highlight the value of working in cross-domains (i.e. general problems relating to different scientific disciplines) for the development of skills and their transferability. Various authors have reported on the effectiveness of their interventions to enhance cross-domain experimental skills in grades 6 or 7, or even lower age groups (e.g. Carey et al., 1989; Ergül et al., 2011; Klahr & Nigam, 2004). Moreover, they all acknowledge that it is necessary to load down cognitive/conceptual complexity of the context to focus on the development of complex scientific thinking skills such as experimental design.

Following up on this debate, there is still much to be explored on effective ways to teach this central skill to children, how students can develop their ability to design experiments on new topics and how permanent this learning can be over time. This study focuses on answering precisely those questions.

We explored the impact of an inquiry-based activity to develop the ability of experimental design in holistic terms, by working in various conceptual domains with low conceptual demand, open experimentation and with close guidance of teachers through key questions and interventions. We analysed the effect of a level III (Herron, 1971) guided inquiry-based eight weeks unit on the development of 4th grade (9- and 10-year-old) students' experimental design skills at a public urban primary school in Argentina.

We looked at student performance on the design of valid experiments related to cross-domain everyday situations before and after the intervention, compared to a control group. We also examined the persistence of this learning after eight months (which included the two-month summer holidays).

## Research questions

To analyse the impact of the implementation of a guided inquiry-based sequence in the development of the experimental design skill in a group of 4th grade students, we addressed the following research questions:

- To what extent did the implementation of an eight-week guided inquiry-based unit focused on the planning of cross-domain and low conceptual load experiments improve the experimental design skills of 4th grade students?
- How well did the students' improved ability in experimental design endure eight months after the implementation?

## Programme description

This study emerged as the result of a partnership programme between a state university in Argentina and a group of urban public primary schools with students from low-income populations. Schools could voluntarily choose to join the programme, which mainly consisted of supporting teachers in the design and implementation of a series of inquiry-based units covering topics of the science curriculum for each grade.

Within the inquiry-based approach, the curriculum units aimed to develop student learning of scientific skills through the resolution of open-ended problems, which placed students in an active role as knowledge generators and teachers as guides to enhance and enrich this process (Harlen, 2000). Within this same framework, units were designed as level III guided inquiry, which implies that students had to develop their own experimental design to respond to a research question and other related structured guiding questions posed by the teacher (Herron, 1971).

In particular, we followed the Scientific Discovery as Dual Search (SDDS) model proposed by Klahr (2000). The SDDS model emphasises three major problem-solving phases: in the first, the problem solver searches the hypothesis space; afterwards, he searches the experiment space; finally, evaluates the evidence. These processes are interconnected and consist of different sub-processes and cognitive activities.

In this study, we analysed the impact of a unit specially designed to enhance the experimental design skill in 4th grade students, which was implemented at one of the participating schools. Students were expected to answer research questions that arose from given everyday problematic situations in different domains, all with low conceptual load, by planning the way to look for pertinent evidence. By low conceptual load problems we refer, in this context, to problems which only require students to use their everyday knowledge (rather than scientific understandings). As suggested by Harlen (2010), each activity engaged children in solving everyday problems with simple experiments in which the dependent and independent variables could be clearly distinguished, in order to foster their understanding of how to establish valid comparisons between tests.

The entire intervention took place over eight 2-h periods taught over eight weeks. It consisted of three different activities: (1) *Should we buy big vegetables?* (2) *What do yeasts feed on?* and (3) *What environment do woodlice prefer?*

**WHAT DO YEASTS FEED ON?**

**1) Read the following problem and answer the questions.**

*Boris read in a Science book that there are some tiny organisms called yeasts. Clarita told Boris that when yeasts feed, they liberate gas which forms little bubbles and that the better they feed, more bubbles they produce. Boris assured Clarita that yeasts feed better on sweetener than on sugar, so she suggests that they should test it through a simple experiment. Let´s help them out!*


a) What is the question the children want to answer?

**2) In groups, discuss and perform the following activities:**

a) What should Boris and Clarita buy to carry out the experiment?
b) What are the children going to measure? How can they measure it?
c) In order to put this experiment into practice and obtain valid results to answer their research questions, the children will have to keep some things constant. What are those things?
d) What is the only thing we need to change between one group of yeasts and the other?

**3) In groups, discuss and draw a sketch of the experiment you are going to perform to help Boris and Clarita find out which food is best for the yeasts.** Think of every detail of the experiment (materials, measuring techniques, etc.)

**4) Draw and write a description of all the possible results Clarita and Boris may obtain.**

**5) Register in a table the results obtained.**

**6) What conclusions can you drawn from the results obtained?**

**Figure 1.** Example of one of the proposed activities: What do yeasts feed on?

Students worked in small groups (four or five students each) to solve the activities. Figure 1 shows an example of one of the activities, including the guiding questions which were common to all the activities in the unit aimed to help students in designing their experiment.

Once every group had elaborated their experimental design, each group presented it to the rest of the class, which served as the *science community* to evaluate the quality of the experimental designs and to discuss how to improve them, with the teacher's moderation.

