

Developing Higher Order Thinking in Elementary School Science: A Narrative Around a Research Group Quest



Melina Furman, Mariana Luzuriaga, and Inés Taylor

The Question Where Our Quest Begins

Imagine the children of today in 10, 20, 50 years' time... What kind of adults do you hope they are?

This is one of the main questions that guide our work and commitment to improving education. At conferences, training sessions, and interviews, we repeatedly ask school principals, teachers, and parents to reflect on their answers to that question.

Frequent responses include “curious,” “creative,” “resilient,” “kind,” “thoughtful,” “critical thinkers,” and “problem-solvers.” Thus, there is a consensus that we aspire for children to have the necessary knowledge and tools to function in the world, to develop a taste for lifelong learning, and to have an awareness of the “common good.” All of these speak to the need for developing higher order thinking skills, which encompass a continuum of cognitive processes, such as understanding, applying, analyzing, evaluating, and creating, across factual, conceptual, procedural, and metacognitive knowledge dimensions (Anderson & Krathwohl, 2001; Bloom et al., 1956; Forehand, 2005), key in developing twenty-first-century citizenship (Scott, 2015).

The next questions we ask are “how do we achieve it?” and “what must we do now to prepare our students for the future we envision?”

One way of doing so is through science education, which has long been considered a field of opportunity to develop higher order thinking skills. In particular, an important goal of science education is the development of scientific thinking,

M. Furman (✉)

School of Education, Universidad de San Andres, Buenos Aires, Argentina

Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

M. Luzuriaga · I. Taylor

School of Education, Universidad de San Andres, Buenos Aires, Argentina

defined as the skill set of “ways of knowing” involved in science, which include higher order thinking skills, such as inductive and deductive reasoning, interpreting and constructing data and models to explain phenomena, designing valid experiments, building evidence-based arguments, and finding and assessing ways to solve socioscientific problems (Klahr et al., 2011).

Although there is evidence that children have some rudiments of scientific thinking skills even at a very young age (Gopnik et al., 2001; Gweon et al., 2010; Gweon & Shultz, 2011), they do not fully develop spontaneously (Kuhn, 2010). Instead, they need to be intentionally taught and put into practice in a systematic and sustained way overtime (Duschl et al., 2007).

An effective science education for all is perhaps even more important within non-WEIRD (Western, Educated, Industrialized, Rich and Democratic) populations, both as a means to foster and increase citizens’ informed participation in critical issues, such as health and environmental matters, and potentially promoting vocational orientations in the fields of science, technology, engineering, and mathematics (STEM), which are fundamental in knowledge-based economies. There is evidence that, in Argentina, the context of this study, as in many other countries, only few students pursue career choices related to STEM, especially among women and populations from socioeconomically vulnerable contexts (Albornoz et al., 2009; Szenkman & Lotitto, 2020).

Consequently, science teaching reform efforts were implemented in several countries over the last decades. As a recent study on the state of affairs regarding science education in primary schools in Latin America shows, starting in the second half of the 2000s, several countries, such as Chile, Colombia, Ecuador, México, Paraguay, and Uruguay, have renewed their science curriculum and implemented professional development programs, all aimed at enhancing science teaching with a particular focus on developing scientific thinking skills (Furman, 2020). This is also the case in Argentina (Tedesco, 2009).

However, low standardized test results and research studies consistently suggest that what happens within classroom contexts is still far from what is expected. Argentine and Latin American students, in general, achieve very low levels of performance in science and particularly struggle to solve questions relating to higher order thinking skills, such as interpreting data and drawing conclusions, analyzing and solving science-related problematic situations, and assessing or proposing experimental designs to answer inquiry questions (Meschengieser & Otero, 2016; UNESCO, 2015). Furthermore, there is great inequity in learning opportunities, with a consistent pattern showing students from wealthier families attaining higher levels of learning achievement than their less privileged peers in many countries (OECD, 2019; Vegas et al., 2014).

Faced with this scenario, we wonder: how can these societies facing structural challenges aspire to thrive when so few students seem to be equipped with the necessary skills? More specifically, the prevalence of such low learning outcomes opens up questions regarding current science teaching practices: are they encouraging children to become the thinkers and problem-solvers we aspire them to be?

Unfortunately, it appears that we aren't. Research on teaching practices, although limited, indicates that teachers across many Latin American countries show persistent difficulties in handling conceptual content and enacting reform-based teaching practices (Furman, 2020; Kisilevsky et al., 2019). In turn, this is consistent with these countries having poor teacher training systems in general (Bruns & Luque, 2015) and in science in particular (Furman & Luzuriaga, 2017; Maiztegui et al., 2000), which seems to negatively affect the quality of science instruction.

Thus, some of the main questions our research projects address are as follows: What do science teachers need to be able to improve their "business as usual" practices? How can we help them move toward enhancing science education and thus student learning of scientific thinking skills?

This is where the "CABA study," which will be described throughout this chapter, came in. Our quest to better understand the current state of scientific thinking skills development in relation to science teaching practices, and how to strengthen them, led us to develop a research project set in a representative sample of schools in the City of Buenos Aires (CABA), Argentina. Through this study we sought to characterize science teaching practices and student learning outcomes, both in general and as a result of teachers receiving specific professional development (the full design and results can be found in Albornoz et al., 2019).

In particular, we set out to rigorously evaluate if providing teachers with educative curriculum materials (ECM), specially designed to foster scientific thinking skills in students, is an effective way to improve science teaching and learning in the local context. ECM are detailed lesson plans, including reading materials, worksheets, and other resources for students, aimed at supporting teachers in the organization, content, and pedagogy of given topics (Ball & Cohen, 1996; Davis & Krajcik, 2005).

ECM became a widespread, regular source of consultation for teachers as well as a key component in curriculum reforms and teacher training policies in general as in science education, endorsed by research that brought attention to their potential, particularly in contexts with low-quality indicators in teacher education and student learning (Bruns & Luque, 2015; Mourshed et al., 2010).

