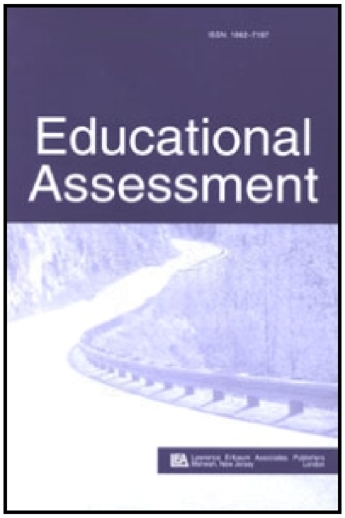

Educational Assessment
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653641

The Analysis of Classroom Discourse: Elementary School Science Curricula Advancing Reasoning With Evidence

Ilonca Hardy[a]; Birgit Kloetzer[b]; Kornelia Moeller[c]; Beate Sodian[b]
[a] University of Frankfurt, Germany [b] University of Munich, Germany [c] University of Muenster, Germany

Online publication date: 10 December 2010

To cite this Article Hardy, Ilonca , Kloetzer, Birgit , Moeller, Kornelia and Sodian, Beate(2010) 'The Analysis of Classroom Discourse: Elementary School Science Curricula Advancing Reasoning With Evidence', Educational Assessment, 15: 3, 197 — 221

To link to this Article: DOI: 10.1080/10627197.2010.530556

URL: http://dx.doi.org/10.1080/10627197.2010.530556

Educational Assessment, 15:197–221, 2010

ISSN: 1062-7197 print/1532-6977 online
DOI: 10.1080/10627197.2010.530556

Routledge
Taylor & Francis Group

The Analysis of Classroom Discourse: Elementary School Science Curricula Advancing Reasoning With Evidence

Ilonca Hardy
University of Frankfurt, Germany

Birgit Kloetzer
University of Munich, Germany

Kornelia Moeller
University of Muenster, Germany

Beate Sodian
University of Munich, Germany

Students' ability to participate in scientific discourse and to appropriately use empirical evidence to support beliefs or conclusions has been consistently stated as a core goal of science education. In the present study, we analyzed the quality of scientific reasoning in elementary school science classrooms, using the Evidence-Based Reasoning (EBR) Video Framework (see Furtak, Hardy, Beinbrech, Shavelson, & Shemwell, this issue). Two data sets from six 3rd-grade and 4th-grade science classrooms drawn from class discussions on floating and sinking were analyzed for the quality of EBR, the kinds of teacher prompts, and the level of conceptual understanding in classroom discourse. We found that the majority of discourse involved unsupported claims about the scientific phenomena. Although there was a clear progression in conceptual understanding over the course of the curriculum, no consistent effects were observed in the level of EBR. There was some evidence for effects of a preceding curriculum that had focused on nature-of-science constructs on the quality of students' EBR. Moderate correlations were obtained between students' conceptual understanding and reasoning level within reasoning units. Teacher prompts for providing support for conclusions and inferences were associated with higher reasoning levels, emphasizing the role of teachers in promoting a culture of productive use of evidence in classroom discourse. The quality of the EBR Video Framework as an assessment tool is discussed.

Correspondence should be sent to Ilonca Hardy, Fachbereich Erziehungswissenschaften, Institut für Pädagogik der Elementar- und Primarstufe, Fach 113, Senckenberganlage 15, D-60054 Frankfurt am Main, Germany. E-mail: hardy@em.uni-frankfurt.de

EVIDENCE-BASED REASONING IN SCIENCE DISCOURSE

Students' ability to participate in scientific discourse and to use empirical evidence to support inferences and conclusions is consistently stated as a core goal of science education (American Association for the Advancement of Science, 1993; Bybee, 2002; National Research Council, 1996). For example, the essential features of scientific inquiry in the National Research Council's (1996) *National Science Education Standards* include the expectation that students formulate explanations from collected evidence and evaluate these explanations with regard to plausible alternatives. Similarly, Duschl and Gitomer's (1997) discussion of the goals of science education includes students learning the methods of science exploration by generating data and evidence as well as learning the argumentative skills to develop theories that link evidence to explanation. These goals frame scientific inquiry as an inductive process that is based on transformations from data to evidence, from evidence to patterns, and from patterns to explanations (Duschl, 2003). Of importance, scientific inquiry is not merely the collection of empirical evidence but also the use of that evidence for the construction, confirmation, or refutation of theories and models (Driver, Newton, & Osborne, 2000). Thus, students' participation in scientific inquiry requires a basic understanding of the nature of science as an iterative process of formulating hypotheses and theories, and collecting evidence that is, in turn, based on theories and hypotheses (e.g., Carey & Smith, 1993).

