Research Article

The Practice of Using Evidence in Kindergarten: The Role of Purposeful Observation

Sabela F. Monteira and María Pilar Jiménez-Aleixandre

Universidade de Santiago de Compostela, Santiago de Compostela, Spain

Received 7 December 2014; Accepted 18 May 2015

Abstract: This article examines kindergarten children's (5–6 years old) engagement in scientific practices, with a focus on generating and using evidence to support claims, during a 5-month project about snails. The research questions are as follows: (1) what meanings do kindergarteners construct for what constitutes evidence? How are those meanings reflected in the development of data into evidence? (2) Which ways of gathering empirical evidence are jointly constructed by children and teacher during the project? (3) How do children use evidence to revise their understandings? The participants are one class of Early Childhood Education children ($N = 25$) and their teacher. They were engaged in a project about snails, involving pursuing their own questions, carrying out experiments and purposeful observations, collecting data and drawing conclusions, under the guidance of the teacher. The results show that children developed meanings of a certain level of sophistication about evidence, that they distinguished between empirical evidence from planned experiments and from prolonged observation, which we call purposeful, and that they combined different types of evidence in the revision of their ideas about snails. We identified two levels in the development of data into evidence—closer to descriptive statements and evaluative judgments. We suggest that purposeful observation, which has a clear focus, is guided by the teacher and explicitly discussed, has affordances in early childhood science. For instance, 30 out of 57 evidence statements relate to purposeful observation. Promoting purposeful observation as a source of evidence at this age may allow studying processes both for children (biology processes) and for researchers (learning processes). The results would support Metz's (2011) contention about the relevance of instructional opportunities over developmental constraints. © 2015 Wiley Periodicals, Inc. J Res Sci Teach
Keywords: purposeful observation; kindergarten science; use of evidence; scientific practices

Résumé : Examínase a participación do alumnado de educación infantil (5–6 anos) nas prácticas científicas, en concreto en xerar e usar probas para sustentar conclusións, durante un proxecto de cinco meses sobre caracois. As preguntas de investigación son: (1) Que significados constrúen os nenos e nenas para o que constitúen probas? Como se reflicten estes significados no desenvolvemento de datos en probas? (2) Que formas de obter probas empíricas son construídas conxuntamente por nenos e mestra durante o proxecto? e (3) Como usan os nenos e nenas as probas para revisar o seu coñecemento? Os participantes son unha clase de terceiro curso de Educación Infantil ($N = 25$) e a súa mestra. Levaron a cabo un proxecto sobre caracois, procurando respostas ás suas propias preguntas, realizando experimentos e observacións cun propósito, recollendo datos e extraendo conclusións, guiados pola mestra. Os resultados mostran que desenvolveron

Contract grant sponsor: Spanish Ministerio de Economía y Competitividad (MINECO); Contract grant number: EDU2012-38022-C02-01.
Correspondence to: S.F. Monteira; E-mail: sabela.fernandez.monteira@usc.es
DOI 10.1002/tea.21259
Published online in Wiley Online Library (wileyonlinelibrary.com).

significados de certa sofisticación sobre as probas, distinguindo entre probas procedentes de experimentos planificados e da observación prolongada, que denominamos cun propósito; e que combinaron diferentes tipos de probas na revisión das súas ideas sobre os caracois. Identificamos dous niveis na transformación de datos en probas, enunciados cercanos a descricións e xuízos avaliativos. Suxerimos que a observación cun propósito, caracterizada por ter un obxectivo definido, estar guiada pola mestra e ser discutida explicitamente, ten potencial no ensino das ciencias en educación infantil e primaria. Por exemplo, 30 dos 57 enunciados sobre probas relaciónanse coa observación cun propósito. Promover a observación cun propósito como fonte de probas nestas idades pode permitir estudar procesos, tanto polos nenos (procesos biolóxicos) como polas investigadoras (procesos de aprendizaxe). Os resultados sutentan a perspectiva de Metz (2011), respecto da relevancia da instrución sobre limitacións debidas ao desenvolvemento.

Scientific Practices in Early Childhood

Interest in students' engagement in epistemic practices has been growing in the last decade, which is reflected in research (e.g., Chinn, Buckland, & Samarapungavan, 2011), complementing studies about epistemic beliefs, or beliefs about science and scientific knowledge. We draw from Kelly's (2008) definition of epistemic practices as "the specific ways members of a community propose, justify, evaluate, and legitimize knowledge claims within a disciplinary framework" (Kelly, p. 99). In policy there has been a move toward situating scientific practices at the center of teaching and learning. This trend is consistent with a model that views science as consisting of a set of scientific practices (Osborne, 2014). As Chinn et al. (2011) point out, the rise of epistemological naturalism has produced a shift from examining whether we can have knowledge at all toward how individuals and communities generate knowledge. Our article examines young children's generation of knowledge through engagement in scientific practices. The focus is on generating and using evidence, on the meanings constructed for what constitutes evidence, and on the role of evidence on the revision of children's understandings. Most studies about epistemic practices focus on secondary or middle school (e.g., Pluta, Chinn, & Duncan, 2011). Although there is a small body of research about inquiry in early primary (e.g., Metz, 2008, 2011; Varelas & Pappas, 2013) and kindergarten (e.g., Mantzicopoulos, Samarapungavan, & Patrick, 2009; Siry & Max, 2013), the use of evidence by young children to test hypotheses or support claims is an understudied issue. What our study seeks to add to the literature is an examination of younger children's engagement with the practice of using evidence, in particular, the meanings constructed for evidence, how is it gathered, and the role of evidence in revising knowledge. This would help to support their engagement with scientific practices at increasing levels of sophistication.

The three interconnected research questions driving the article are as follows:

(1) What meanings do kindergarteners construct for what constitutes evidence? How are those meanings reflected in the development of data into evidence?
(2) Which ways of gathering empirical evidence are jointly constructed by children and teacher during the project?
(3) How do children use evidence to revise their understandings?

In this article, we use the term Early Childhood Education (ECE) to refer to the three first years of school, from 3 to 6 years of age, in Spain and most European countries. The Spanish educational system provides opportunities for exploring young children's learning, because (i)

ECE is part of state schooling from three years of age (teachers fought the "pre-school" denomination, arguing that it is part of school); (ii) children stay with the same teacher during the 3 years of ECE.

Rationale: Children's Engagement in Practices and the Use of Evidence

In this section, we first of all review work about children's engagement in science in early primary and ECE; secondly, we discuss the relationships between observation and evidence construction; and thirdly, we address the use of evidence in early ages.

Promoting Young Children's Engagement in Science

In the literature, there is an on-going debate about young children's reasoning skills and their abilities to coordinate claims with evidence. Two decades ago Kathleen Metz (1995) raised criticisms against assumptions of developmental constraints that would limit young children's engagement in inquiry and scientific practices. A debate ensued in the *Review of Educational Research* between Deanna Kuhn (1997), who suggested that research in developmental psychology should be viewed more as guideposts than as constraints, and Metz (1997) who attributed reported children's failures in scientific reasoning to weak knowledge rather than to developmental deficiencies. At the heart of the controversy is the issue of whether young children are able to engage in epistemic reasoning or not. Metz (2008) has been carrying out a research program exploring to what extent primary grade children's reasoning capacities are sensitive to instructional opportunities. Metz's (2011) conclusion is that some cognitive developmental approaches underestimate children's capabilities because they ignore the impact of instruction. Other authors reached similar conclusions: Sandoval, Sodian, Koerber, and Wong's (2014) review of cognitive development research showed that young children possess capabilities that can be productively built upon by science instruction. Work in the program *Integrated Science Literacy Enactments* (ISLE) (Varelas & Pappas, 2013) provided evidence that primary pupils are able to engage in explanatory reasoning.

Data from the Program for International Students' Assessment (PISA) show the critical role of ECE in the school performance of 15-year-old students (OECD, 2012). The few studies about pre-primary also point to the need for a change of focus from deficits to preschoolers' competence (Gelman & Brenneman, 2012). Early science interests and potential gender differences in them are considered a relevant dimension: Leibham, Alexander, and Johnson's (2013) longitudinal study showed that science interests between the ages of 4 and 6 were related to higher science self-concepts and achievement by age 8 for girls (not for boys). They found no evidence of gender differences in science achievement at age 8. Other work suggests that when kindergartners participate in reform-oriented curriculum, such as the Scientific Literacy Project (SLP), there are no motivation or achievement differences between genders (Patrick, Mantzicopoulos, & Samarapungavan, 2009). A science curriculum including investigations mediated by kindergarteners' interests (Siry & Max, 2013), supported children in developing and refining explanations. These results point to the relevance of providing rich environments and starting from children's interests.

The Relevance of Observation in Gathering Data and Constructing Evidence

The relevance of observation for young learners has been generally acknowledged. In the NRC (2012) framework the progression of practices across grades stresses observations related to direct experience in grades K-2. Gelman and Brenneman (2012) highlight the central role of observation in constructing science knowledge. Their *Preschool Pathways to Science* (PrePS) program focuses on five key practices, paying attention to observation skills. The program introduces children to systematic observation, to noticing something instead of merely seeing it, so they "come to think differently about an item that they are observing instead of just glancing at" (Gelman & Brenneman, 2012, p. 162). Observation is not identified as one of the scientific practices in the NRC (2012) framework, but it is part of some of them, in particular of planning and carrying out investigations, (PCOI) (Duschl & Bybee, 2014).