However, while studies show mixed results on the effects of providing teachers with ECM on student learning and teaching practices (Davis et al., 2016), there is a dearth of evidence to this regard in the region. Considering that, within these contexts of resource scarcity, providing ECM represents a significant investment, we identified a pressing need to contribute with large-scale, high-quality, rigorous studies on the implementation and effects of science ECM in student learning of higher order scientific thinking skills as well as in teaching practices.

In this chapter we present a narrative of how this research project unfolded, from our initial questions to its design, the findings and new questions that emerged along the way, to the lessons learned at the end of the process. Although the overall findings for the wider randomized controlled trial were published (Albornoz et al., 2019), as are several more in-depth papers considering different aspects of the pedagogical intervention (Furman et al., 2021; Taylor et al., 2020), for this chapter we conducted further data analysis and present new results focusing on the implementation of ECM.

Designing Our Action Steps

We proposed a randomized controlled trial to evaluate both student learning and teachers' teaching practices in seventh grade (the final year of primary schooling) science, comparing everyday teaching against the effects of supporting teachers in the implementation of scientific, higher order thinking skills oriented teaching practices through the provision of ECM. Two experimental groups were defined: the control group, in which "business as usual" teaching practices and learning outcomes would be observed, and the intervention group, which we provided with ECM.

The study was set in the City of Buenos Aires, the federal capital of Argentina (CABA, for its Spanish acronym). CABA is the most densely populated city and also one of the districts with better socioeconomic and educational indicators at the national level. However, this is within a context of low levels of academic performance and high degrees of educational inequity to the detriment of students from disadvantaged contexts (Di Virgilio & Serrati, 2019; Meschengieser & Otero, 2016).

Mimicking Argentine education policies more generally, participating in this research program was mandatory yet not enforced, and there were no formal consequences if schools or teachers chose not to abide by what was requested (i.e., there were no financial or administrative incentives or penalties). Also, given that conducting research within school contexts is not a well-established practice in the country, this required involving and consulting senior officials of the CABA Ministry of Education, stakeholders from different ministerial areas, and school district superintendents and principals, as their support was essential to access schools and promote teachers' participation.

A sample of 47 state primary schools located across six (out of 21) school districts in the city participated in the study. These schools were representative of schools in CABA in terms of socioeconomic conditions, their total school and seventh grade enrollment, pass and repetition rates, overage rate, and previous district test scores (see Albornoz et al., 2019 for full sample balances).

As shown in Table 1, 24 schools were randomly assigned to the control group and another 23 to the intervention group. Within each group, all seventh grade science teachers participated in the study. After analyzing the differences in the means along with p-values from two-tailed t-tests of equality of means, no significant differences were observed between each group in student-level variables (e.g., gender, socioeconomic background, and academic performance), teacher-level variables (e.g., gender, seniority in seventh grade teaching, and seniority in the current school), nor school-level variables (e.g., location, total enrollment, amount of

Table 1 Sample description

	Control group	Intervention group
Number of schools	24	23
Total seventh grade classrooms	50	44
Total seventh grade students	1086	917
Total seventh grade science teachers	36	32

seventh grade classrooms, percentage of schools with double school day, and school dropout and overage student rates).

We chose to focus our study on one particular science seventh grade curricular topic: the human body, in order to limit the scope of the intervention and favor the comparison between both experimental groups. Besides being one of the prescribed contents in the national and district curricular guidelines (Argentine Ministry of Education, 2005; CABA Ministry of Education, 2012), the human body is regularly taught in primary level science lessons for it's a well-versed topic, often valued by teachers as relevant and interesting for students.

Teachers in the control group addressed the topic of “the human body” as they regularly did in science lessons, allowing us to capture common teaching practices and learning outcomes, while teachers in the intervention group were provided with science ECM to address the topic, focusing on the basics of the digestive, circulatory, and respiratory systems and how they contribute to cellular nutrition.

ECM were developed by the team of researchers based on curriculum guidelines and informed by research-based, best practice science principles (Davis et al., 2014; Davis & Krajcik, 2005; Harlen, 2013) and then discussed with nonparticipating primary science teachers and validated by the curricular team of experts at the CABA Ministry of Education. They ought to support teachers by organizing science content, with detailed indications on how to guide science activities with specific learning goals. Activities were designed to have students explore different topics and carry out diverse tasks following an inquiry-based approach (Harlen & Qualter, 2018). As Table 2 shows, as well as some more traditional activities such as reading from informative texts or diagrams and answering simple questions, the ECM included activities promoting higher order thinking skills, associated with scientific thinking (Anderson & Krathwohl, 2001; Li & Klahr, 2006). The ECM were designed for implementation in an estimated 38.5 hours of science lessons over a maximum of 12 weeks (the original Spanish version of the ECM can be found here (shorturl.at/gnsAR)).

The intervention began at the beginning of the school year, when all participating science teachers were summoned to one in-service 4-hour informative meeting aimed at presenting the research project, encouraging teachers' commitment to participate, and establishing what they were expected to do. All teachers were asked to teach the “human body” in their science lessons over the following 12 weeks, at the end of which students would complete an external assessment in the form of a written test.¹

The written test was developed by our research team based on the content and skills covered in the local curriculum regarding the topic, so that students of teachers who did not use the ECM, but followed the national and district curriculum frameworks as expected, should be able to achieve good learning outcomes. It was reviewed by science teaching experts and then piloted in two seventh grade

¹At the end of the intervention, teachers in the control group received a copy of the ECM.