Then, each group carried out the experiment according to their own design. The teacher helped students with the observation and measurement-taking. Finally, the results were recorded in crosstabs and shared with different groups. The whole class discussed the conclusions obtained and the answer to the research question. The unit is summarised in Table 2.

## Methodology

In order to examine the effect of the implementation of a cross-domain guided inquiry unit on students' ability to design experiments, we conducted a quasi-experimental and

**Table 2.** Inquiry-based unit summary.

| Meetings | Activity | Common steps for all activities |
|---|---|---|
| 1 and 2 | Should we buy big vegetables? | 1. Children are presented with a problematic situation and asked to recognise the research question involved in the problem |
| 3, 4 and 5 | What do yeasts feed on? | 2. In groups, children work on the elaboration of an experimental design to answer the research question. Teachers guide each group in this process |
| 6, 7 and 8 | What environment do woodlice prefer? | 3. Each group revises the experimental design of other groups and proposes improvements |
| | | 4. Students conduct the experiment and collect data |
| | | 5. Each group presents the results obtained and draws conclusions |
| | | 6. The whole class reflects on the conclusions drawn for the original research question and on the processes that was undertaken in order to answer it |

longitudinal study with pre-test/post-test/delayed post-test measures. Study participants involved 30 4th grade students in each class from an urban public primary school with a low-income student population in Argentina. The experimental group worked for eight weeks with the unit, whereas the other (control group) worked with their regular science curriculum involving the ordinary pedagogical approach. In the control group, children worked mostly with texts and questionnaires, and only occasionally their teacher performed experimental activities, always in a demonstrative manner. Both groups were comparable in terms of student socioeconomic background, gender ratio and prior performance in science.

Students' experimental design skill level was assessed in both groups of students before and after the eight-week intervention. Finally, we conducted a follow-up test (delayed post-test) with both groups eight months after the end of the intervention (which included a two-month summer recess) to evaluate to what extent learning was sustained over time.

In addition, to deepen our understanding of the students' answers on the pre- and post-tests, we conducted semi-structured interviews with five randomly selected students from the experimental group.

The methodological design is summarised in Figure 2.

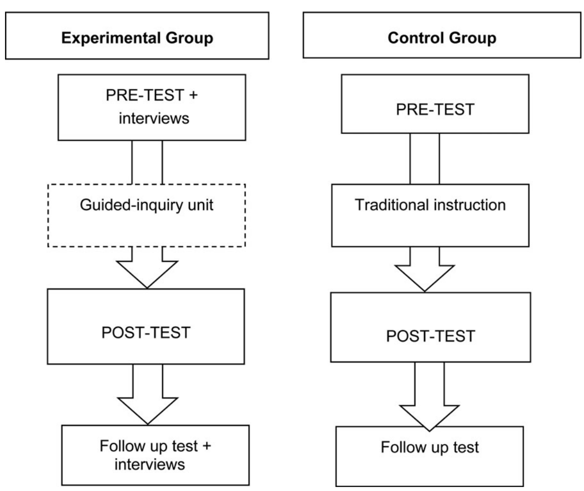


**Figure 2.** Methodological design.

## *Assessment instruments*

In order to assess the students' experimental design skills, we used problematic situations described as simple, everyday stories. We designed three tests with similar structure, format and level of difficulty but with different content and application domains for the pre-, post- and delayed post-test.

Problems were designed by taking into account the learning goals from the curricular guidelines of the province of Buenos Aires (Dirección General de Cultura y Educación de la Provincia de Buenos Aires [DGCE-PBA], 2008) and previously validated assessment instruments such as the ones described in Furman (2012).

A pilot test was conducted prior to the intervention with students of the same age and from a similar educational context. Teachers and researchers jointly revised the activities' suitability to test the students' scientific skills and made small adjustments.

The tests consisted of three activities with open-ended questions that assessed different scientific skills. For the purpose of this study, we focused on one of the activities related to experimental design. We chose open-ended questions in order to obtain qualitative evidence on the way students structured and justified their answers.

Figure 3 shows an example of the pre-test instrument.

## *Interviews*

After each assessment (pre- and post-tests), we conducted semi-structured, individual interviews with five students in order to enrich our analysis. These were randomly selected from the experimental group as a representative sample. We focused on students within the experimental group since we were interested in gathering qualitative evidence about their answers and how they linked them to their work within the curriculum unit. Each interview lasted about 20 minutes and involved going over with the students their answers to the tests, asking them to explain what they had drawn and written. We specifically inquired on the meaning students gave to the activities and why they solved them in a particular way. In the post-test interview, we also asked the children to reflect on how their work within the unit had helped them answer the problems on the test.

Boris wants to find out if a fabric dye dissolves best in hot or cold water. He suggests to his friend Claire to test it with an experiment.

1. What experiment can Boris do to find this out?
   a. Draw the experiment in the box below and label the materials used (remember to describe every detail).
   b. Explain the experiment you´ve drawn.