We are focusing on the role of observation not just as a part of PCOI, but on its own, as a process providing opportunities for collecting and analyzing data, in other words, for constructing empirical evidence. In their work with primary children, Varelas and Pappas (2013) acknowledge that data can be gathered and interpreted either in the context of experiments or observation. In both cases, the students themselves generate first-hand data rather than second-hand data generated by others. Comparisons of the use of first- and second-hand data showed that middle school students voiced a higher sense of ownership with regard to first-hand data (Delen & Krajcik, in press; Hug & McNeill, 2008). In Delen and Krajcik's study, students created stronger explanations when analyzing their first-hand data. However, Hug and McNeill found benefits and limitations in both types of data.

In order to be productive for constructing empirical evidence, observation needs to meet criteria as systematic (Gelman & Brenneman, 2012), active, and having a purpose. Alexander (2008) considers purpose as a relevant feature of dialogic teaching seeking to promote certain types of talk and discourse. Prolonged observation has affordances that short-term observation is lacking; for instance, it provides opportunities to explore processes, and it may be used and revised in order to refine ideas. We call *purposeful observation* one that is prolonged, systematic, with a clear focus, discussed and used to test claims, and compare initial ideas with later ones. The notion of purposeful observation is drawn from medical training (Morris, 2007), where it is described as considering (1) what the learner is asked to observe; (2) what prior knowledge the observation activity assumes; and (3) what the intended learning outcomes (i.e., the purpose) of the observation activity are. It is a notion connected to observation as research instrument (Merriam, 2009), which serves a research purpose, is deliberately planned and systematically recorded. We suggest that this idea can be extended to children's observation of beings or phenomena.

Scientific Practices and Use of Evidence in Kindergarten: Goals and Challenges

Recent understandings of science view it as a set of scientific practices (Osborne, 2014). Osborne argues that science education needs to include explaining why we know what we know, for doing so will contribute to a commitment to evidence as the epistemic basis of beliefs. Policy reforms in the United States are aligned with this approach, adopting a focus on scientific practices alongside crosscutting concepts and core ideas as exemplified by the New Generation Science Standards (NGSS) (Achieve, 2013; NRC, 2012). The examination of the use of evidence is framed in this notion of scientific practices, as a part of practice 7 *Engaging in argument from evidence*.

The focus is on children's construction of the meanings for evidence and on the practice of its use. Evidence evaluation is a central component in argumentation (Jiménez-Aleixandre, 2008) and in the construction of explanations (McNeill & Krajcik, 2008). Most of the work about the use of evidence in primary and secondary classrooms is framed in the construction of explanations. Duschl's (2008) Evidence-Explanation (E-E) continuum involves three critical transformations or students' judgments: (1) selecting or generating data to become evidence; (2) using evidence to ascertain patterns of evidence and models; and (3) employing the models and patterns to propose explanations. As Duschl points out, each of these transitions involves making epistemic judgments about "what counts" as data, evidence or explanations. In her study with fifth grade, McNeill (2011) investigated changes in pupils' ideas about explanation, argument and evidence. Gotwals, Songer, and Bullard (2012) and Gotwals and Songer (2013) discussed a progression for evidence-based explanations from fourth to sixth grade, focusing on the following: (1) articulation of claims; (2) use of appropriate and sufficient evidence to support these claims; and (3) use of reasoning that draws on scientific principles to link the evidence to the claim. We acknowledge building explanations as a goal of using evidence; however, our work with kindergarteners focuses on the first stages: articulating claims, selecting or generating data to become evidence, identifying patterns, and using evidence to support claims. We have only located one paper about entry points in argumentation in kindergarten (Gotwals, Hokayem, & Wright, 2014) whose preliminary results showed that, given appropriate learning opportunities, children may engage in supporting their claims with evidence.

About the terms related to evidence, like McNeill and Krajcik (2008), we consider a claim "a statement or conclusion that answers the original question" (p. 60). Kuhn and Pearsall (2000) consider that a statement is a theoretical claim if it is potentially falsifiable by evidence. For McNeill (2011), evidence is data that is used to answer a question, solve a problem, or make a decision, but many authors restrict its meaning to data meeting certain conditions. An influential characterization of evidence is scientific data that is both appropriate and sufficient (McNeill & Krajcik, 2008), used for instance by Gotwals and Songer (2013). An illuminating discussion about in which cases data would become evidence is Aikenhead's (2005), who emphasizes the judgment of the data's significance as evidence. In the context of kindergarten, we are focusing first on this "what counts" as evidence dimension, in other words, data are considered evidence if their discursive role in children's talk is to evaluate a claim; and second, on the appropriateness of data to support the claim.

Research reports difficulties experienced by middle and high school students in coordinating evidence with claims. For instance, often students did not support claims with evidence (Jiménez-Aleixandre, Bugallo, & Duschl, 2000), used inappropriate criteria in evaluating the significance of pieces of evidence (Hogan & Maglienti, 2001), or struggled with explaining how a given piece of evidence supported a claim (Sandoval & Millwood, 2005). Reasoning was identified as the most difficult aspect for sixth graders (Gotwals & Songer, 2013). Reasoning is also challenging for upper elementary students as well as understanding what counts as evidence. For example, McNeill (2011) found that the majority of fifth graders did not explicitly talk about data in their discussion of evidence, talking about it rather as supporting an answer to a question. In Songer and Gotwals' (2012) study, the most difficult aspect for fourth graders was generating evidence, in particular, providing two pieces of valid evidence that matched their claims, and for fifth graders it was generating reasoning.

The question is what our goal should be in early childhood. It should be noted that our focus is on the enactment of scientific and epistemic practices, or children's practical epistemologies (Metz, 2011; Wickman, 2004). In other words, the focus is on functional understanding of the nature of science rather than on declarative understandings (Allchin, 2011). Previous studies (Lehrer & Schauble, 2012; Mantzicopoulos et al., 2009; Metz, 2011) support children's potential for engaging in epistemic reasoning. Similarly to Gotwals et al. (2014), we favor a focus on what students bring with them, rather than on identifying their difficulties with argument components. What we seek to add to this literature is to document how children use evidence to support claims at different epistemic levels, from closer to data to evaluative judgments, and how they use evidence to evaluate and revise knowledge.

Research Methodology, Context, and Participants

Research Design

Framed in qualitative methods, this article reports a case study, appropriate in cases when little is known about one issue (Yin, 2003). Its approach is interpretive, seeking to understand the participants' meanings in context (Merriam, 2009). A range of data were collected through immersion of the first author in the classroom, taking field notes, videotaping nine sessions, collecting children's productions as drawings or classroom displays (see Table 1 below). For the purposes of this article, the six video recordings related to the snails project were selected, corresponding to numbers 1–6 in Table 1.

Participants

The participants are 25 children (16 girls/9 boys) in one class of the third year of Early Childhood Education (ECE3), from 5 to 6 years of age (mean age at the beginning of the snails project 5 years, 6 months, 23 days; standard deviation 3 m 1 d), in an urban state school, and their teacher. They have been with her since ECE1. In the class, there is only one child of immigrant origin, from North Africa, although she speaks fluent Spanish. The children's names are pseudonyms; the teacher is identified with her real name, Dolores Vázquez, because we author papers together. It must be noted that Galicia is a bilingual community where both Spanish and Galician (belonging to the Portuguese stem; as seen in our abstract) are co-official languages and consequently taught in school and used as vehicular languages for instruction. Due to their interlinguistic similarity, all speakers understand both, and children use one or the other in their talk, which here has been translated to English. The teacher belongs to *Torque*, a team of six kindergarten women teachers, a professional learning community meeting twice a month, and she is one of the two more experienced members, having taught for more than 2 decades. The less experienced teachers are enculturated in the profession through their participation in the group. Every year a science project is carried out in the six classes, in collaboration with the researchers, who attend the meetings and provide input, for instance about the science content. However, this is not an intervention study, as the teachers are responsible for the design of the project.

Context: the Snails Project

The teachers design a science project every year. Like the previous ones, the snails project implemented during the 2013–2014 school year has a flexible design, modified depending on children's questions and interests. It is framed in design principles parallel to Metz's (2011) principles for science in early primary, reworded to fit the context of the study.

(1) *Engagement in practices triggered by curiosity and question-based*: curiosity as the motor that interests children in science. Children's questions are driving the project.
(2) *Deep prolonged immersion in a problem*: the snails project spans 5 months, from mid-January to mid-June.
(3) *Rich domain knowledge entwined with building-knowledge practices*: biology issues explored during the project included the fact that snails are hermaphrodite, their body plan, their mouthparts (radula), and the process by which a broken shell is healed.