Table 2 Characterization of the activities present in the provided ECM, according to their cognitive demand

	Activities associated with lower order science thinking skills	Activities associated with higher order science thinking skills
General description	Reading and writing activities based on content-based texts on different aspects related to the human body, specially focused on the learning of scientific vocabulary and facts, and activities involving observing and describing phenomena	Activities based on the analysis of history of science episodes and on designing and conducting experiments, as well as where students have to explain more complex scientific phenomena and reflect on their own understanding
Scientific thinking skills involved	Recall Read information from simple tables and diagrams Observe Describe	Analyze Explain Apply Interpret and draw conclusions from data Design a valid experiment Synthesize information Metacognition (reflect on one's own learning)
Examples of activities present in the provided science ECM	Read and answer simple questions about the importance of a balanced diet Write a sentence about the role of each organ in the digestive system Observe a heart dissection	Design and carry out an experiment on changes in human heart rate under different conditions Read about the experiments conducted by Robert Boyle on the role of oxygen in living cells, identifying his research questions and procedures and explaining how his experiments led to changes in understanding at the time Compare the role of the three systems (digestive, respiratory, and circulatory), and create a concept map on how they work together

classrooms from comparable, nonparticipating schools, based on which necessary adjustments on language, clarity, and content levels were made.

The result was an 11-item, 90-minute written science test, consisting of both multiple-choice and open-ended questions. This combination allowed evaluators to capture a wider range of student responses, including stronger evidence of critical thinking skills, than is typically associated with only multiple-choice tests (Stanger-Hall, 2012). In addition, items aimed to capture learning gains with different levels of complexity according to the demanded thinking skills (see Table 3).

Tests were conducted in class with the classroom teacher and an external researcher present following international examination standards (such as completing the examination individually on separate tables in silence and teachers only allowed to answer questions relating to clarifying the question). Then, student tests were anonymized and randomly allocated to members of the research team for the marking and grading process. The tests were graded on a 10-point scale according to a shared rubric, which had been piloted by researchers and shared with graders.

Table 3 Description and examples of the science test items, according to the cognitive demand involved in each question

Cognitive demand of evaluated thinking skills	Complexity level of the questions, according to the demanded thinking skills	Examples of test items for each level	Maximum score given to correct answers per question
Lower order thinking skills	Level 1: Recalling scientific concepts. Reading data from simple tables and diagrams (4 items)	John ate a plate of ravioli and is digesting. The following diagram shows the organs that form John’s digestive system. Name each organ and briefly describe its function	1
Higher order thinking skills	Level 2: Analyzing and explaining a given situation involving scientific phenomena based on the concepts learned and establishing relationships between them (5 items)	7th grade students were in a physical education class. The teacher asked them to run for 5 minutes. When they finished, everyone noticed that their hearts were beating faster than at the beginning. Why does our heart beat faster when we exercise?	2
	Level 3: Interpreting science-related problematic situations, proposing feasible solutions based on evidence. Designing valid experiments to answer research questions (2 items)	7th graders wanted to investigate whether there are differences in the average heart rate between males and females when they are sitting and resting. Describe in as much detail as you can what experiment they could do to find it out	3

Graders participated in a training session and practiced marking a set of students’ examinations, achieving inter-marker reliability of over 85%.

As shown in Table 3, the science test questions were weighted according to difficulty, with higher order questions scoring 2 or 3 points and lower order questions scoring 1 point. Answers were classified as either “correct” (given full marks), “partially correct” (given half the maximum marks), “incorrect” (no marks given), or “omitted” when no answer was provided by students (no marks given). Overall test results were calculated, and average levels of performance for each group were compared using ordinary least squares, including control variables for student, teacher, and school characteristics (Albornoz et al., 2019).

Reaching a Milestone: A Look at Students’ Science Learning

So what did we find in terms of student learning? Students in the intervention group, in which teachers received ECM, learned more on the topic of the human body than their peers in the control group. When considering the total average test score for

each group, as shown in Fig. 1, we found that the intervention group students significantly outperformed ($p < 0.01$) those in the control group. While students in the control group had an average score of 3.74 points over 10 (SD = 2.08), the average score in the intervention group was 4.79 points over 10 (SD = 2.33). Considering that the pass score in CABA is 4 (out of 10), this means that control group students did not reach the minimum level of performance considered acceptable and that the treatment produced a gain of 0.55 standard deviations in the average test score.

Our next step was to try to understand what happened in terms of students' science thinking skills, for which we analyzed the types of questions students in each group could solve correctly. We found that students in the intervention group achieved a significantly ($p < 0.01$) greater average percentage of correct answers for questions involving both lower and higher order science thinking skills (see Fig. 1). Coincidentally, the average percentage of omitted answers, frequently regarded as an indication of what students find too distant or complex to even risk a possible