**Figure 3.** An example of a pre-test instrument question.

### Data analysis

In order to assess the impact of the unit on students' experimental design skills, we established four different performance levels: level 1: absent; level 2: incipient; level 3: developing; level 4: advanced (see Table 3). The definition of such levels was determined from a preliminary analysis of the students' answers in the pre-test and the identification of key aspects of the experimental design skill as proposed by Zimmerman (2007).

In order to define performance levels, we first established an 'optimal' level (level 4) according to prior research in 4th graders' experimental design skills (Harlen, 2007; Zimmerman, 2007) and to local learning standards. The advanced level (level 4) was our reference level of experimental design, that is, what students that age are expected to achieve according to the national curriculum goals (Federal Board of Education [FBE], 2004). As described in Table 3, students at level 4 must show the capacity to establish a valid comparison between at least two different experimental groups, including proposing a way to measure the effects of the variables involved, and the consideration of at least one condition that should remain constant for the experiment design to avoid possible confounders.

All the remaining levels (levels 1–3) established a progression towards the optimal level (level 4) and were defined by classifying children's responses in the pre-test. We also took into account prior research on children's learning process for this skill in order to help define these levels of progression (Zimmerman, 2007). In doing so, we grouped children's pre-test responses in an inductive manner, looking for increasing levels of development of the experimental design skill, as we described in a previous article (Di Mauro, Furman, & Bravo, 2015).

Two independent observers classified the students' answers based on a common assessment rubric. Differences among observers, if any, were discussed in order to reach an agreement. Statistical analysis was performed using a chi-square test, comparing the distribution of levels of experimental design skills between the control and experimental groups.

The above-described performance levels regarding the experimental design skill were also used as themes for a thematic analysis of students' interviews. We looked for evidence of the four established levels of student performance in children's responses. We also looked for evidence of children's transfer of the experimental design skill to other contexts, especially to everyday life problems.

**Table 3.** Levels of experimental design skills.

| Levels | Description |
|---|---|
| Level 1: absent | There is no correct comparison between two or more groups nor is there a coherent plan to solve the problem |
| Level 2: incipient | Proposes a correct comparison between groups, but does not include other aspects of the experiment |
| Level 3: developing | Proposes a correct comparison between groups and includes only one of the following elements: a valid measuring strategy OR the identification of a condition that must remain constant |
| Level 4: advanced | Proposes a correct comparison between groups, a valid measuring strategy AND identifies at least one condition that must remain constant |

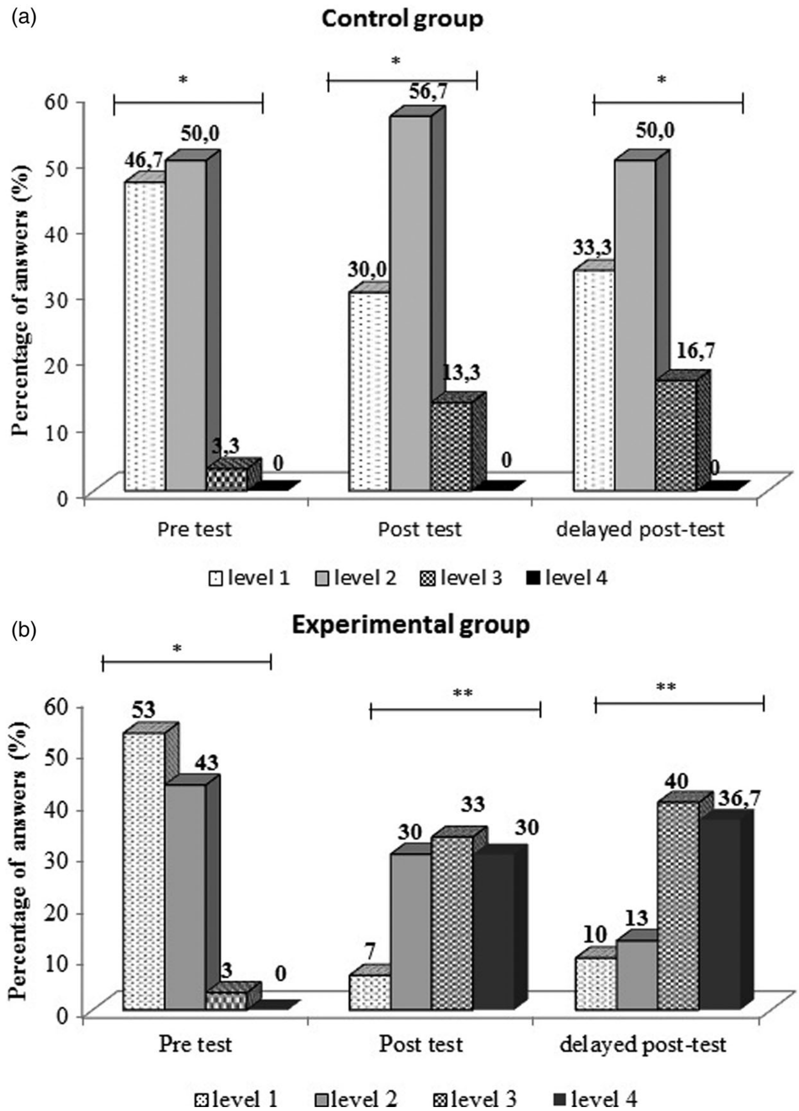


**Figure 4.** Experimental and control group pre-, post- and delayed post-tests results. Percentage of student answers classified in four levels according to the level of development of the experimental design skill. Level 1: absent. Level 2: incipient. Level 3: developing. Level 4: advanced. (a) Results obtained by the control group ($n = 30$). They do not show statistical differences over time. (b) Results obtained by the experimental group ($n = 30$). They showed statistical differences between pre-test, when compared to post-test and delayed post-test results. No statistical differences were found between the post and delayed post-tests. The initial results obtained by both groups in the pre-test did not show significant differences ($\chi^2 = 0.276$, $p = .871$).