The teachers' approach in their own words is reproduced in Table S1. It involves the following: (1) *Motivating, eliciting children's interest*: introducing the phenomenon or living being to the class. In this project, bringing a box with garden snails (the big European snail *Helix aspersa*, reaching a length of 28–35 mm); (2) *Collecting children's ideas and producing questions*:

Table 1
Timeline of the snails project: sessions analyzed 1–6

Month	Experiences and Observations	Classroom Talk in Videotaped Sessions	Children's Products
January	Introducing snails		Three classroom displays: (1) driving questions; (2) what snails eat; (3) hypotheses experiment 1
	Experiment 1: smell		23 initial drawings of snails
			Six modeling clay models
			22 drawings experiment 1
February	Experiments 2, taste and 3, hearing	*Videotaped session 1*:	Two classroom displays: hypotheses for experiments 2 and 3
	Observation: snails' healing from salt	Snails are hermaphrodite	20 drawings mouthparts
	Observation: broken shell recovery	Snails' shells grow with them	23 drawings experiment 2
	Observation: color of snails' "poo" related to food	Broken shell and healing	23 drawings experiment 3
		Experiment 3: data, claims	
		Colored "poo"	
		First ideas about mouthparts	
		Meaning of hypotheses	
March	Experiment 4: "surfaces"	*Videotaped session 2*:	One classroom display: hypotheses experiment 4
	Observation: snails' tooth marks on food	Report: collected information	23 Drawings experiment 4
	Watching a video of snails' mouths	Snails' "tongue," radula	
	Giving flour to snails to see the radula	Snails' internal organs	
		Experiment 4: data, claims	
		Videotaped session 3:	
		Snails are mollusks	
		Snails' "tongue," "teeth"	
		New question: Why do little snails disappear from the box?	
		Revising 20 earlier drawings of mouths	
April	Weighing snails with scales	*Videotaped session 4*:	
	Using spring and digital scales,	Snails' weight	
		Problem of scale's sensitivity	
		Making predictions	
		Snails' mucus	

Month	Experiences and Observations	Classroom Talk in Videotaped Sessions	Children's Products
		Videotaped session April 28 (not about snails)	
May	Experiment 5, strength and 6, walking on tightrope	*Videotaped session 5*:	Two Classroom displays: hypotheses for experiments 5 and 6
	Observation of a limpet's radula with an e-amplifying lens	Experiment 5 and 6: predictions, data, claims	19 Drawings of experiments 5
	Experience 7A: forces	Functions of slime	20 Drawings of experiments 6
		Videotaped session 6:	
		Comparison of limpet' s radula with their ideas of snails' mouthparts	
		Explanation for how snails use their radulae	
		Videotaped session May 19	
		Forces (not about snails)	
June	Experience 7B: forces	*Videotaped session June 5*	17 Drawings of experience 7
	Bringing snails back to the garden	*Forces (not about snails)*	

producing responses to three questions, used for driving the project: (a) What do we know about snails? (b) What do we want to know about snails? (c) What did we learn? And (3) *Engaging children in scientific practices guided by teachers*: the starting point are the children's own questions.

The children are familiar with snails. The school is located in a small city with no clear limits with the countryside and many houses have gardens where snails are abundant (and a pest). They do know for instance that snails eat greens such as cabbage or collard leaves or that they can withdraw into their shells. Everyday routine included counting up snails and checking what they had eaten. Each day, a team of two children was in charge of cleaning up the terrarium and wetting the snails, so they would come out of their shells. Table 1 summarizes the project's timeline. The sustained participation created opportunities for a range of experiments, observations and children's productions. Attendance was not always constant, ranging from 25 to 19; they took their productions home and sometimes did not return them, accounting for the differences in numbers of drawings in Table 1.

Two experiments may illustrate the children's engagement with the project: in experiment 3, to test hearing, they placed four snails on a lid, and stayed in silence for a few minutes. Then they made noises by (1) shouting; (2) banging two sticks; and (3) playing a tambourine. The snails' behavior in the conditions of silence and noise were compared. Experiment 5, snails' strength, tested if snails would spontaneously move toward a piece of lettuce pulling a potato either through a plastic strainer or through a cardboard "cart" attached to the snail's shell (see photos in Figures S1 and S2).

Data Analysis

Data were examined through discourse analysis (Gee, 2005), and the use of the constant comparative method (Glaser & Strauss, 1967). The six sessions, comprising 5 hours (5:5:58) of video recording, were transcribed by the first author. The unit of analysis was the turn of speech, defined as each intervention by the participants in the talk; in cases when a single turn included different elements (e.g., claim, evidence), it was segmented into utterances, representing a unique contribution to the discussion. In order to construct a map of events for each session, turns were grouped into episodes, defined as one or several turns of speech related to the same topic or action. Transcriptions and children's productions were analyzed through prolonged immersion in the data of both authors, elaboration of initial repertoires of categories drawing from the literature and independently assigning a tentative code to each unit, comparing the codes and resolving the differences; and collapsing and refining the categories. Using these revised categories, data were subjected to several cycles of analysis.

Coding categories emerged from the interaction of dimensions from argumentation literature with data in successive iterations. The first level of analyses focuses on the identification of argument components: claim—evidence—justification (reasoning). We draw from research examining argumentation and evidence (Aikenhead, 2005; Duschl, 2008; McNeill & Krajcik, 2008). In this coding scheme, we also included a category for *raw data* that we define as description of first-hand observation, experiment or second-hand information, but unrelated to a claim or to a question. We believe it is relevant to distinguish between these and data that constitute evidence. In the context of kindergarten, we define evidence as data whose discursive role is to support a claim, in other words, which "count as" evidence in the students' discourse, and are appropriate. It needs to be noted that we do not include McNeill and Krajcik's (2008) "sufficient" condition in our coding scheme. This condition seems to be more difficult for primary students than appropriateness. For instance, in Gotwals et al. (2012) practice progression for fourth to sixth grades, levels 1 and 2 are defined by the use of "appropriate but insufficient (partial) evidence" (p.

187). For the purposes of identifying entry points for the use of evidence in kindergarten, we suggest leaving out sufficiency. Table 2 presents the specific codes for argument components, illustrated with examples from children's talk. Repetitions, which are frequent at this age, such as repeating the piece of evidence another child has offered, were not counted; in other words, we considered each different element only once in each argumentative episode, although they were counted when occurring in a different episode or session. Both authors coded all the transcripts, and the percent agreement for the last version of the codes was 82%. Disagreements were resolved through discussion.

The second level of analysis, related to question 1, focuses on the 57 argumentative components identified as evidence (see Table 4 in the findings). They were subjected to two further analyses: first, in order to identify the meaning, in particular the distinction found by McNeill (2011) among fifth graders' ideas of evidence in terms of *support an answer to a question*, between (1) having evidence to find something out, and (2) using evidence to support a claim. It needs to be noted that McNeill's findings are based on interview responses to direct questions about the meaning of evidence, while our study examines the practice of using evidence. The criterion used to distribute the occurrence of evidence in these two categories was the discursive context: answering a question or evaluating (supporting or rebutting) a claim, taking into account that they involve overlapping. Second, we analyzed the 57 evidence statements seeking to identify different stages or levels in the development of data into evidence. The coding scheme draws from Aikenhead (2005) and from Duschl's (2008) first critical transformation in the E-E continuum. Similarly to the process described above, it was developed in successive iterations. It distinguishes two levels of epistemic judgment: (L1) statements closer to data; (L2) statements involving evaluative judgments, meeting one of these criteria: (1) identifying patterns in data; (2) connecting data and claim through justification; (3) establishing comparison with other data; (4) explicitly evaluating one or several alternative claims. Criteria (1) and (3) are drawn from Aikenhead (2005). It needs to be noted that although for analytic purposes, we distinguish two levels, they should be seen as a continuum. The percent agreement was 86%, and disagreements were resolved through discussion. Table 3 presents the levels, their definitions, and examples from children's talk.

For the purposes of this article, the analysis of children's drawings focused (1) on written texts within the drawings, for question 1; and (2) on conceptual content analysis, for question 3. An analysis of children's drawings as semiotic resources (Jewitt & Oyama, 2001) is carried out elsewhere. For question 1, the focus was on the use of argumentative connectors in the written

Table 2
Argumentation coding scheme

Code	Description	Student Examples
Claim	A statement or conclusion that answers the original question	"If a snail grows, its shell grows with it"
Raw data	Description of first-hand observation, experiment or secondhand information unrelated to a claim or question	"(baby snails are) very soft"
Evidence	Data used to support a claim, judged as significant ("count" for evaluating the claim), and appropriate	"All of them can lay eggs" (no distinction female/male; supporting the claim "snails are hermaphrodite")
Justification (reasoning)	Connects the evidence to the claim	"if they do not hide it is because they do not hear"

Table 3
Evidence coding scheme

Code	Description	Children Examples
Level 1	Descriptive evidence statements closer to data	"(snails have) heart, brain" (claim: It is almost like us)
Level 2	Evidence statements involving evaluative judgments, meeting one of these criteria:	
	– (a) identifying patterns in data	"Poo of the color of what they eat"
	– (b) connecting data and claim through justification;	"If it is not rough, it doesn't scrape"
	– (c) establishing comparison with other data;	"It [the snail] does not eat bit after bit, it makes little holes"
	– (d) explicitly evaluating one or several alternative claims	"No, they are snails [not excrements], they have little horns"

texts within the drawings of experiments, under the printed heading "conclusions." Argumentative connectors are defined by Ducrot (1983) as signs that link two or several statements, assigning them a particular role in the argumentative discourse. In this case, we examined the use of the connector *because* to link evidence and conclusions in drawings of experiments 1–5, selected because they share a common structure (see Table 6 in the findings). For question 3, the analysis of children's drawings of the snails' "mouths" focused on its representational meaning, drawing from Kress and van Leeuwen's (2006) semiotic functions. Both authors analyzed the drawings and the agreement was 100%.