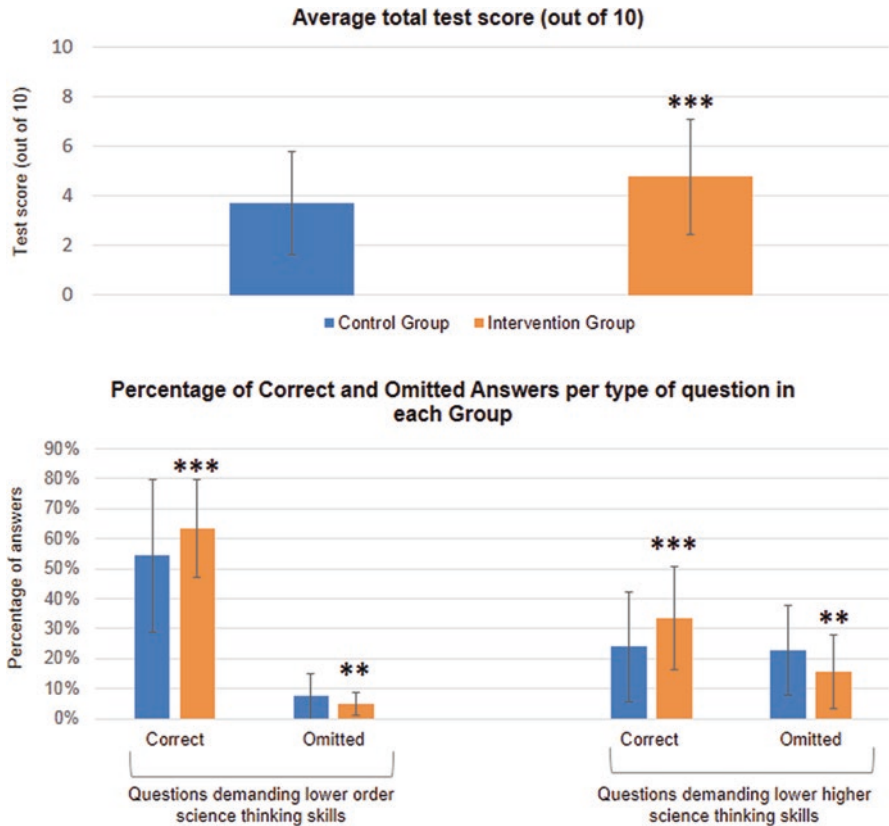


Fig. 1 Average total test score and percentage of correct and omitted answers per type of question for each group (***) = $p < 0.01$; ** = $p < 0.05$). Error bars show standard deviation

answer (Jakwerth et al., 1999; Köhler et al., 2015) was significantly lower in the intervention group ($p < 0.05$).

Encouragingly, our results show that when teachers are provided with science ECM to guide their teaching practices in ways that support higher order science thinking, student learning increases. Yet, overall results remain low and still far from what could be expected. Thus, here, we encountered a new major “think point”: although the provision of ECM contributed to student learning, why were the effects limited?

Perhaps, as accounted for in the literature, we were in the presence of an “implementation gap”, where the ways in which ECM are implemented in classroom contexts differ from their original design (Davis et al., 2016; Penuel et al., 2009). Analyzing ECM implementation requires addressing questions such as the following: Are the proposed activities implemented during instruction? Do teachers choose to implement certain activities in particular over others? How do they carry out these activities? Are the ways in which teachers use ECM consistent with the intended pedagogical goals and rationale?

These led us to want to “open the black box” of the ECM implementation. As we describe below, doing so implied characterizing and comparing what science teaching practices are usually like (as observed in the control group), against teaching practices when teachers are provided with ECM. Thus, a new milestone on the road emerged for us in our quest to understand the current state and opportunities to enhance science education.

Looking Deeper: Understanding Science Teachers’ Teaching Practices

To understand teachers’ teaching practices in each group, we decided to use student notebooks as the main data source. In general, student notebooks are endorsed as a legitimate data source for teaching practices for, despite capturing only written work, particularly at the primary level, they are extensively used during lessons to record classroom activities (Badanelli Rubio & Mahamud Angulo, 2007; del Pozo Andrés & Ramos Zamora, 2012; Gvirtz, 1997). Therefore, at the end of the intervention, all teachers were asked to choose a student notebook they considered to be most complete and thus representative of their science teaching. Student notebooks were photographed page by page and then analyzed by researchers. All information regarding the students’ and schools’ identities was blinded to contribute to performing an unbiased analysis.

Considering an “activity” (i.e., a distinct task specified by the teacher, with a specific learning objective – Cañal de León, 2000) as our basic observation unit, we determined (a) an estimation of the time dedicated to each activity and (b) what types of teaching activities were present in the notebooks. The estimation of the time dedicated to each activity was determined based on previous, similar

interventions and in consultation with experienced science teachers. Three possible values of time were allocated to each activity: for example, 0.5 hour for closing activities aimed at reflecting on learning, 1 hour to complete a short set of questions based on a science text, or 2 hours for experimental activities. In this way, it was possible to calculate the total time destined to science teaching in each class during the intervention.

Also, following the same criteria used to classify ECM activities and test items, the activities found in student notebooks were identified as either promoting lower or higher order science thinking skills (Albornoz et al., 2019). Then, by calculating the percentage of time dedicated to activities demanding higher order science thinking skills versus those demanding lower order science thinking skills, we determined what we called the “cognitive blueprint” of science lessons. Making an analogy with fingerprints, this allowed us to characterize the “cognitive identity” of each science classroom, with a particular focus on what thinking skills students are learning, for both the control and the intervention groups.

Comparing the Cognitive Blueprints of Control and Intervention Group Lessons

When comparing the cognitive blueprint in the control and intervention groups, and that of the proposed science ECM, we found some clear differences, which led us to two important insights. For one, the cognitive blueprints of science lessons in the control and intervention groups were very different. As Fig. 2 shows, in the control group, on average only 20% of the time was dedicated to activities demanding higher order science thinking skills, and 80% of the time was focused on activities involving lower order thinking skills. In other words, “business as usual” science

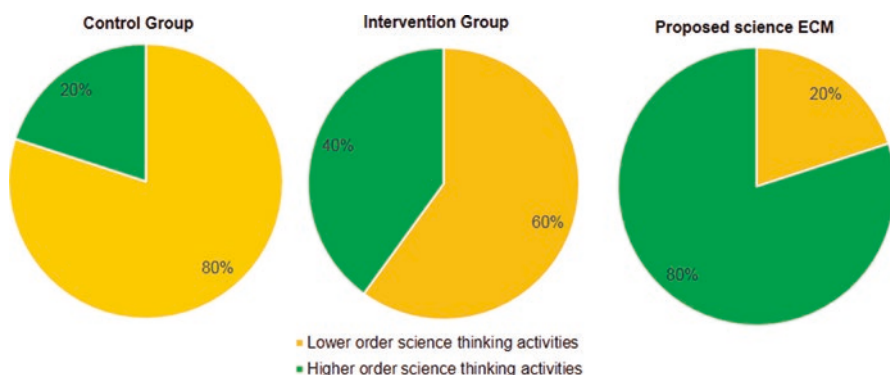


Fig. 2 Cognitive blueprint. Proportion of time dedicated to activities demanding higher and lower order science thinking skills during science lessons in the control and intervention groups and in the proposed science ECM

lessons are characterized by a prevalence of lower order science thinking, with a striking cognitive blueprint of 20/80.