## Results

### *Pre- and post-test performance*

At the pre-test, there were no significant differences in the level of performance of students in the control and experimental groups ($\chi^2 = 0.276$, $p = .871$), as shown in Figure 4(a,b) (pre-test), supporting the comparability of the control and experimental groups. Consistent with what other researchers and what national and international standardised tests have shown, the levels of student ability in this domain were very low. In both groups, more than 95% of the students performed at level 1 (absent) and level 2 (incipient) for the experimental design skill.

However, the post-test results showed important differences between the performance of both groups. As shown in Figure 4(b), after the eight-week unit implementation, the experimental group showed significant levels of improvement at the end of the programme ($\chi^2 = 27.980$, $p < .001$), with no differences for the control group (Figure 4(a)).

A closer look at the data reveals that in the experimental group 66.3% of the students reached levels 3 (in progress) or 4 (advanced). Moreover, only those students who participated in the inquiry-based instruction were able to reach an advanced level in their ability to design experiments (level 4).

Figure 5 shows representative examples of student responses for each assessment category to illustrate in more concrete terms what the level of improvement achieved by the students in the experimental group implied. The first is the case of a girl that at the beginning (pre-test) did not recognise that she had to compare two situations in order to determine which one was better (in this case, compare hot or cold water in order to find out which of them dissolved a certain colourant the best – see Figure 3). This answer was assessed as the lowest possible level (level 1 – absent) in the ability to plan experimental designs (Figure 5(a)). After the instruction, the same student was able to establish a comparison between the two options and list the materials needed (including measuring instruments, such as the watch) to draw valid conclusions about a question related to the use of two different materials to make a soup spoon that would not conduct heat very fast. In the post-test, therefore, she was able to reach performance level 3 (developing) (Figure 5(b)).

The second case was the one of a boy, who at the beginning reached a performance level 2 (incipient). This means that he could propose a simple comparison between groups (Figure 6(a)), although he could not give any further details of how he would conduct the experiment. After the intervention, this student was able to describe a more complete experimental design, that is, to compare the two situations, establish a measurement method for the dependent variable and consider which variables should remain constant for the experiment to be valid, showing a performance of level 4 (advanced) (Figure 6(b)).

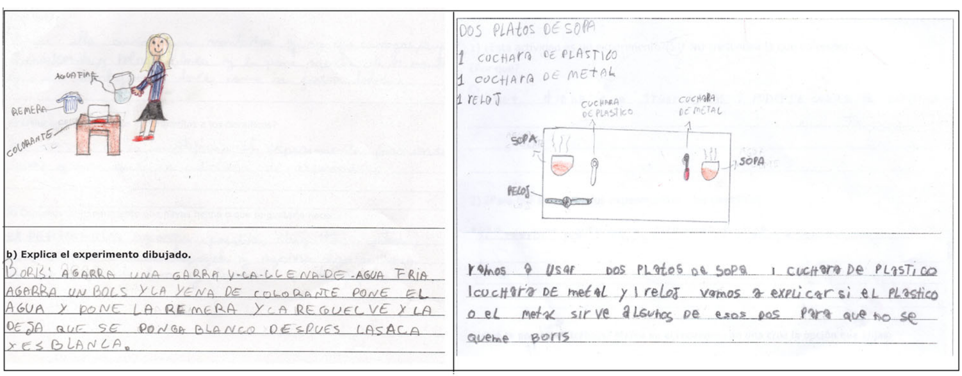


Figure 5. Example of a student's answers before and after the intervention.