How is productive scientific inquiry fostered? In a recent review, Duschl (2008) suggested that important roles for teachers include creating learning contexts in which students can use evidence in scientifically meaningful ways as well as creating opportunities for monitoring and assessing students' appropriate construction and use of evidence. Duschl also emphasized that assessment practices should provide students and teachers with feedback on the achievement of conceptual, epistemic, and social learning goals to promote scientific reasoning. In particular, teacher prompts can play a pivotal role in the construction of formative assessment routines, as they can provide opportunities for students to establish scientifically valid argumentation patterns. Prompting and modeling the use of evidence to support scientific theories and inferences by students may contribute to a classroom climate fostering Evidence-Based Reasoning (EBR) and may also contribute to the development of individual students' abilities in scientific reasoning.

THE DEVELOPMENT OF SCIENTIFIC REASONING IN ELEMENTARY SCHOOL

Scientific explanations for phenomena need to be constructed by relating information to underlying theories and their supporting evidence rather than on an ad hoc basis. Kuhn and Franklin (2006) noted three requirements that need to be fulfilled for children to coordinate theory and evidence. Children need to realize that: (a) a theoretical claim can be falsified; (b) evidence can be used as a means of falsification; and (c) evidence and claim are different epistemological categories. Thus, for students to use appropriate arguments in science discourse, they must realize that the construction, testing, and revision of theories and hypotheses about scientific phenomena constitute a fundamental part of the scientific endeavor.

In developmental psychology, scientific reasoning has traditionally been seen as a late developing ability and a hallmark of intellectual maturity (Inhelder & Piaget, 1958; Kuhn, Amsel, & O'Laughlin, 1988). Scientific reasoning is also multifaceted. It includes: (a) the application of appropriate research methods to test hypotheses, which might involve using a control-of-variables strategy when designing experiments, or using correlation methods when data sets are being compared. Scientific reasoning also includes: (b) the ability to evaluate data patterns and to use representations and models; as well as (c) a metaconceptual understanding of the nature of science, or the ability to differentiate between theoretical claims, hypotheses, and evidence to support claims and hypotheses.

Contrary to traditional assumptions, recent research has demonstrated that elementary school children and even preschoolers have some relevant component scientific reasoning skills (see Bullock, Sodian, & Koerber, 2009; Kuhn & Pearsall, 2000; Zimmerman, 2007, for a review). Given contextual support, third and fourth graders can distinguish between a controlled and a confounded experiment and understand why a controlled experiment is preferable when testing a hypothesis about the causal effects of one variable on another. Without contextual support, the large majority of elementary school children do not spontaneously apply appropriate research methods, such as a control-of-variables scheme (Bullock & Ziegler, 1999), although teaching interventions show that third graders can be trained to use this strategy (Strand-Cary & Klahr, 2008). First and second graders are able to differentiate between tests with conclusive and inconclusive results as well as explain their choices if they are prompted to formulate a hypothesis (Sodian, Zaitchik, & Carey, 1991). Similarly, basic skills in data evaluation have been shown even in preschoolers (Koerber, Sodian, Thoermer, & Nett, 2005). Nevertheless, children also tend to rely heavily on their domain-specific knowledge when interpreting data, often ignoring or distorting evidence that is inconsistent with their prior beliefs (Chinn & Malhotra, 2002; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995). Therefore, students' domain-specific knowledge should be assessed alongside attempts to fully understand how children develop the ability to reason scientifically.

Although young children's basic abilities to apply experimental strategies and skills have been demonstrated in a substantial number of studies, an explicit metaconceptual understanding of the nature of science develops more slowly and matures only in early adulthood, if at all. Research has consistently indicated that students in early secondary school believe knowledge is the result of direct observation or factual information, an understanding that corresponds to the most basic of the levels of understanding described by Carey, Evans, Honda, Jay, and Unger (1989). At this most basic level (Level 1), there is no distinction between ideas, theories, or hypotheses and evidence. Science is conceived of either in terms of producing positive effects or in terms of collecting factual knowledge. At Level 2, students understand that scientists search for explanations of natural phenomena, and they have a basic understanding of testing causal hypotheses. However, students see the process of hypothesis testing as a series of largely unconnected experiments and fail to understand the iterative and cumulative nature of the construction of scientific knowledge (see Driver, Leach, Millar, & Scott, 1996; Lederman, 1992; McComas, 1998, for similar descriptions). A mature epistemology of science (Carey et al.'s Level 3) recognizes the role of theories or interpretive frameworks in knowledge construction. The highest level of this understanding of the cumulative and cyclical nature of knowledge construction is almost never found in students. Even adult university students often only implicitly acknowledge the role of theories

(Thoermer & Sodian, 2002). In a long-term longitudinal study of the development of scientific reasoning from elementary school to adulthood, Bullock et al. (2009) found a slow and protracted development of metaconceptual understanding of science, with progress from Level 1 to Level 2 in adolescence and young adulthood, but only rare instances of reasoning at Level 3.