On the other hand, in the intervention group, we found a cognitive blueprint of 40/60, meaning that 40% of the time was dedicated to activities demanding higher order thinking skills, exactly double that seen in the control group. However, this was still far from the original cognitive blueprint in the proposed science ECM (80/20).

So what did this mean in terms of how science lessons are being taught? In the control group, students spend most of their time conducting activities that involved lower order thinking skills: filling their notebooks with information, typically definitions of concepts, and even copying informative texts from their textbooks by hand. Activities based on reading and working with texts were also one of the most predominant across all notebooks, but in the vast majority of cases, the posed questions only demanded for students to reproduce explicit information with a particular focus in specific terminology. The latter was also seen in the great presence of word search puzzles, anagrams, crosswords, and “fill-in-the-blanks” exercises (see, e.g., Fig. 3).

On the other hand, in the intervention group notebooks, there was more frequent evidence of activities that are associated with the promotion of higher order science thinking skills. For example, even when they involved reading texts and answering questions, these usually included questions that implied analyzing and/or inferring information, explaining, and/or drawing conclusions. Experimental activities also had greater presence in the intervention group notebooks and were distinctive, in that in general they followed an inquiry-based approach.

Using the ECM: Patterns of Implementation by Teachers

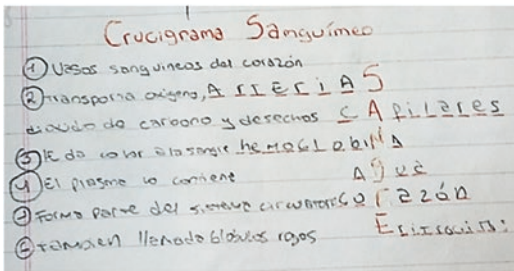
These results led us once again to a new “think point,” around how teachers used the ECM. Given that the cognitive blueprint in the intervention group was 40/60, still far from the original 80/20 of the ECM, we realized that teachers adapted the ECM by lowering its cognitive load.

Considering that previous research raises certain “red flags” around teachers’ implementation of ECM, pointing out that lowering the cognitive load of proposed activities is a regular practice (see, e.g., Davis et al., 2016), we wondered: With what frequency and intensity did teachers use the ECM provided? Which activities did they select to implement with their students? How did they adapt them? Did they propose other activities besides those from the ECM?

A closer look at how teachers in the intervention group used the ECM allowed us to address these questions. For this, we analyzed which of the activities proposed in the ECM teachers chose to implement, identifying whether activities present in the intervention group student notebooks had come from the provided ECM or if they used activities from elsewhere.

Overall, we found great diversity in how much teachers used the ECM in their science lessons but, at the same time, certain common patterns as to the ways in

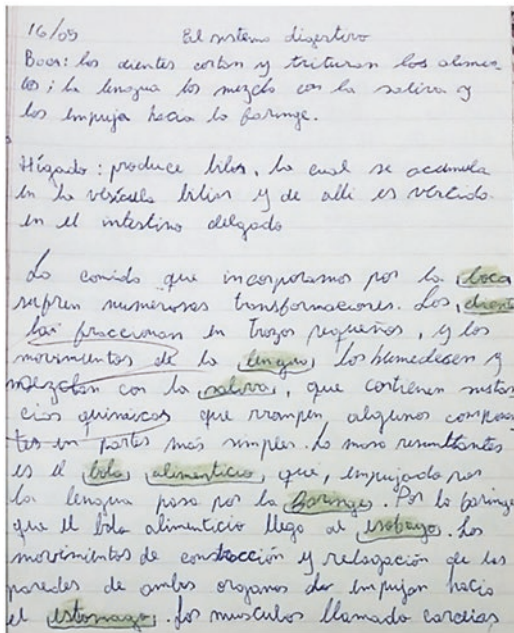
A. Illustrative examples of lower order activities that predominated in the Control Group student science notebooks



Example of lower order, terminology centered activity. In this case, a crossword puzzle on the topic of the circulatory system. Based on the six provided definitions (on the left), students had to complete the crossword on the right.

For example:

(1) Blood vessels in the heart - answer: "arteries"



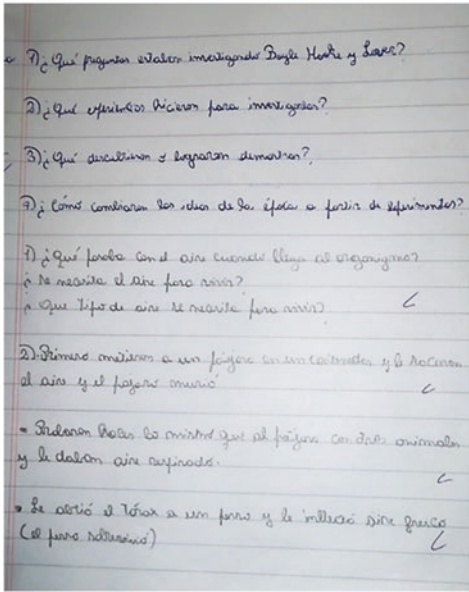
Example of an activity based on copying information (typically from textbooks, the board or dictated by the teacher). In this case, under the title "The digestive system", it defines the functions of the mouth and the liver. E.g.: "Mouth: teeth cut and grind food; the tongue mixes them with saliva and pushes them into the pharynx".

Then, the digestive process is described. Terms referring to the involved organs ("mouth", "tongue", "stomach") and other keywords ("saliva", "alimentary bolus") are highlighted.