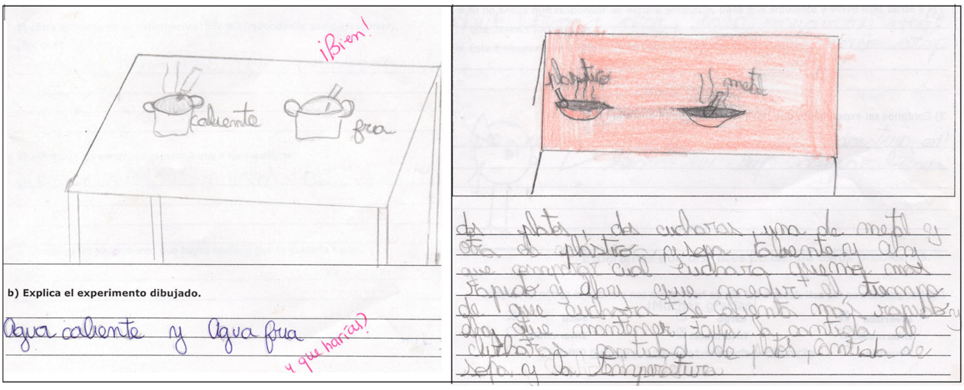


**Figure 6.** Example of another student's answers before and after the intervention.

Students' testimonies coincided with the results obtained in the tests and offered us further evidence of the improved ability to think and plan experiments to answer research questions among those who participated in the programme. For example, the following transcript shows the answers of a student (student A) who initially could neither state a comparison between the two situations nor consider the possibility to design an experiment to answer simple research questions (level 1: absent). Although the interviewer rephrased the question several times, the student gave the expected result without proposing a way to find that answer, a very frequent situation among students at the beginning.

Interviewer: Let's imagine the following situation: you are at home and find a white T-shirt that you want to dye red. You want to try if cold water or hot water dyes the T-shirt better …
Student A: Cold water.
Interviewer: Why?
Student A: Because I think so.
Interviewer: But, what would you do to find that out?
Student A: I would put the T-shirt with cold water and dye.

After the intervention, this same student showed that he was capable of stating a comparison between the two situations and to establish the procedure to collect results to answer the research question beyond his beliefs. As shown below, he also recognised that some factors have to remain constant for the comparison to be valid.

Interviewer: Now let's attend to this problem, what experiment would you do to find out which material, plastic or metal, is better to have soup without burning your hand?
Student A: Put hot water and two spoons, one metal and one plastic, and wait.
Interviewer: Wait for what?
Student A: 70 minutes.
Interviewer: What for?
Student A: To know which of the two spoons get hotter.
Interviewer: And how would you know?
Student A: I don't know.
Interviewer: What would you do?
Student A: Touch them after the 70 minutes.
Interviewer: What for?
Student A: To know which spoon is hotter.

In the following case, the student (student B) initially recognised that he had to establish a comparison between two situations (level 2: incipient), but despite the interviewer asking him repeatedly, he could not state any other characteristic of the experiment. After reading the pre-test problem in which they had to find out if cold or hot water is better for dissolving fabric dye (see Figure 3), the student stated the following:

| | |
|---|---|
| Student B: | Boris gets two boilers and in one he puts hot water with fabric dye and in the other cold water with fabric dye, to see in which one it dries faster. |
| Interviewer: | Dries or what? |
| Student B: | Which one dissolves better. |
| Interviewer: | And how will you know which one dissolves better? |
| Student B: | By doing the experiment. |
| Interviewer: | And how would you do it? |
| Student B: | I get two boilers, one with hot water and one with cold water and fabric dye and wait for the results. |
| Interviewer: | And how will you realize which one is best? |
| Student B: | From the results of which one is best. |
| Interviewer: | Which results can you obtain? |
| Student B: | I don't know. |

In the interview after the intervention, this same student showed a great improvement in planning the experiment to answer a given research question. Besides being able to design a more complete experiment (level 4), he was also able to state how to measure the results and repeatedly pointed out that some variables must remain constant to allow comparisons, both aspects that he did not consider before.

| | |
|---|---|
| Interviewer: | Let's see … what would you do to find out about the best spoon for making soup? |
| Student B: | An experiment. |
| Interviewer: | How? |
| Student B: | I put two dishes on the table, put the same amount of soup in each plate and put the two spoons and wait to see how much time it takes for each spoon to become so hot that I cannot handle it. |
| Interviewer: | How? |
| Student B: | I measure when the spoons become too hot to be handled. |
| Interviewer: | How? |
| Student B: | I … with my hands. |
| Interviewer: | Ok, good … Is there something else we should consider? |
| Student B: | The possible results. |
| Interviewer: | What else? Before thinking on the possible results, you said in your written response that we should put the same amount of soup in two plates, and then? |
| Student B: | I put the same amount of soup and two spoons of the same length, because if not one becomes too hot to handle because it is smaller and the other takes longer because it is bigger and it does not work that way. |

Another interesting thing to point out is that, besides stating that some things should be the same to be able to compare them, when the interviewer asked him about the advantages of putting both spoons in the same dish of hot soup, the student was able to explain the importance of controlling variables in an experiment, as shown in the transcript below:

| | |
|---|---|
| Interviewer: | If I were to put both spoons on the same plate, will I be able to perform the experiment? |

| | |
|---|---|
| Student B: | I don't think so … |
| Interviewer: | Yes or no? |
| Student B: | Yes, because I can measure the same things in the same dish. |
| Interviewer: | Why? |
| Student B: | Because I wouldn't need to measure. |
| Interviewer: | How is that? |
| Student B: | Because it would give the same results as long as the spoons are different. |
| Interviewer: | Different? |
| Student B: | No, of the same length and width, but as long as one spoon is plastic and the other metal it can be done. |
| Interviewer: | Let's see … |
| Student B: | I know, because he will obtain the same results and there will be no differences between the spoons. |
| Interviewer: | I don't understand … would you choose to use one or two dishes? |
| Student B: | One because I will have the same results as with two and I can do the experiment just the same. |
| Interviewer: | So then what is the advantage of using one plate? Or is it the same of using two? |
| Student B: | No, it is not the same because if I put the spoons in two different plates it can occur that one is hotter than the other or that I put one before and one after and the results are because I put one before and one later. So, I can be more certain of the results I will obtain. |

During the interview, student B also revealed that he had managed to design different experiments at home to solve research questions of his own. For example, he explained how he came up with a question on bubbles to perform an experiment on which liquid was best in order to produce bubbles that lasted the longest: 'I wanted to plan the experiment on bubbles because one day when I was taking a bath I saw that bubbles were produced with the soap and I also knew that they could be done with detergent'.