For question 2, on ways of gathering evidence, the analysis draws on the literature about first- and second-hand data (e.g., Hug & McNeill, 2008), and about the relevance of observation, discussed in the rationale. We propose calling purposeful observation one that is prolonged, systematic and with a clear focus. The coding scheme distinguishes three types of questions addressed in this classroom: (1) empirical first-hand data gathered through purposeful observation; (2) empirical first-hand data gathered through experiments; and (3) second-hand data found on the Internet, books, or from families. First, the 20 questions produced by children, prompted by the teacher's question "What do we want to know about snails?" and collected in a classroom display were independently analyzed by both authors according to the type of evidence required to answer them, and the agreement was 100% (see Table 5 in the findings about question 1). Second, in order to answer questions 2 and 3, we conducted a thematic analysis (discussed in another paper) drawing from multiple data sources, classroom transcripts, children's drawings, and classroom displays. Twenty-one conceptual topics were identified, which were subsequently grouped into 13 conceptual issues distributed in four broad themes, respectively: a snail's body; a snail's biology; the classification of snails; and its skills and behavior. In addition, talk about knowledge, hypotheses, or prediction was coded separately. For question 2, the 21 topics were examined in terms of the evidence used to address them: purposeful observation, experiments, or second-hand data (or combinations of these). An independent review by both authors yielded a 96.8% agreement. From the 10 topics addressed through purposeful observation, three related to processes were selected to illustrate the findings.

For question 3, *the use of evidence to evaluate and revise knowledge*, we reviewed eight of the 21 conceptual topics identified in the thematic analysis that were recurring through two, three, or

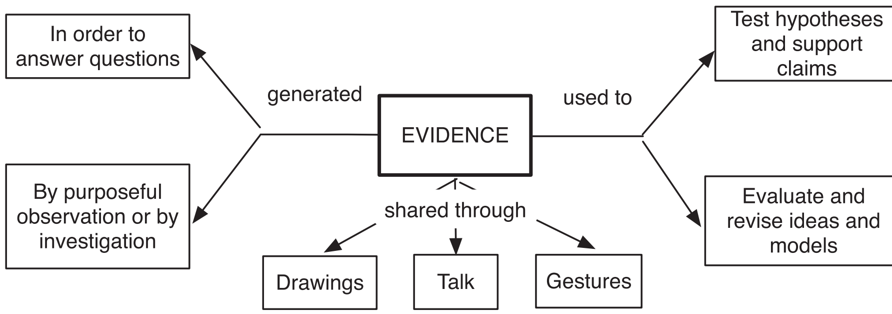

Figure 1. Meanings for evidence and uses of evidence in the classroom.

five sessions. From these eight, the question of mouthparts, discussed in five out of the six sessions, was selected to illustrate the process.

Findings: Evidence in a Kindergarten Classroom

The three research questions (RQ) are interconnected, as summarized in Figure 1: RQ1 examines the meanings for evidence in this classroom, represented in the upper part of the figure (on the one hand, evidence is generated in order to answers questions, in other words, in the context of seeking how to find out about questions posed by children, and on the other, it is used to test hypotheses and support claims), where generation of evidence and its use are interconnected; RQ2 examines ways of gathering empirical, first-hand evidence, which in this classroom are either through purposeful observation or through investigation; and RQ3 examines how children use evidence to evaluate and revise ideas and models. Evidence is shared through multiple modes of communication, drawings, talk, and gestures.

The Meaning of Evidence in Kindergarten

Question 1 examines the meanings for what constitutes evidence constructed by children, and how are they reflected in the development of data into evidence. In order to frame the findings, we first present the overall quantitative results of the analysis of classroom talk. From the 937 turns of speech in the six sessions, 276 were identified as argumentative talk. Their distribution in argumentative components is summarized in Table 4. The purpose is not a quantitative analysis, but rather to give a sense of how often they occurred. What becomes apparent first of all is that there are more claims than evidence; so, similarly to previous studies (e.g., Jiménez-Aleixandre et al., 2000), many claims are not supported by evidence. Secondly, not all data are developed into evidence, as shown by 30.8% of the statements we coded under raw data.

Evidence is Generated in Order to Answer Questions. This meaning for evidence is supported by the following: (1) children's initial questions and those addressed during the project and (2) the

Table 4
Argument components

Components	Claim	Raw Data	Evidence	Justification
N = 276	125 (45.3%)	85 (30.8%)	57 (20.6%)	9 (3.3%)

Table 5
Types of initial questions: In italics, questions addressed in the project

Type: Can Be Answered by	Children's Initial Questions $N=20$
1. Empirical first-hand data gathered through purposeful observation: $N=8$	– *Do they have mouth?* – *Do they have teeth?* – The little things under their head: are they legs? – *What are the tentacles for?* – Are snails born with shell? – *What do they need the shell for?* – *Where does snail slime come from?* – Why are eggshells put there [in the box]?
2. Empirical first-hand data gathered through an investigation or experiment: $N=4$	– *Can they hear?* [later expanded to: *Can they smell? Can they taste? Do they have a sense of touch?*] – What is the shell made of? – *Can they live if their shell is broken?*[a] – *Is it slime what sticks the shell to the body?* [later modified as: *What are the functions of slime?*]
3. Second-hand data, found on the Internet, in books or from family: $N=8$	– *How do we take care of them?* – *Where do they live?* – How many types of snails are there? – Whether or not there are underground snails – *Where is the penis?* – Do they need calcium? – *What do sea snails eat?* – Why do they need water and sunlight?

[a]it could be answered by an experiment, but involving harm to snails.

context for the 57 statements coded as evidence; as well as by explicit talk about evidence, claims, and testing, for evidence and testing are linked to particular questions.

First, the project was organized around driving questions suggested by children. Formulating empirically answerable questions about phenomena is a basic scientific practice (NRC, 2012). In the NGSS (Achieve, 2013), the goals for K-2 include (1) asking questions based on observations to find more information about the natural world; and (2) asking or identifying questions that can be answered by an investigation. The teachers from Torque begin all science projects eliciting children's questions, by asking "What do we want to know about ... (snails) ?" (see Table S1). This happened in the last week of January when the snails were brought to the class. Children's questions collected in the classroom display are translated in Table 5, distributed in three types according to how can they be answered—an issue discussed with research question 2—by empirical first-hand data gathered either through purposeful observation or through experiments, or by seeking second-hand information. The questions range from features of a snail's body to a snail's biology or well-being. Some questions seem grounded on previous knowledge, for instance the need for calcium.

In Table 5, the questions subsequently examined during the project are italicized. All eight questions that can be answered by purposeful observation were addressed, although in many cases they were modified. From the four questions that can be answered by an investigation, there is one ("Can they live if the shell is broken?") that was not the object of an experiment, for breaking a shell on purpose was out of the question. Its examination through prolonged observation was made possible by an accident. "Can they hear?" was expanded into three more questions about the senses, explored through experiments 1 (smell), 2 (taste), 3 (hearing), and 4 (walking over

Table 6
Frequency of use of the connector because in the drawings: experiments 1–5

Experiment	1. Smell $N = 22$	2. Taste $N = 23$	3. Hearing $N = 23$	4. Surfaces $N = 23$	5. Strength $N = 19$
Use of *because* # Because / # drawings: 65/110	16/22	10/23	22/23	1/23	16/19

different surfaces). Two more questions emerged about snails' capacities: their strength (experiment 5) and their ability to balance on narrow objects (experiment 6). For each question in this category, children were asked to suggest experiments that were planned with strong input from the teacher. Then they were asked to generate hypotheses contextualized in the experiment (but before carrying them out). Each question was dealt with in one or several days, sometimes over different months, as seen in the timeline in Table 1. From the eight questions that required information from secondary sources, only four were addressed in the project. This emphasis on certain types of questions shows the teacher's focus on first-hand data and direct evidence.

Secondly, for the 57 statements coded as evidence, in terms of the distinction reported by McNeill (2011), we examined the context of the argumentative episodes when these evidence statements were produced: (1) In 11 cases, the meaning was *having evidence to find out something*, such as "I played the tambourine and they didn't hide" (finding out if snails can hear); (2) in 24 cases, the meaning was *using evidence to support a claim*, such as "they have little horns [tentacles]" (backing up the claim that tiny things in the box were baby snails not excrements); (3) in 22 cases, both meanings overlapped, as in "[snails have] a heart, a brain" (finding out which internal organs snails possess, and also supporting the claim that they are "almost like us").

Explicit talk about evidence, claims, and testing also supports this meaning, as testing and hypotheses are connected to questions. As a summary, it may be said that a dimension of the meaning of evidence in this classroom is being generated in order to answer questions (see Figure 1).

Evidence is Used to Test Hypotheses and Support Claims. This meaning for evidence is supported by the following: (1) children's interpretation of the results of experiments; (2) how evidence is connected to conclusions through justifications, and (3) the occurrence of explicit talk about evidence, claims and testing.

First, on how children interpret the results of experiments: after a classroom discussion, they were asked to produce individual drawings for each experiment and to write the conclusion in their own words in a template with "conclusion" printed at the bottom. The examination of drawings focuses on experiments 1–5, sharing common features, namely that they convey narrative meanings. In their discussion of visual designs as semiotic resources, Kress and van Leeuwen (2006) distinguish two types of representational meanings: narrative and conceptual. Narrative representations are characterized by designing actions, and by the presence of a vector. For instance, in Alberto's drawing, reproduced in Figure S3, a story unfolds horizontally, and all participants are represented as doing something, children are shouting or playing the tambourine, snails are out of their shells. A detailed analysis of children's drawings through this social semiotic lens exceeds the purposes of this article and is carried out elsewhere. Children's writing skills were uneven, and the teacher's help was needed to interpret the texts. We examine the use of the lexical connector *because* (Ducrot, 1983) to connect evidence and conclusions. Some drawings were missing from the portfolios, either because children missed the session or because they were not returned when they took the portfolios home to share them. Table 6 shows the frequencies of use in each experiment.