Fig. 3 Illustrative examples of activities present in the control group and intervention group science student notebooks

which they were implemented. On average, teachers implemented 27.17% (SD = 17.98) of the activities in the ECM, ranging from teachers that implemented 0% ($n = 3$) to over 60% ($n = 2$) of the total activities. Most teachers predominantly used the ECM (as opposed to other activities of their own choice) in their science lessons. When calculating what percentage of the activities present in each notebook belonged to the provided ECM, the median was 73.2% (SD = 30.59). This means that, although teachers used a limited percentage of the activities proposed in the ECM, in general, they did not add other activities beyond those present in the

B. Illustrative examples of higher order activities found in the Intervention Group student science notebooks



Questionnaire based on a history of science episode, in this case, the experiments performed by Boyle, Hooke and Lower on the respiratory system.

Questions included: (1) What questions were Boyle, Hooke and Lower investigating? (2) What experiments did they perform to investigate them? (3) What did they find out and managed to show? (4) How did the ideas from their time change based on their experiments?



Posters produced by the students, showing the results of an experimental activity. In groups, they had to define a research question (e.g. Is there a difference in the average height of female and male students? Are left-handed students' hands bigger than right-handed students?), acquire the necessary data (measuring students' heights and length of their hands), analyze them, take averages, build graphs, discuss results in their teams, and draw conclusions.

Fig. 19.3 (continued)

ECM. They simply used fewer activities than those suggested in the ECM to teach the human body.

But which activities did they choose to implement? Interestingly, we found salient patterns in the types of activities that teachers selected (and those that they did not). In this way, we distinguished between what we called “popular” activities, that is, those that the vast majority of teachers implemented, and “unpopular” activities, which were rarely chosen. Popular activities were mainly those based on reading and answering questions on texts, while metacognitive (i.e., students reflecting

on their own learning) and some experimental activities were identified as unpopular. Given that the popular activities involved lower order thinking skills, this explains why the cognitive blueprint of the intervention group showed a higher proportion of lower order thinking activities than originally intended by the ECM (40/60 versus 80/20).

Identifying popular and unpopular activities led us to think on what makes teachers choose to implement or skip certain ECM activities. Our results show that when teachers received ECM predominantly based on more demanding activities than those they regularly propose to students, they managed to at least partially incorporate them into their teaching practices. However, they also adapted ECM by “cherry-picking” activities, possibly favoring those that were closer to their regular practice, with which they might have felt most comfortable or confident about, and that they considered their students would be able to solve without major difficulties.

Conclusions: Lessons Learned and Opening of New Horizons

“What do we want the children and youth of today to be like in the future, and how do we achieve this?” This is the big question that initiated our quest and still sets the course of our work as a research team. In this chapter we narrate the questions and insights that emerged when conducting a research project on primary level student learning and teaching practices in science, in a context where there is a pressing need to enhance science education as an opportunity to foster higher order thinking skills. Along this way, we learned many valuable things, relevant in Argentina as in other non-WEIRD contexts.

First, we learned that opening the “black box” of science lessons in a rigorous manner contributes to understanding why student learning outcomes in the area are lower than expected. In Argentina, as in much of Latin America, research on the insides of classrooms and science teaching practices is still insufficient. Beyond those based on national or international standardized assessment programs, in general, few studies attend to student learning, and even less so that experimentally evaluate the effects of teaching and learning improvement interventions. In this sense, our study broadens the field of knowledge, potentially favoring the design, implementation, and evaluation of evidence-based interventions, so necessary in contexts such as ours.

Our findings when analyzing student notebooks and the cognitive blueprint of science lessons reveal that, in its regular form (as shown in the control group), science teaching in CABA is far from what is promoted by the literature and curricular policy, with a clear predominance of lower order activities focused on recalling factual information. They present an alarming picture of what and how our students are learning in science, showing that their opportunities to develop higher order science thinking skills are very limited, thus explaining the observed learning results. Having students spend 80% of their time in science lessons performing tasks that involve lower order science thinking skills is a matter that needs to be urgently

addressed. This is of particular importance within non-WEIRD populations, considering that developing higher order science thinking skills is key to promoting critical thinking; citizens' informed and active participation in social, health, environmental, and economic issues; and the centrality of science and technology in economic development.

From our results it becomes evident that there is still a long way to go for science teaching practices to contribute to attaining the goal of fostering scientific thinking for all. Perhaps this is not entirely surprising, given that it has been reported that preservice teaching programs in the region are deficient in this regard. They predominantly focus on teaching contents also through traditional approaches rather than modeling and promoting the enactment and reflection on reform-based teaching practices oriented toward the learning of higher order science thinking skills (Cofré et al., 2015; Furman & Luzuriaga, 2017). Therefore, our study points toward the need to provide support to both preservice and in-service science teachers to strengthen their knowledge and practices.

Along this line, the second thing we learned is that research-based ECM are effective to provide this needed support to teachers to move toward higher order thinking skills oriented science teaching practices and that this has positive effects on student learning. What we identified as an alarming state of affairs of science teaching in the control group was improved in the intervention group where the cognitive blueprint of science lessons showed a higher proportion of activities demanding higher order science thinking skills. In turn, these changes in teaching practices are consistent with the improvement in the intervention group students' performance in the science test, both overall and specifically in the higher order questions. In other words, even when we are faced with a critical scenario, the condition of science education in CABA and, arguably, in similar contexts is not a lost cause. We found proof that it is possible to start up the path of improvement in the short term through the provision of ECM.

However, we also learned that ECM are not enough to profoundly transform science teaching practices and maximize student learning. Similarly to what was found in other contexts and educative levels (Beyer & Davis, 2012; Davis et al., 2016; Furman et al., 2017), in this study teachers in the intervention group selected and adapted the provided ECM in ways that resulted in lowering the cognitive load of the suggested activities. This scenario opened up new questions that are worth addressing in future research.