When providing details of his experimental designs, this student demonstrated a great capacity to transfer the scientific skills developed in class and apply similar procedures to answer the research questions he came up with. For example, in the experiment on bubbles he identified the need to use two containers to measure the same amount of the liquids he would use (soap and detergent) for the comparison to be valid, as opposed to what happened in the previous problem on heat conductivity: 'I thought I needed two cups … in this case I needed two cups because they are different things, two cups and a feeding bottle'.

Secondly, he considered every detail in the experimental design, for example, the use of a feeding bottle to take measurements and other materials.

| | |
|---|---|
| Interviewer: | You need two cups and one … ? |
| Student B: | … feeding bottle, detergent and soap. |
| Interviewer: | Good. |
| Student B: | I get the feeding bottle so that they have the same quantity in each cup and the experiment is done correctly. |
| Interviewer: | And you would use a feeding bottle for that? |
| Studentt B: | Yes, to measure, because feeding bottles always have centimetres to see the amount of milk. So I put a little bit of water to the soap and I crush it and measure the amount of soap that was left. In the other one, I put detergent and also measure it. I put the same amount of water in each cup and put detergent in one and soap in the other. And I also need two wires. |

| | |
|---|---|
| Interviewer: | For what? |
| Student B: | To blow the bubbles! |
| Interviewer: | Oh, right! How nice! |
| Student B: | I stand up on top of a chair, but have to stand up and measure in both because it can last less because of the height from which I blew the bubble and explode before the other. |
| Interviewer: | I don't understand, how is that? For what do you need the chair? |
| Student B: | I blow both bubbles on top of a chair because if I do one on top of the chair and one standing on the floor, the one I did on top of the chair can explode before. |
| Interviewer: | But why? What do you want to see? |
| Student B: | If the soap bubble lasts longer than the detergent one. |
| Interviewer: | Okay, fantastic! And what would you measure? |
| Student B: | The same measure of … |
| Interviewer: | To get the result, what would you measure? |
| Student B: | Which bubble explodes first. |
| Interviewer: | What do you think working with this unit, and with experiments in general, has helped you do better? |
| Student B: | I feel I can answer my own questions, I can learn more things and I have fun (…) Before, I asked myself questions but could not answer them. |

As this transcript shows, not only the student was able to plan the complete design and predict the possible expected results for all of his experiments, but he also expressed the relevance of conducting experiments to satisfy his own curiosity.

We summarise in Table 4 the thematic analysis of student responses at the interviews.

## *Delayed post-tests*

With reference to our second research question, we were interested in examining if the experimental design skills developed by the students were sustained several months after the intervention was completed. In order to do this, we conducted a delayed post-test with both the experimental and the control groups eight months later, including the summer recess.

After eight months, we saw no significant differences in the control group in their level of skill in the experimental design when comparing the post-test with the delayed post-test ($\chi^2 = 0.289$, $p = .866$) (Figure 4(a)). This result is important, since it means that even when students turn to 5th grade (that is, after the eight months period), performance does not significantly improve if they do not specifically work towards learning the experimental design skills, and supports the idea that the development of this ability does not evolve naturally without specific teaching.

Moreover, in the case of the experimental group, our data showed no significant differences between the post-test and delayed post-test results ($\chi^2 = 2.505$, $p = .474$) (Figure 4 (b)). This result is especially important, since it means that the students' levels of experimental design skills were sustained after eight months, and points towards the stability of this learning. Here, we provide evidence of the impact of an eight-week intervention based on a problem-based, guided inquiry approach with low conceptual load on building enduring scientific skills. As research has shown, this is a goal that has often proved difficult to achieve. As Dean and Kuhn (2007) have described, although explicit and direct