As previously suggested, students can benefit from curricular intervention. For example, Carey et al. (1989) designed a seventh-grade curriculum on the nature of science constructs that included the explicit testing of hypotheses and reflection on evidence and theory building. This curriculum concerned the role of yeast formation for baking bread. In a comparison of pre- and postcurriculum interviews, students on average showed an increase of about half of a level on Carey's Nature of Science Interview (Carey et al., 1989). In a similar curricular intervention study in elementary school, Sodian, Thoermer, Kircher, Grygier, and Günther (2002) observed an increase in fourth-graders' average level of understanding of the nature of science, with students changing from Level 1 (science as collecting facts) in preinstructional interviews toward Level 2 (science as search for explanations) in the postinstructional interviews. Similar to Carey et al.'s (1989) study, students' improvement was toward an implicit Level 2 understanding of science, rather than a full or explicit one (Level 3). More specifically, students were able to provide Level 2 ideas only when working with concrete examples of scientific inquiry or when the curriculum prompted reflection on theory-construction and the use of evidence (Grygier, 2008).

Smith, Maclin, Houghton, and Hennessey (2000) demonstrated the long-term importance of metaconceptual reflection for a basic understanding of the nature of scientific knowledge at the elementary school level. This study compared the outcomes of a traditional curriculum with an inquiry-based classroom, in which reflection on the theory–evidence relation was emphasized. Students in the inquiry-based classroom showed a significantly higher level of understanding the nature of science in sixth grade.

ASSESSMENT OF STUDENTS' SCIENTIFIC REASONING

Developmental research on scientific reasoning has provided instruments to assess students' skills in experimentation and data evaluation as well as their understanding of the nature of science. This research has relied largely on data from individual students rather than whole-class discussions. Assessments have included interview studies on the nature of science (Carey et al., 1989), tasks involving students' application of control-of-variables strategies, or tasks addressing students' ability to evaluate data patterns in experimental designs with varying hints and contextual support (Koerber et al., 2005). Whereas the application of these instruments in research has demonstrated that even young children can "reason scientifically," the processes underlying the acquisition of skills, strategies, and metaconceptual knowledge are poorly understood. Microgenetic studies of scientific reasoning (Schauble, 1990) have also focused on individual strategy acquisition.

In contrast, within the literature on science education, as well as in empirical studies of scientists' real-world reasoning processes (Dunbar, 1994), there is an explicit focus on scientific reasoning as a fundamentally social process mediated by epistemological, cultural, and technological factors and amenable to scaffolding by teachers and curriculum materials. To facilitate reasoning processes in a pedagogical context, Duschl (2003, 2008) has suggested that "conversations should mediate the transition from evidence to explanation and vice versa" (Duschl, 2008, p. 280). More specifically, learners should be engaged in conversations regarding observations, data, and theory; empirically test their knowledge; and theorize about their results. This emphasis on scientific reasoning as an inherently social discourse gives special importance, within research on science education, to the role of argumentation and explanation in prompting and promoting scientific reasoning (see also Erduran & Jimenez-Aleixandre, 2007, on the contribution of argumentation for facets of science learning). In this literature, science instruction is designed to allow students opportunities to move along within an Evidence-Explanation continuum (Duschl, 2008) where data become evidence, evidence is used to ascertain patterns and models, and models are employed to propose explanations. This process thereby provides students an opportunity to experience how these transformations are a critical part of the nature and practice of science.

Implementing the model of an Evidence-Explanation continuum in science classrooms requires assessment tools that allow teachers to measure and guide students' construction of arguments and explanations, and their argumentation exchanges. Whereas there are a number of adaptations of Toulmin's (1958) framework of argumentation for classroom discourse (see Erduran & Jimenez-Aleixandre, 2007; Sampson & Clark, 2006), finding appropriate assessment tools for students' scientific reasoning that have the potential to provide feedback to both students and teachers remains a challenge (Duschl, 2008). The demand for new assessment tools has been recognized especially within reform-based teaching approaches (National Research Council, 2001).