In particular, one of the main concerns that emerged is what is needed to go "the extra mile" toward enhancing the science education we aspire to offer our students? Is it a matter of time, in the sense that these types of interventions based on the provision of ECM need to be longer and sustained overtime? Are there other strategies that can be put in place to complement and deepen the effects of teacher professional development interventions? What other changes are called for?

As both previous research and our own study point out, the need to provide significant and sustained support for teachers is of clear importance. On the one hand, this has implications for in-service teacher professional development programs based on ECM. Particularly, to consider, design, and evaluate professional

development strategies that may enhance the use of ECM, providing tools and learning opportunities to teachers that, while contributing to their professional autonomy, foster their abilities to select, curate, adapt, and design curriculum materials focused on the development of higher order science thinking skills. For instance, mentoring and instructional coaching programs, which provide teachers with individualized, relationship-based, context-specific, intensive, and sustained support (Desimone & Pak, 2017; Knight, 2007; Tschannen-Moran & Carter, 2016), may be promising to facilitate teachers to embody knowledge and discrete skills associated with research-based instructional practices (Joyce & Showers, 2002).

There is also evidence that suggests the need to take into account the institutional dimension of educational change. Considering the importance of evaluating the long-lasting effect of teacher professional development efforts, 1 year after the end of our intervention, we performed a follow-up study to reveal if participating teachers continued using the provided ECM. Surprisingly, we found that only 27% of the teachers in the intervention group continued to teach science in the seventh grade the following year. Even when all of those remaining teachers reported that they had used again the given ECM to teach the topic of the human body, which advocates the sustainability of our intervention, concerns arise regarding teaching turnover, which may dissipate part of the science teaching reform efforts. This could be indicating the importance of framing these types of interventions within the schools' institutional projects with the principals' support and endorsement to help sustain changes in teaching practices and thus student learning.

In non-WEIRD contexts such as ours, investing in in-service science teacher professional development is crucial as a large-scale, economically viable solution to ensure that students who are currently going through their schooling do not miss valuable learning opportunities while working together with teachers to strengthen their knowledge and skills (Bruns & Luque, 2015). But this is only part of the equation; deeper, systemic, long-term changes are also necessary, which demand attending to preservice teacher education and improving teachers' working conditions.

In all, what did we learn as researchers invested in science education in non-WEIRD contexts? The stakes are high; the need for change is urgent. But it is also possible, and that will continue to be our quest.

References

- Albornoz, F., Anauati, M. V., Furman, M., Luzuriaga, M., Podestá, M. E., & Taylor, I. (2019). Training to teach science: Experimental evidence from Argentina. *The World Bank Economic Review*, 34(2), 393–417.
- Albornoz, M., Marchesi, A., & Arana, L. (2009). *Cultura científica en Iberoamérica. Encuesta en grandes núcleos urbanos*. Madrid, España: FECYT, OEI, RICYT.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Argentine Ministry of Education. (2005). *Núcleos de Aprendizaje Prioritarios Séptimo Año*. Buenos Aires, Argentina: Consejo Federal de Educación, Ministerio de Educación.

- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—Or might be—The role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, 25(9), 6–14.
- Badanelli Rubio, A. M., & Mahamud Angulo, K. (2007). Posibilidades y limitaciones del cuaderno escolar como material curricular: un estudio de caso. *Avances en Supervisión Educativa*, 6(junio), 1–16.
- Beyer, C. J., & Davis, E. A. (2012). Developing preservice elementary teachers' pedagogical design capacity for reform-based curriculum design. *Curriculum Inquiry*, 42(3), 386–413.
- Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). Taxonomy of educational objectives: The classification of educational goals. In *Handbook I: Cognitive domain*. Longmans.
- Bruns, B., & Luque, J. (2015). *Profesores Excelentes*. Cómo mejorar el aprendizaje en América Latina y el Caribe.
- CABA Ministry of Education. (2012). *Diseño curricular para la Escuela Primaria (segundo ciclo)*. CABA, Argentina: Ministerio de Educación del Gobierno de la Ciudad Autónoma de Buenos Aires.
- Cañal de León, P. (2000) Las actividades de enseñanza. Un esquema de clasificación. *Investigación en la escuela*, 40, 5–21.
- Cofré, H., González-Weil, C., Vergara, C., Santibáñez, D., Ahumada, G., Furman, M., ... Pérez, R. (2015). Science teacher education in South America: The case of Argentina, Colombia and Chile. *Journal of Science Teacher Education*, 26(1), 45–63.
- Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, 34(3), 3–14.
- Davis, E., Palincsar, A. S., Arias, A. M., Bismack, A. S., Marulis, L., & Iwashyna, S. (2014). Designing educative curriculum materials: A theoretically and empirically driven process. *Harvard Educational Review*, 84(1), 24–52.
- Davis, E. A., Janssen, F. J. J. M., & Van Driel, J. H. (2016). Teachers and science curriculum materials: Where we are and where we need to go. *Studies in Science Education*, 52(2), 127–160. <https://doi.org/10.1080/03057267.2016.1161701>
- del Pozo Andrés, M. D. M., & Ramos Zamora, S. (2012). Reconstruir la historia de la escuela a través de los cuadernos escolares. *Revista de ciencias de la educación*, 231–232(julio-diciembre), 415–434.
- Desimone, L. M., & Pak, K. (2017). Instructional coaching as high-quality professional development. *Theory Into Practice*, 56(1), 3–12.
- Di Virgilio, M., & Serrati, P. (2019). *Las desigualdades educativas en clave territorial*. CABA, Argentina: Unidad de Evaluación Integral de la Calidad y Equidad Educativa, Ministerio de Educación e Innovación, GCABA y Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura.
- Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Takin science to school: Learning and teaching science in grades K-8*. National Academies Press.
- Forehand, M. (2005). Bloom's taxonomy: Original and revised. In M. Orey (Ed.), *Emerging perspectives on learning, teaching and technology*. Global Text Project.
- Furman, M. (2020). *Aprender ciencias en las escuelas primarias de América Latina*. Montevideo, Uruguay: Oficina Regional de Ciencias para América Latina y el Caribe, Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.
- Furman, M., & Luzuriaga, M. (2017). Science teacher education in Argentina. In Pedersen, J., Isozaki, T., y Hirano, T. (Eds.) *Model Science Teacher Preparation Programs. An International Comparison of What Works*. Charlotte, NC, United States of America: Information Age Publishing.
- Furman, M., Luzuriaga, M., Taylor, I., Podestá, M. E., & Jarvis, D. (2017). From inception to implementation: An argentine case study of teachers enacting early years inquiry-based science. *Early Years*, 39(4), 408–425.