**Table 4.** Summary of the thematic analysis of student interviews.

| Theme | Pre-test interview<br>Example from student responses | Post-test interview<br>Example from student responses |
|---|---|---|
| Level 1 | *I would put the T-shirt with cold water and dye* | No examples of this performance level in the interviews |
| Level 2 | *Boris gets two boilers and in one he puts hot water with fabric dye and in the other cold water with fabric dye, to see in which one it […] dissolves the best* | No examples of this performance level in the interviews |
| Level 3 | *I take the dye, put it in a bowl with hot water, then put some more in another bowl with cold water and wait. Then I see how long it took for the T-shirt to become all red* | *Put hot water and two spoons, one metal and one plastic, and wait […] for 70 minutes […] to know which of the two spoons get hotter […] Touch them after the 70 minutes […] to know which spoon is hotter* |
| Level 4 | No examples of this performance level in the interviews | *I put two dishes on the table, put the same amount of soup in each plate and put the two spoons (of the same length) and wait to see how much time it takes for each spoon to become so hot that I cannot handle it* |
| Transfer of the experimental design skill to everyday situations | No examples of this performance level in the interviews | *One day when I was taking a bath I saw that bubbles were produced with the soap and I also knew that they could be done with detergent […] I put the same amount of water in each cup and put detergent in one and soap in the other. [I wanted to see] if the soap bubble lasts longer than the detergent one […] Which bubble explodes first* |

instruction methods achieve immediate positive results on children's learning, but these achievements are fragile, since they fade out after only three months.

## Conclusions and discussion

As we have discussed, in many parts of the world, including Argentina, in the context of this study the key science learning goal of primary school students currently includes the development of scientific thinking skills, which, as research shows, has not been an easy goal to attain. Consequently, there is an emergent line of research that responds to the need of compiling evidence on teaching strategies that foster the development of such skills. They aim to elucidate, among other issues, the on-going debate on how much guidance should teachers provide to students, what kind of activities students should be engaged in and the amount of time needed for students to learn (Toth, Klahr, & Chen, 2000; Zimmerman, 2007).

Our study contributes important evidence to this debate. In general terms, we have shown that working for eight weeks with a guided inquiry-based unit that involved a series of everyday problematic situations anchored in different content domains produced a significant improvement in the development of student experimental design skills among 4th graders. In particular, we believe that working with low conceptual load problems set in cross-domain everyday situations, which combined student independent work through a series of guiding questions posed by the teacher with moments of more explicit teacher guidance, was a key element to foster this development.

In line with the findings of international evaluation programmes, the results of the diagnostic tests showed that student performance on experimental design was very poor.

Initially, many students performed in level 1 (absent) and tended to give the correct answer to the problem based on their beliefs rather than proposing a method to find the solution. This was confirmed during the pre-test interviews, where despite students being asked repeatedly and in different ways how they could they find out the answers, they responded to the question without proposing a pertinent plan to follow. However, results showed that after the intervention, students who participated in the experimental group were able to identify two situations to compare, propose an experiment to test them and even determine which variables should remain constant (levels 3 and 4). Moreover, they could anticipate different possible results to answer the problem and thus provide an understanding of the concept.

This paper shows evidence of the possibility to implement strategies that foster enduring scientific thinking skills in children in real classroom contexts. We have also shown the positive impact of using a guided inquiry approach, which integrates student independent group work guided by a series of written questions offered by the teacher, with a more direct orientation of the teacher and whole class discussions.

Regarding transfer, it has been suggested that inquiry skills are firmly entrenched in the domain in which they are learned (Van Joolingen, de Jong, & Dimitrakopoulout, 2007). However, our study points out the value of having students work with low conceptual load problems, which do not require a great deal of conceptual understanding, as a way to focus their learning on scientific skills which, in turn, can be transferred to other conceptual domains, as we saw both in the students' post-test responses (which presented problems in different conceptual domains) and in the case of student B, who was able to plan his own experiments for questions he was intrigued about.

Furthermore, one of the main challenges that Khun and Dean (2005) attribute to science teaching is for students not only to acquire scientific skills but to be able to use them in everyday life situations to understand their environment. In this sense, we found encouraging evidence in the case of student B, who proved to be able and motivated to use the skills and tools learned beyond the classroom to satisfy his curiosity and explore phenomena around him. Although this was not a specific objective of the intervention, it is interesting to point it out particularly in a context marked by certain disinterest in scientific careers.

Our study adds to the evidence that shows that long-term and repeated practice of inquiry can lead to successful acquisition and transfer to novel tasks (Dean & Kuhn, 2007). What is especially relevant, however, is that student improvement occurred after only eight weeks of lessons.

Along these lines, the fact that students could sustain their high levels of performance is extremely relevant, for it implies that their learning was enduring. As we have stated before, many authors have pointed out the importance of developing enduring learning, as well as the difficulties in sustaining learning presents in the long-term (Wiggins & McTighe, 2005). The fact that student learning lasted for at least eight months, including the summer recess, is especially meaningful in this context.

In all, this study adds to the conversation about the most effective ways to develop scientific thinking skills in young children. Our findings suggest the need to develop curriculum materials and instruction which specifically focus on the developing of scientific skills as a requisite to develop these skills in children, and provide evidence that the development of the ability to design simple experiments can be developed as early as in 4th grade. They also suggest the value of working with students in problems that require low-level of conceptual knowledge in order to focus on the development of experimental skills, with the ultimate aim of fostering the transfer of those skills to different conceptual domains. Finally, our data point out to the importance of scaffolding student planning of experimental designs by providing guiding questions and teacher feedback regarding basic aspects of the process, as well as allowing for opportunities for open discussion and experimentation.