From the 16 children who handed in the five drawings, five used *because* in four cases, seven in three, two in two and two in one drawing. A similar trend appears in the cases of children handing in fewer drawings (see Table S2). Two examples of conclusions with a connector (children's spelling mirrored in translation): Carmen: "snails can smell because we put binegar [*sic*] and water ant they went to water" (experiment 1). Alberto: "s[n]ails donthear b[e]c[au]se we s[ho]uted andthey didnot [h]ide we play[ed]thetamb[our]ine and they didnot hi[d]e)" (experiment 3, see drawing in Figure S3). One example without a connector: Anton, "sna[ils] don['t] he[ar] s[ho]uted and din['t] hi[de] (experiment 3).

Although over half of the drawings (59%) show the use of *because* to link claim and evidence, its use is uneven across experiments. The weakest results are found in experiment 4 on the question, "Are snails able to walk over all surfaces?" We interpret this as a consequence of the different nature of data in the experiments. In this case, most children provided a general claim "Snails are able to walk over all surfaces." Only one of them worded it in connection with data: "beqa[u]se they w[e]alk overrr all of them, eve[n] over the pins." The use of *because* to connect data and claims is also found in the transcriptions, as seen in the excerpts about "poo" and experiment 3 reproduced below. This presence alone does not mean that children fully understand the role of evidence in supporting or falsifying claims; we consider it an indicator connected to others.

Secondly, connection between evidence and conclusions through explicit justifications occurred in nine cases (see Table 4). Eight out of the 25 children offered them. It is a low number of cases, but it should be noted that this is one of the most difficult aspects of the use of evidence for older students (e.g., Gotwals & Songer, 2013). This is illustrated by two examples, one from experiment 3 and one from observation:

Marta: They don't hear, because we shouted, and we banged sticks, but they didn't hide [in the shells]. And if they don't hide it is because they don't hear.

Lua: I played the tambourine and they didn't hide

Several children: They are deaf!

Marta uses evidence from the experiment to support her claim (snails do not hear), and she makes her argument stronger by appealing to a justification that connects evidence and claim ("if they don't hide it is because they don't hear"). This justification is implicitly grounded in previous knowledge: firstly, that "snails are timid and withdraw into the shell," a statement from the "What do we know?" display; secondly, that danger or threats cause snails to retreat into their shells. This second piece of knowledge is related to an accident discussed in the next section. In other instances, justifications draw from theory, as in the case of shell regeneration. Marta's argument is represented in Toulmin's format in Figure 2.

Another example, after directly observing a radula (snails' mouth part) in session 6, and combining this observation with previous data, is related to the question about how the radula works on food, discussed in detail in the section on RQ3:

Danilo: I think that [it does] like this [he mimics with his hands a back and forth movement]

Ester: [They] scrape food

Alberto: It [the radula] has to be rough in order to scrape. Because if it is not rough it doesn't, it doesn't scrape

The children cooperate in the argument, they have been discussing other data, such as the little spikes over the radula surface, or the marks observed on food. Danilo mimics a movement

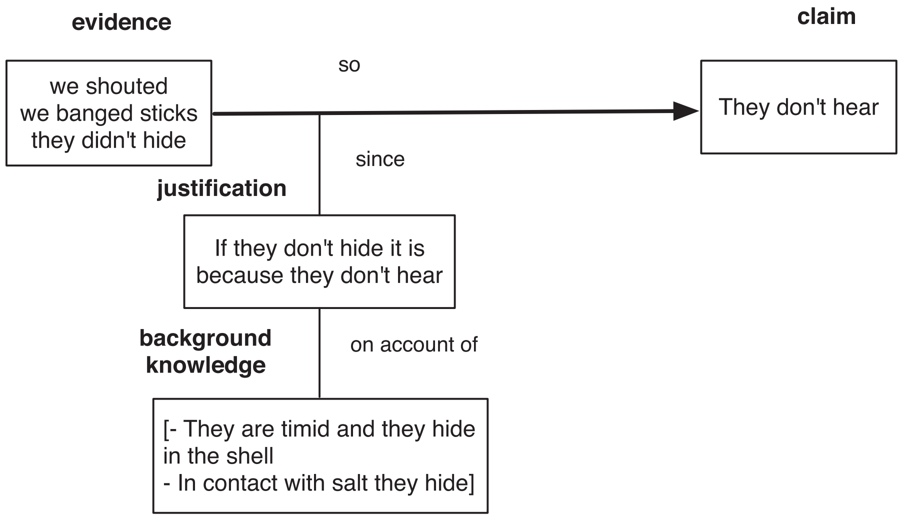

Figure 2. Marta's argument about hearing. Implicit knowledge between brackets.

they saw in a video, a non-verbal form of communication used in three sessions, and Ester offers a claim about how the radula works. Alberto connects this claim to previous evidence through a justification based on knowledge about materials or tools that scrape, such as sandpaper.

Thirdly, explicit talk about evidence, claims, and testing occurred in all the sessions. It must be noted that the terms used in Spanish (S) and Galician (G) for *evidence* pose challenges for translation: in both languages *evidencia* means something that does not need to be tested or proved; the terms used in this classroom and in argumentation literature, to refer to it are *prueba* (S) and *proba* (G), and for testing, the terms are *probar* or *comprobar* in both languages. Furthermore, in both languages, there is only one word, *investigar*, for research and investigate. There were 15 episodes of such explicit talk across the six sessions, 11 initiated by the teacher, for example, prompting them to talk about hypotheses or recalling that in order to know something, an experiment is needed, as illustrated below in two examples; one was initiated by the researcher, as seen in the excerpt about the color of excrements in the next section, and three initiated by children.

In this classroom, claims are not accepted unless there is evidence to support them, as illustrated by this discussion about young snails and adults in session 2.

Carmen: Some [snails] go faster than others.

Teacher: Which ones?

Carmen: The little ones.

Unidentified child: Smaller ones go faster because they weigh less.

Teacher: But we don't know whether this is true or not, we would need an experiment.

In classroom talk, there is a distinction between hypotheses, before carrying out the experiments, and conclusions, after the experiments. Children appropriate the terms that had been introduced by the teacher.

In summary, the context for the generation of evidence was seeking how to answer children's own questions, with a focus on direct, first-hand evidence. This is what we interpret as a meaning for evidence being generated in order to answer questions. We think it shows the connection between the practices of asking questions and engaging in argument from evidence. This evidence is used to test hypotheses and support claims, as seen in actions and in both written and spoken speech.

How Are These Meanings Reflected in the Development of Data Into Evidence?. Data are not always developed into evidence, as shown by the 85 statements coded as raw data (see Table 4). The examination of the 57 cases when it does points to differences in epistemic judgment, to different levels in the transformation of raw data into evidence (Duschl, 2008). We distinguish two levels, L1 closer to data, and L2 involving evaluative judgments, either by identifying patterns, using justifications or explicitly comparing data or claims. Table 7 summarizes the distribution of evidence statements in the two levels and across the three different sources of evidence (see the graphic representation in Figure S4). As the table shows, 63.2% of the statements are coded in the lower level, in many cases corresponding to support for claims related to features, such as color, size, or categories, for example, snails are mollusks, and 36.8% are coded in the higher level, in many cases corresponding to support for functions, such as eating, or processes, such as shell regeneration. Higher level evaluative statements are distributed across the three types of sources of evidence, although it needs to be noted that the proportion L2 versus L1 is higher for experimental sources, an issue addressed in the discussion. Examples are discussed in the final findings section.

Evidence Can Be Gathered Either by Investigation or by Purposeful Observation

The second research question examines which ways of gathering empirical evidence are jointly constructed by children and teacher during the project. Questions such as whether or not snails can hear or smell can be answered by an investigation or experiment. There are other questions such as "Are snails born with a shell?" that refer to *processes*, so the best way to answer them is to watch the snails for a sustained period. The question "Can they live if their shell is broken?" leads to observation until the shell had grown again. On the other hand, questions such as "Do they have a mouth?" or "Do they have teeth?" look as if they could be answered by a simple yes or no. However, the way of addressing them in the classroom changed them to "What are snails' mouths like? How do they work?" examined through observation, as discussed below.

We call *purposeful observation* that which takes place in this classroom. We use it to refer to prolonged observation that had a particular focus, was guided by the teacher, discussed and used to test claims and to compare initial models with later ones. The notion of active purposeful observation draws from dialogic teaching (Alexander, 2008), medical training (Morris, 2007), and educational research (Merriam, 2009). Our suggestion is that it can be extended to children's observation of beings or phenomena.

It should be noted that both experiments and purposeful observation generate empirical first-hand data, in contrast to second-hand data acquired from other sources. In this classroom, pupils and teacher referred to these two ways of generating evidence by different names: "to investigate," for short-term planned experiments, and "to discover," for purposeful observation (although it

Table 7
Levels in the development of data into evidence across sources of evidence

Levels/Source of Evidence	Experiment	Observation	Secondary	Total (%)
L2 Higher—Evaluative	6	7	8	21 (36.8%)
L1 Lower	3	23	10	36 (63.2%)
Total $N = 57$	9	30	18	

does not mean that observations were non-planned). This is a distinction introduced by the teacher.