- Furman, M., Luzuriaga, M., Taylor, I., & Podestá, M. E. (2021). How does coaching influence teacher implementation of a science programme? Evidence from an experimental study. *International Journal of Mentoring and Coaching in Education*, 10(4), 449–465. <https://doi.org/10.1108/IJMCE-09-2020-0059>
- Gvirtz, S. (1997). *Del currículum prescripto al currículum enseñado. Una mirada a los cuadernos de clase*. Buenos Aires: Aique.
- Gopnik, A., Sobel, D., Schultz, L., & Gylmour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37(5), 620.
- Gweon, H., & Shultz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science*, 332, 1524.
- Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, 107(20), 9066–9071.
- Harlen, W. (2013). Inquiry-based learning in science and mathematics. *Review of Science, Mathematics and ICT Education*, 7(2), 9–33.
- Harlen, W., & Qualter, A. (2018). *The teaching of science in primary schools*. Routledge.
- Jakwerth, P. M., Stancavage, F. B., & Reed, E. D. (1999). *An investigation of why students do not respond to questions, NAEP validity studies, working paper series*. American Institutes for Research.
- Joyce, B., & Showers, B. (2002). *Designing training and peer coaching: Our needs for learning*. Alexandria, VA, United States of America: ASCD.
- Kisilevsky, M., Zapata, C., Fridman, M., & Rinaldi, L. (Eds.). (2019). *Estudio Nacional 2016-2017. Prácticas de enseñanza en el contexto del aula en los niveles primario y secundario*. CABA, Argentina: Instituto Nacional de Formación Docente, Secretaría de Innovación y Calidad Educativa, Ministerio de Educación, Cultura, Ciencia y Tecnología de Argentina.
- Klahr, D., Zimmerman, C., & Jiorut, J. (2011). Educational interventions to advance Children's scientific thinking. *Science*, 333, 971–975.
- Knight, J. (2007). *Instructional coaching: A partnership approach to improving instruction*. Thousand Oaks, CA, United States of America: Corwin Press.
- Köhler, C., Pohl, S., & Carstensen, C. H. (2015). Investigating mechanisms for missing responses in competence tests. *Psychological Test and Assessment Modeling*, 57(4), 499–522.
- Kuhn, D. (2010). What is scientific thinking and how does it develop? In Goswami, U. (Ed.) *Handbook of childhood cognitive development*. West Sussex, United Kingdom: Blackwell Publishing Ltd.
- Li, J., & Klahr, D. (2006). The psychology of scientific thinking: Implications for science teaching and learning. In J. Rhoton & P. Shane (Eds.) *Teaching science in the 21st century*. Arlington, VA, United States of America: NSTA Press.
- Maiztegui, A., González, E., Tricárico, H., Salinas, J., de Carvalho, A. P., & Gil, D. (2000). La formación de los profesores de ciencias en Iberoamérica. *Revista Iberoamericana de Educación*, 24, 163–187.
- Meschengieser, C., & Otero, M. P. (2016). *Informe de resultados TIMSS 2015. Ciudad Autónoma de Buenos Aires*. CABA, Argentina: Unidad de Evaluación Integral de la Calidad y Equidad Educativa.
- Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. McKinsey & Company.
- OECD. (2019). *PISA 2018 results. What students know and can do*. Paris, France: OECD.
- Penuel, W., Fishman, B. J., Gallagher, L. P., Korbak, C., & Lopez-Prado, B. (2009). Is alignment enough? Investigating the effects of state policies and professional development on science curriculum implementation. *Science Education*, 93(4), 656–677. <https://doi.org/10.1002/sce.20321>
- Scott, C. L. (2015). *El futuro del aprendizaje 2 ¿Qué tipo de aprendizaje se necesita en el siglo XXI? Investigación y prospectiva en educación*. Documentos de Trabajo.

- Stanger-Hall, K. F. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE-Life Sciences Education*, 11(3), 294–306.
- Szenkman, P., & Lotitto, E. (2020). *Mujeres en STEM: cómo romper con el círculo vicioso. Documento de políticas públicas 224*. Buenos Aires, Argentina: CIPPEC.
- Taylor, I., Sánchez, M., Luzuriaga, M., Podestá, M. E. & Furman, M. (2020). Marking classwork in primary science: written feedback practices in state schools in Buenos Aires, Argentina. *Assessment in Education: Principles, Policy & Practice*, <https://doi.org/10.1080/0969594X.2020.172969>
- Tedesco, J. C. (2009). Prioridad a la enseñanza de las Ciencias: Una decisión política. In *Educación, Ciencia, Tecnología y Sociedad* (Vol. 03). Madrid, España: OEI, Centro de altos estudios universitarios.
- Tschannen-Moran, M., & Carter, C. B. (2016). Cultivating the emotional intelligence of instructional coaches. *International Journal of Mentoring and Coaching in Education*, 5(4), 287–303.
- UNESCO. (2015). *Informe de resultados TERCE. Logros de aprendizaje*. Santiago de Chile, Chile: OREALC/UNESCO.
- Vegas, E., Ganimian, A., & Bos, M. S. (2014). *América Latina en PISA 2012: ¿Cuántos estudiantes tienen bajo desempeño?* Washington, DC: Banco Interamericano de Desarrollo (BID).