We believe this study also calls for a reflection on the current teaching practices in science at the elementary school level. As we have mentioned, the ability to design valid experiments to answer research questions is a key attribute of what has been defined as scientific thinking, and is often part of the learning standards expected by many countries of the world. However, both international assessments and research studies have shown that, in many regions of the world, most primary school children finish school without having developed even the foundations of this ability.

Yet, this study provides evidence that advancing elementary school children's learning on the experimental design skill is an attainable goal, even in a relative short amount of time. What is more important, perhaps, is that this learning goal was attainable within the context of a public low-income urban school with large-classroom sizes (around 30 students), using low-cost materials and working with children holding heterogeneous levels of performance. Therefore, this study brings 'a proof of possibility' (Cochran-Smith, 2004) of what students can learn provided they are offered suitable learning opportunities.

Regarding such learning opportunities, we have shown the value of working for several weeks with different kinds of everyday life problems, as well as the importance of having students solve these problems in groups following a structured guide which scaffolded their problem-solving process.

A question remains open, however, on the most effective teacher education strategies to support primary school teachers in offering children these learning opportunities. Given the central role teachers have in guiding and scaffolding student problem-solving and experimental design processes, finding successful ways of helping teachers work with students in problems which require the ability to design a valid experiment becomes both urgent and imperative.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

Brownell, S. E., Wenderoth, M. P., Theobald, R., Okoroafor, N., Koval, M., Freeman, S., & Crowe, A. J. (2014). How students think about experimental design: Novel conceptions revealed by in-class activities. *Bioscience*, *64*(2), 125–137.

Cañal, P. (2007). La investigación escolar hoy. *Alambique*, *52*, 9–19.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). 'An experiment is when you try it and see if it works': A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, *11*(5), 514–529.

Chen, Z., & Klahr, D. (1999). All other things being equal: Children's acquisition of the control of variables strategy. *Child Development*, *70*, 1098–1120.

Cochran-Smith, M. (2004). *Walking the road: Race, diversity and social justice in teacher education*. New York: Teachers College Press.

Dean Jr., D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, *91*(3), 384–397.

Di Mauro, M. F., Furman, M., & Bravo, B. (2015). Las habilidades científicas en la escuela primaria: un estudio del nivel de desempeño en niños de 4to año. *Revista Electrónica de Investigación en Educación en Ciencias*, *10*(2), 1–10.

Dirección General de Cultura y Educación de la Provincia de Buenos Aires. (2008). *Diseño Curricular para la Educación Primaria. Segundo Ciclo*. [Curricular guidelines for the elementary level]. La Plata: Author.

Ergül, R., Şımşeklı, Y., Çaliş, S., Özdılek, Z., Göçmençelebı, Ş., & Şanli, M. (2011). The effects of inquiry-based science teaching on elementary school students' science process skills and science attitudes. *Bulgarian Journal of Science & Education Policy*, *5*(1), 48–68.

Federal Board of Education [Consejo Federal de Cultura y Educación]. (2004). *Núcleos de Aprendizaje Prioritarios*. Buenos Aires: Ministerio de Educación, Ciencia y Tecnología.

Furman, M. (2012). ¿Qué ciencia estamos enseñando en escuelas de contextos de pobreza? [Which science are we teaching in schools from vulnerable contexts?] *Praxis & Saber*. *Revista de Investigación Y Pedagogia*, *3*(5), 15–52.

Harlen, W. (2000). *The teaching of science in primary schools*. London: David Fulton Publichers.

Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation*, *33*(1), 15–28.

Harlen, W. (2010). *Principles and big ideas of science education*. Hatfield: ASE.

Herron, M. D. (1971). The nature of scientific inquiry. *The School Review*, *79*(2), 171–212.

Hidalgo, V. H. (2010). Diseños experimentales: breve semblanza de su importancia en las ciencias naturales. *Ciencia y Mar*, *XIV*(40), 45–52.

Khun, D. (2007). Jumping to conclusions. *Scientific American Mind*, *18*(1), 44–51.

Khun, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*(11), 866–870.

Klahr, D. (2000). *Exploring science. The cognition and development of discovery processes*. Cambridge: MIT Press.

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*, 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*, 661–667.

Metz, K. E. (2004). Children's understanding of scientifc inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction*, *22*, 219–290.

Minner, D. D., Levy, A. J., & Century, J. (2009). Inquiry-based science instruction-what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, *47*(4), 474–496.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Osborne, J. F., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079.

Roesch, F., Nerb, J., & Riess, W. (2015). Promoting experimental problem-solving ability in sixth-grade students through problem-oriented teaching of ecology: Findings of an intervention study in a complex domain. *International Journal of Science Education*, *37*(4), 1–22.

The Royal Society. (2010). *Primary science and mathematics education: Getting the basics right*. London: The Royal Society.

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

UK Department for Education. (Ed.). (2013). *The national curriculum in England. Key stages 1 and 2 framework document*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/335133/PRIMARY_national_curriculum_220714.pdf
Van Joolingen, W., de Jong, T., & Dimitrakopoulout, A. (2007). Issues in computer supported inquiry learning in science. *Journal of Computer Assisted Learning*, *23*(2), 111–119.
Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. Baltimore, MD: ASCD.
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172–223.