Twenty-one conceptual topics identified in a thematic analysis were coded in terms of the evidence used to address them in the classroom: purposeful observation, experiments, or second-hand data (or combinations of these). From the 10 topics addressed through purposeful observation, three instances related to processes were selected to illustrate it, beginning with the regeneration of a broken shell.

Teacher: What happened that day a boy grabbed a snail and then you were sad?

Pupils (several): The shell broke.

Teacher: And what did we think?

Pupils: That it would die.

Teacher: And what did we discover?

Pupils: That the shell grew again!

Ester: That with the eggshell the calcium is put in the shell. Because it has calcium.

Hector: Because they eat it [*the eggshell*].

Marta: Yes, and then it is not smashed because it is tougher.

This is an example of how evidence is gathered through observation, and of a collaborative effort of the learning community in analyzing and interpreting data, resulting in the claim that the shell grew. The initial question, framed in an opposition between being alive or dead and worded as a yes/no issue, is transformed into the examination of the process of shell regeneration. Ester adds a justification (called reasoning by McNeill & Krajcik, 2008) about the role of calcium in this process, and Marta complements it with a statement about the effect of calcium in a shell's toughness, grounded on scientific knowledge; this connects the process to two other initial questions about calcium and eggshells. As this example illustrates, data come from different sources. The word *calcium* was introduced by a child who brought it from home, after a web search at the beginning of the project. This word was given a meaning through its *function*: calcium is a component of structures such as eggshells and it makes snails' shells tougher. The prolonged observation of this recovery provided opportunities for an initial contact with the crosscutting concept of stability and change. As the NRC (2012) acknowledges, larger time scales are needed in order to observe changes.

In this excerpt Carmen is reporting about experiment 2, to test if snails can taste.

Carmen: [*we carried out*] An experiment with flour and salt. And we placed three snails in the middle [*amid flour and salt*] and they went towards the flour and then they ate it. But someone let a snail fall into the salt and we didn't know whether it was dead or not.

Researcher: And what happened?

Carmen: It was foaming! [...]

Elena: Silvia healed it. Silvia is a girl who poured a lot of water over it [...] First, she cleaned it very well and then we put it in a paper.

Roberto: Then she poured so much water that it stopped foaming.

Elena: And she cleaned all that slime... and then we were worried about if it was dead or not. And Dolores [*the teacher*] marked it with a red cross like in hospitals, and then next day we discovered that it was alive.

Researcher: How did you know that it was alive?

Elena: Because we saw it! [...] It was hidden in the shell.

Researcher: And did it come out of the shell?

Elena: Yeees!

That unexpected event drove pupils to gather evidence through observation to test whether the snail would survive. As in the instance of the broken shell, Carmen and Elena switched from considering just two different extreme states, that is, dead or alive, to take into account the healing process.

Observation in some cases leads to experiments. Children observed that the color of excrements was related to the color of food. Then they experimented giving snails food of a single color, recorded the outcomes, described as "we investigated it," and identified a pattern, as shown in Ester's generalization:

Ester: With [*eating*] carrot they poo orange, because carrots are orange.

Researcher: And if they eat lettuce?

Several children: Green!

Ester: Poo of the color of what they eat.

Researcher: And what did you do to know that?

Several children: We investigated.

Ester: [*We saw*] Poo of different colors and then we investigated it.

Gathering evidence in this classroom has two complementary meanings, doing experiments, "to investigate," and carrying out purposeful observation, "to discover" (see Figure 1). In the next section, we discuss the role of purposeful observation in evaluating ideas.

Evidence From Purposeful Observation is Used to Evaluate and Revise Ideas

The third research question examines how children use evidence to revise their understandings. The examination of how they used data from purposeful observation to evaluate or revise their emerging models about snails is framed in an approach that considers the articulation of practices with core ideas. In the thematic analysis, 21 conceptual topics were identified. Because of the teacher's focus on continuity, eight of them recurred over two, three or five sessions, for instance (1) parts and features of a snail's body (theme 1), such as their two pairs of tentacles, one of them carrying the eyes, their ribbon-like mouthpart ("radula"), their feet and shell; (2) a snail's biology (theme 2), such as what they eat, that they are hermaphrodite and lay eggs, how they are born and develop, the function of slime in a snail's movement, or the fact that shells grow with the snails. In seven cases, these recurring topics were explored through a combination of purposeful observation, experiments, and second-hand information. Purposeful observation was a driving force in the revision of ideas. Certainly, mere observation does not produce conceptual change. The way the teacher scaffolds purposeful observation is illustrated with the process of how children revise their ideas about snails "teeth" and "tongue," which recurred through five videotaped sessions. Table 8 summarizes its timeline.

In the last week of January, after they have observed the snails for 2 weeks, but not gathered data or information about their organs, the teacher asks them to draw "what they thought was inside the snail's mouths." Twenty drawings are returned, all of them representing the mouth as a semi-elliptical shape, like a human tongue, and 10 of them with teeth around it or at the end. Eighteen labeled it "tongue,“ a term used to refer to mouthparts that can be projected outside. As Inagaki and Hatano (2006) acknowledge, human-based inferences or person analogies are useful for biological understanding, and should be viewed positively, as reflecting a child's adaptive mind. Figure S5 reproduces a representative drawing. Unlike drawings about experiments discussed above, these convey conceptual meanings (Kress & van Leeuwen, 2006).

While conducting experiment 2, the teacher prompts the children to observe the snails feeding on flour. Daily observation also generates a discussion on February 24 about the deep holes ("tunnels") in food.

Ester: They don't eat carrots like we do.

Pupils (*several, talking at the same time*): They make little holes! / Yes, they do. / Why do they make holes? / Because they have teeth.

Table 8
Timeline of the process of revision of ideas about snails' mouthparts

Date	Observation, Experiences, Drawings, Talk, Gestures	Children's Ideas and Appeals to Evidence
January, last week	– Initial drawings of their ideas of snails' "inside of the mouth"	– Snails' "tongue" similar to human tongue with "teeth" around
February, 17	– Initial observation of "tongue" during experiment 2	– A snail's "tongue" is very small
24	– Observation of marks in food	– A snail's teeth are different from ours
	– Talk: holes in carrots	– Tooth marks as evidence of existence of teeth
March, 17	– Second-hand information: radula	– Radula is like a ribbon with spikes
21	– Watching video of snail eating	– Radula is like a saw
24	– Experience with flour, photos	– Spikes over radula not around it
	– Children mimic radula movements with their tongues	– Snails scrape off, they cannot chew (spikes shape as evidence)
	– Critical revision of initial drawings	
May, 12	– Observation of a limpet's radula through the digital microscope	– Radula is coiled – Analogy with zip fastener – Mechanism: holes as evidence of spinning movement of radula
	– Talk about mouthparts: analogies, names, movement	

Teacher: We need to study what their teeth are like. Because they are not like ours: Are they?

Elena: Oh my, if they have them, we don't know that yet.

Teacher: True, we don't know that yet.

Alberto: They are smaller.

Teacher: That is what Alberto imagines. We will need to test it.

Marta: I think they do have teeth, because otherwise they could not make these tunnels.

Several scientific practices are enacted in combination. Children interpret data, holes in food and glimpses of mouthparts, to construct their explanations, communicated through class talk: snails have teeth, but they are not like ours, they are smaller because they make little holes. Data are based on observation, in particular, indirect data, through the tooth marks, used as evidence of the existence of teeth, in a way similar to their use by biologists to identify animals.

In the second week of March, the teacher asks them to collect information about snails' mouths at home. On March 17, the children share this second-hand information from web searches, such as the term *radula*, its ribbon shape and the little spikes on it.

Teacher: Let's see, Luis.

Luis: It is shaped like a ribbon.

Teacher: It is shaped like a ribbon, and: What are the little spikes for?

Luis: To scrap off food.

Teacher: To scrap off food. Then: Does it have teeth?

Luis: No, it has little spikes.

Children revise their ideas, discarding anthropomorphic names: the chitinous spikes are no longer "teeth," radula is the name of the "tongue" (although this term continues to be used), and snails do not chew, but rather scrape off food. Differentiation among similar structures is a step in the construction of science concepts. They use analogies, like ribbon, to share information. First-hand data are combined with second-hand data, as on March 21st when they watch a YouTube video of a snail feeding, in which, due to magnifying and good lighting, the radula can be clearly observed.

A critical revision of their previous models occurs on March 24th. The teacher asks each child to discuss her or his drawing, and to explain why they had drawn it like that. Fourteen out of 20 justify their initial drawings saying that they thought it was “like ours.“ These models are then compared with new data from observation and the video.

Marta: They [*the spikes*] had hooks [*shape*]

Roberto: We were impressed [*by the video*]!

Teacher: How did it work?

[*Children mimic radula’s movements sticking their tongues in and out*]

Teacher: Were any of us right when we imagined what the radula was like?

Children: Nooo!

Teacher: And now: How would you draw it?

Children (several): Shaped like a ribbon.

Teacher: Does it have anything around it?

Children: No.

The teacher prompts an explicit comparison of observations with their previous ideas, both in general terms, and in specific issues, such as where the spikes are placed. Children communicate new knowledge through multimodal discourse, for instance mimicking the movement of a radula with their tongues several times during this session.

On May 12th, the researcher brings in the radula of a limpet, and they have the opportunity of directly observing it with the digital stereomicroscope. A new revision of their ideas takes place, focusing on the new notions and their connection with evidence:

Teacher: What did snails do to food?

Children (several): Little holes

Teacher: Little holes. So we said that the “tongue” would need to have.

Children (several): Spikes.

Teacher: And: What would they do in order to make holes?

Ester: They would stretch it, pick up food, and withdraw it into the mouth.

Marta: True, like butterflies.

[...]

Alberto: Certainly while the radula is spinning it is digging because it makes deep holes.

The revision of ideas in this session includes the shape of the radula and spikes, revised with data from direct observation that are compared to second-hand data, and its movements, evidenced by the deep holes they have been observing throughout the project. The teacher prompts them to propose explanations about how this mouthpart with its tiny spikes would be able to make holes. Several children make proposals or analogies, like Marta who compares Ester’s explanation to butterflies, or a zip fastener. It is noteworthy that Alberto proposes a mechanism that accounts for the deep holes or “tunnels” observed in food. Explanations that include mechanisms are more challenging for students to construct. This lesson was one of the only times during the snails project in which pupils produced mechanistic explanations. Figure 3 summarizes the revision of their models about mouthparts. This supports the role of purposeful observation in the evaluation of their ideas (see Figure 1).

Discussion: Purposeful Prolonged Observation and Its Role in Kindergarten

This classroom study seeks to shed light on the meanings kindergarten children construct about evidence and about how they use it in the context of a project about snails spanning 5 months. The meanings are more explicit, while the uses are examined through the enactment of practices.

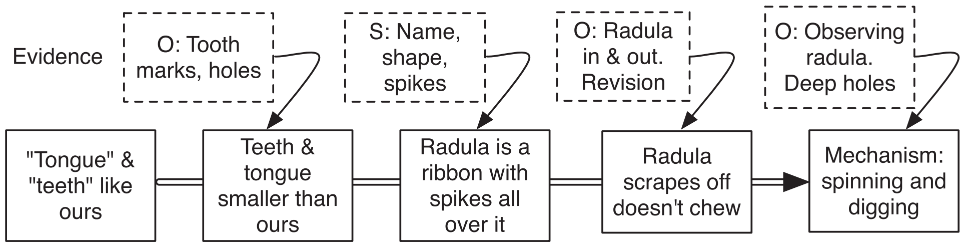

Figure 3. Evolution of children's ideas about mouthparts across the project. Observation (O), Second-hand evidence (S).

Since this is a case study, there are limitations: for example, we are unable to generalize our findings; however, a number of important issues do emerge from this work.

First of all, among the meanings of evidence in this classroom, we think importance should be placed on the difference established in teacher and children's discourse between evidence gathered through "investigating" (evidence from planned experiments) and through "discovering" (evidence from prolonged observation). We call the second process *purposeful observation*, drawing from Alexander's (2008) notion of purpose in dialogic teaching, and from Morris (2007) and Merriam (2009), and define it as prolonged systematic observation that has a clear focus, is guided by the teacher, recorded, explicitly discussed, and used to test claims and revise initial models. It must be noted that the evidence collected in both cases is empirical first-hand data, which according to previous studies (Delen & Krajcik, in press; Hug & McNeill, 2008) evokes in pupils a higher sense of ownership. In the project, this first-hand evidence, from experiments or observation, is combined with second-hand evidence from web searches or family knowledge.

An indication of the significance of purposeful observation in this kindergarten classroom is that over half of the evidence statements, 30 out of 57, correspond to the context of purposeful observation. Because this is the first study focusing on it, it is difficult to ascertain whether this frequency is related to students' age and developmental reasons or to the particular context of this project and the teacher's approach. We suggest that these findings point to the interest of paying attention to promoting systematic, prolonged observation as a context for constructing empirical evidence in early ages.

Secondly, with regard to how these young children use evidence, one relevant finding is the role of evidence from purposeful observation in the *evaluation and revision* of their ideas about snails. An instance of how initial models are contrasted with evidence is the evolution of children's ideas about snails' mouths. As Gelman and Brenneman (2012) point out, through systematic observation children come to think differently about what they are observing. Thus, kindergarteners were able not only to revise their ideas about the form and other external features of the radula, but also to propose a mechanism, an explanation to account for the tooth marks observed in food. We believe that the notion of purposeful observation and its role in the revision of ideas, are new and they are an original contribution of our study. It is known that long-term projects provide opportunities to build understandings (Gelman & Brenneman, 2012), and that epistemic and social elements are most effectively incorporated as part of extended sequences of instruction (Duschl, 2008). What our study adds is a characterization of the features of observation brought by this extended time. Purposeful observation over an extended period enabled the study of *processes*: first, it enabled the children's exploration of biological processes, such as the development of newborn snails, healing from contact with salt or the regeneration of a broken shell. Second, it enabled the researchers' examination of learning processes, such as the evolution of ideas across several months, rather than only the difference between initial and final ideas (products). Our interest lies not only in what children can learn, but also how we can characterize learning environments and strategies that support learning.

Third, in reference to the *development of data into evidence*, our findings point to the distinction between two levels or stages in the transformation of raw data into evidence (Aikenhead, 2005; Duschl, 2008). First, we suggest that studies about early childhood and primary schooling should identify descriptive statements or raw data, alongside argumentative components, such as evidence, in order to better document how the transition from data to evidence occurs. Second, the identification of these two levels may have potential interest for argumentation progressions, and in particular for entry points in kindergarten. In their work about practice progressions for evidence-based explanations beginning in fourth grade, Gotwals et al. (2012) place in level 1 "student makes a claim," with two sub-levels, with and without scaffolding; and in level 2 "student makes a claim and backs it up with appropriate but insufficient (partial) evidence," also with two sub-levels. Gotwals and Songer (2013) identify levels 1–3 with scaffolded practices, beginning with a question provided to students. We suggest that, in kindergarten and early primary, two levels that would be previous to those from Gotwals and colleagues, or overlapping with them, could be the following: (1) selecting data appropriate for being transformed into evidence related to a claim; and (2) identifying potential (appropriate) evidence that could confirm or disconfirm a claim. Both processes would be scaffolded, as is the case in our study.

As noted above, 30 out of 57 evidence statements identified in the study are related to purposeful observation. However, only seven of them correspond to the higher evaluative level, L2. On the other hand, six out of nine statements related to experiments correspond to L1. More studies are needed in order to ascertain if this is a pattern, or if the level depends on other features of the setting. We suggest that experiments provide a frame where the relations between claim and evidence are more explicit and clear-cut from the beginning. In the case of purposeful observation, the claim may be derived from evidence, emerge later in the process, and the relations may be more diffuse. If this is so, the implication is a need for framing purposeful observation in the process of constructing evidence-based explanations more explicitly. It should be noted that our suggestion is to combine experiments and purposeful observation, not to focus only on the second.

Teachers' scaffolding is essential in supporting children's engagement in scientific practices, and in particular, in the processes from evidence to explanation. Although the focus of this article is not on the teacher's support, discussed in another paper, three interconnected teaching strategies identified as relevant for supporting engagement in scientific practices and in using evidence are as follows: (1) *Reflection*: one feature of the teacher's approach is to provide children with many opportunities to think back about their observations and experiences, to talk about them and to reformulate their meaning. Engaging students in discussions about their observations is a feature highlighted by Zangori, Forbes, and Biggers (2013). In our study, the time devoted to these discussions and reflections was substantially longer than the actual time devoted to carrying out the experiment or the observation; (2) *Recurrence*: connected to reflection is the recurrence, over all the sessions, of a few questions and topics that are addressed again, in the light of new evidence either from experiments, second-hand sources or, in particular, from purposeful observation. This recurrence, illustrated in the findings with the revision of children's ideas and drawings of mouthparts, provides *continuity* through the project and may have an influence similar to the effect of science journals reported by Gelman and Brenneman (2012), "(to) solidify their understandings because they provide a chance for learners to think again about a science experience" (ibid, p. 166). Mere observation does not lead to change, unless there is reflection about data, theoretical claims and their connections. As Lehrer and Schauble (2012) point out, the notion of practice implies engagement with the epistemic culture of modeling (and, we will add, of argumentation); (3) *Explicit talk* about evidence, hypotheses, claims, and testing: the teacher initiated 11 out of the 15 episodes of such explicit talk across the six sessions, for instance with an emphasis on the need for evidence in order to make a claim. This emphasis contributed to a commitment to evidence as the epistemic basis of beliefs (Osborne, 2014). Another dimension of the meaning of evidence is being generated in order to answer the children's own questions, and it is in line with their interests, which is an important trait for developing explanations in kindergarten (Siry & Max, 2013).

Other features of the teacher's approach, such as prompting students to identify evidence, or providing hints about evidence and claims, are similar to the ones discussed by Gotwals et al. (2012). This classroom environment is aligned with characteristics of dialogic teaching identified by Alexander (2008), for example, it is *supportive* of children's discourse; they may articulate their ideas freely without worrying about "wrong" answers; it is *reciprocal*, teacher and children listen to each other, share ideas and consider alternative viewpoints; it is *purposeful*, teachers steer classroom talk with specific goals in view.

Educational Implications

We suggest the importance of promoting purposeful observation as a source of evidence in kindergarten and in the first years of elementary education, in particular, in life sciences because it supports students in collecting and interpreting data, in the transformation of data into evidence, and in using evidence in order to revise their understandings. Purposeful observation is complementary to investigations and experiments; it poses, perhaps, fewer difficulties for young children. As research shows, even adolescents have problems when planning investigations (Jiménez-Aleixandre & Crujeiras, 2014). What we are proposing is to use them in combination, not to focus only on purposeful observation; however, we suggest that in early ages purposeful observation should be given more emphasis.

We think that our results support Metz (2011) and Gotwals et al. (2014) regarding the relevance of instructional opportunities over developmental constraints. We suggest these findings have implications for the potential influence of the NGSS (Achieve, 2013), not just in the United States, but internationally. The practices enacted by these children are aligned with NGSS recommendations; however, the teacher and her professional learning community are not familiar with the standards. This would indicate that the NGSS recommendations for practices might be extended to other countries and contexts.

More research is needed in order to understand how to support young children's engagement with the use of evidence. From the indicators of cognitive control over coordination of claims with evidence proposed by Kuhn and Pearsall (2000), children in this study showed awareness of their ideas and of the fact that they underwent revision. However, other reasoning practices required for this coordination (Sandoval et al., 2014) were not identified. Our next goal is to examine which of these practices can be performed in kindergarten, and which specific scaffolds are appropriate to promote them, and we are currently involved in a 3-year longitudinal study on this issue.

This study was supported by Spanish Ministerio de Economía y Competitividad (MINECO) (EDU2012-38022-C02-01). To the teacher Dolores Vázquez and her students. Sabela F. Monteira's work is supported by a BES-2013-062873 scholarship from the Spanish MINECO. The authors thank Ibrahim Delen, and JRST Editors, Associate Editor and anonymous reviewers for their suggestions to the first draft.

References

Achieve. (2013). Next generation science standards: For states, by states. Washington, DC: National Academies Press.

Aikenhead, G. S. (2005). Science-based occupations and the science curriculum: Concepts of evidence. Science Education, 89, 242–275.

Alexander, R. J. (2008). Towards dialogic teaching. Rethinking classroom talk (4th edition). York: Dialogos.

Allchin, D. (2011). Evaluating knowledge of the nature of (whole) science. Science Education, 95(3), 518–542.

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. L. A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. Educational Psychologist, 46(3), 141–167.

Delen, I., & Krajcik, J. (in press). What do students' explanations look like when they use second-hand data? International Journal of Science Education.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic and social learning goals. Review of Research in Education, 32, 268–291.

Duschl, R. A., & Bybee, R. W. (2014). Planning and carrying out investigations: An entry to learning and to teacher professional development around NGSS science and engineering practices. International Journal of STEM education, 1:12. doi: 10.1186/s40594-014-0012-6

Ducrot, O. (1983). Opérateurs argumentatifs et visée argumentative. Cahiers de Linguistique Française, 5, 7–36.

Gee, J. P. (2005). An introduction to discourse analysis: Theory and methods. London: Routledge.

Gelman, R., & Brenneman, K. (2012). Moving young "scientists-in-waiting" onto science learning pathways: Focus on observation. In J. Shrager, & S. Carver (Eds.), The journey from child to scientist: Integrating cognitive development and the education sciences (pp. 155–169). Washington, D.C.: American Psychological Association.

Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: Strategies for qualitative research. London: Weidenfeld and Nicholson.

Gotwals, A. W., & Songer, N. B. (2013). Validity evidence for learning progression-based assessment items that fuse core disciplinary ideas and science practices. Journal of Research in Science Teaching, 50(5), 597–626.

Gotwals, A. W., Hokayem, H., & Wright T. (2014). Argumentation at the start of school: Characterizing the entry points into a learning progression for argumentation. Paper presented at the NARST Annual Meeting, Pittsburgh, March 30–April 2.

Gotwals, A. W., Songer, N. B., & Bullard, L. (2012). Assessing students' progressing abilities to construct scientific explanations. In A. C. Alonzo, & A. W. Gotwals (Eds.), Learning progressions in science: Current challenges and future directions (pp. 183–210). Rotterdam: Sense Publishers.

Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. Journal of Research in Science Teaching, 38(6), 663–687.

Hug, B., & McNeill, K. L. (2008). Use of first-hand and second-hand data in science: Does data type influence classroom conversations? International Journal of Science Education, 30(13), 1725–1751.

Inagaki, K., & Hatano, G. (2006). Young children's conception of the biological world. Current directions in psychological science, 15(4), 177–181.

Jewitt, C., & Oyama, R. (2001). Visual meaning: A social semiotic approach. In T. van Leeuwen, & C. Jewitt (Eds.), Handbook of visual analysis (pp. 134–156). London: Sage.

Jiménez-Aleixandre, M. P. (2008). Designing argumentation learning environments. In S. Erduran, & M. P. Jiménez-Aleixandre (Eds.), Argumentation in science education. Perspectives from classroom-based research (pp. 91–115). Dordrecht: Springer.

Jiménez-Aleixandre, M. P., & Crujeiras, B. (2014). Interpreting anomalous primary data in the laboratory: Findings from a longitudinal study. Paper presented at the AERA Annual Meeting, Philadelphia, April 3–7.

Jiménez-Aleixandre, M. P., Bugallo Rodríguez, A, & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. Science Education, 84, 757–792.

Kelly, G. J. (2008). Inquiry, activity and epistemic practice. In R. A. Duschl, & R. E. Grandy (Eds.), Teaching scientific inquiry: Recommendations for research and implementation (pp. 99–117). Rotterdam: Sense Publishers.

Kress, G., & van Leeuwen, T. (2006). Reading images. The grammar of visual design (2nd edition). London: Routledge.

Kuhn, D. (1997). Constraints or guideposts? Developmental psychology and science education. Review of Educational Research, 67(1), 141–150.

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. Journal of Cognition and Development, 1, 113–129.

Lehrer, R., & Schauble, L. (2012). Supporting inquiry about the foundations of evolutionary thinking in the elementary grades. In S. Carver, & J. Shrager (Eds.), The journey from child to scientist: Integrating cognitive development and the education sciences (pp. 171–205). Washington, D.C.: American Psychological Association.

Leibham, M. B., Alexander, J. M., & Johnson, K. E. (2013). Science interests in preschool boys and girls: Relations to later self-concept and science achievement. Science Education, 97, 574–593.

Mantzicopoulos, P., Samarapungavan, A., & Patrick, H. (2009). "We learn how to predict and be a scientist": Early science experiences and Kindergarten children's social meanings about science. Cognition and Instruction, 27(4), 312–369.

McNeill, K. L. (2011). Elementary students' views on explanation, argumentation and evidence, and their abilities to construct arguments over the school year. Journal of Research in Science Teaching, 48(7), 793–823.

McNeill, K. L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. Journal of Research in Science Teaching, 45(1), 53–78.

Metz, K. E. (1995). Reassessment of developmental constraints on children's science instruction. Review of Educational Research, 65(2), 93–127.

Metz, K. E. (1997). On the complex relation between cognitive developmental research and childrens's science curricula. Review of Educational Research, 67(1), 151–163.

Metz, K. E. (2008). Narrowing the gulf between the practices of science and the elementary school science classroom. The Elementary School Journal, 109, 138–161.

Metz, K. E. (2011). Disentangling robust developmental constraints from the instructionally mutable: Young children's epistemic reasoning about a study of their own design. Journal of the Learning Sciences, 20(1), 50–110.

Merriam, S. (2009). Qualitative research: A guide to design and implementation. San Francisco, CA: Jossey-Bass.

Morris, C. (2007). Teaching and learning through active observation. Creating and supporting opportunities to learn through work participation. London Deanery. http://www.faculty.londondeanery.ac.uk/e-learning/feedback/files/T-L_through_active_observation.pdf. (Retrieved, November 2014).

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Washington, DC: National Academies Press.

OECD. (2012). Access to early childhood education. In Education at a glance, 2012: Highlights. OECD Publishing.

Osborne, J. (2014). Scientific practices and inquiry in the science classroom. In N. G. Lederman & S. K. Abell (Eds.), Handbook of research on science education, Volume II (pp. 1835–1901). New York: Routledge.

Patrick, H., Mantzicopoulos, P., & Samarapungavan, A. (2009). Motivation for learning science in Kindergarten: Is there a gender gap and does integrated inquiry and literacy instruction make a difference. Journal of Research in Science Teaching, 46(2), 166–191.

Pluta, W. J., Chinn, C. A., & Duncan, R. G. (2011). Learners' epistemic criteria for good scientific models. Journal of Research in Science Teaching, 48(5), 486–511.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. Cognition and Instruction, 23(1), 23–55.

Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. Educational Psychologist, 49(2), 139–152.

Siry, C., & Max, C. (2013). The collective construction of a science unit: Framing curricula as emergent from kindergarteners' wondering. Science Education, 97, 878–902.

Songer, N. B., & Gotwals, A. W. (2012). Guiding explanation construction by children at the entry points of learning progressions. Journal of Research in Science Teaching, 49(2), 141–165.

Varelas, M., & Pappas, C. C. (2013). Integrating science and literacy: Forms and functions. In M. Varelas, & C. C. Pappas (Eds.), Children's ways with science and literacy (pp. 3–19). New York: Routledge.

Wickman, P.-O. (2004). The practical epistemologies of the classroom: A study of laboratory work. Science Education, 88(3), 325–344.

Yin, R. (2003). Case study research: Design and methods (3rd edition). Thousand Oaks, CA: Sage Publications.

Zangori, L., Forbes, C. T., & Biggers, M. (2013). Fostering student sense making in elementary science learning environments: Elementary teachers' use of science curriculum materials to promote explanation construction. Journal of Research in Science Teaching, 50(8), 989–1017.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.