A starting point for conceptualizing the assessment of scientific reasoning within classroom discourse is Duschl and Gitomer's (1997) construct of "assessment conversations." Assessment conversations engage students in discussion of a diversity of ideas and representations and prompt them to construct, coordinate, and reflect upon accompanying evidence, thereby achieving learning goals based on the coordination of theory and evidence. The term "conversation" is used in a broad way, including argumentation, modeling, drawing, and writing as representations of scientific ideas. Assessment conversations are thus conceived as an instructional dialogue that embeds assessment into the activity structure of the classroom. In Project SEPIA (Duschl & Gitomer, 1997), assessment conversations are employed on the basis of student products with the goal of developing questions and activities to promote conceptual understanding and provide assessment information to teachers. Following three stages of assessment conversations (receiving information, recognizing information, and using information), specific teacher prompts led students to focus on essential elements of scientific reasoning, such as looking at relationships, consistency, or use of examples as a means of promoting discourse concerned with reasoning processes rather than facts. Thus, Duschl and Gitomer (1997) found that teacher prompts, such as asking students to think about whether an explanation is supported by evidence, can serve as a formative assessment procedure to provide information and scaffolding to students and teachers.

However, merely encouraging students to articulate and justify their views may not be sufficient to promote appropriate argumentation, if not accompanied by a focus on counterargument and rebuttal (Simon, Osborne, & Erduran, 2003). Findings from a year-long investigation by Simon, Erduran, and Osborne (2006) on the teaching of argumentation in scientific contexts showed that teachers who attained high levels of argumentation in their classrooms also encouraged their students to engage in higher order processes such as evaluating arguments and thinking of alternatives. To date, the specific effects of teacher prompts in promoting scientific reasoning at an individual or classroom level are still poorly understood.

SCIENTIFIC REASONING IN CONCEPTUALLY CHALLENGING DOMAINS: THE TOPIC OF FLOATING AND SINKING

The relationship between students' level of scientific reasoning and their conceptual understanding has received little systematic research attention. Tytler and Peterson (2004) suggested that students with only a limited understanding of the concepts and mechanisms involved in producing a phenomenon may be constrained in their consideration of patterns of evidence to support their initial explanations of the phenomenon. Similarly, von Aufschnaiter, Erduran, Osborne, and Simon (2008) found a covariance of conceptual understanding and student reasoning in case studies of students in which a more sophisticated understanding of concepts seemed to support more sophisticated uses of evidence. In turn, a limited general understanding of the theory–evidence relationship may pose a domain-general impediment to the acquisition of complex science concepts (e.g., Carey et al., 1989). Alternatively, attempts to observe and explain complex phenomena may in themselves promote scientific reasoning, especially if scaffolding is provided. Science curricula that involve students in the investigation of complex scientific phenomena may promote students' level of scientific reasoning as well as their conceptual understanding of complex phenomena. It remains an open question as to whether and to what extent challenging involvement in the construction of adequate explanations for scientific phenomena can improve students' level of reasoning.

The literature on conceptual change has been concerned with individual processes of knowledge construction with regard to scientific phenomena, in particular by investigating the role of students' naive conceptions in their successful construction of adequate scientific explanations. Students' explanations for natural phenomena (naive conceptions) are based on everyday experiences. In many cases, these naive explanations are not compatible with scientific models and need to be revised or fundamentally restructured into new, scientifically valid ideas (diSessa, 2006; Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001). To date, research addressing the collective and individual processes of conceptual change in classroom discourse is rare because of methodological challenges such as separating the levels of conceptual understanding from reasoning levels.

In this article, we describe an analysis of elementary school classroom discourse from instructional units on floating and sinking. Because this is a challenging instructional topic, it provides a useful context for the investigation of the relationship between scientific reasoning processes and conceptual understanding.

Answering why something floats or sinks can be framed in terms of "material kind" (e.g., wood floats; metal sinks) or in terms of relative density and buoyancy. Progressing from material kind explanations to the concepts of density and buoyancy is challenging, as it requires reasoning about relationships, not single attributes. Understanding Archimedes' principle, for example, requires considering the relationship between an object and its surrounding fluid, and to compare densities (of object and fluid) and forces (of gravity and buoyancy). This is a challenge to most elementary school students as they tend to consider only one dimension, in this case focusing exclusively on the mass, volume, or shape of objects to determine whether they will float or sink in water (Hardy, Jonen, Möller, & Stern, 2006; Möller, 1999; Smith, Carey, & Wiser, 1985).

RESEARCH QUESTIONS

Although conceptual understanding and level of reasoning are intertwined in classroom discourse, it is possible to analyze these two dimensions separately. In the present research, we did so by applying the EBR Video Framework (Furtak et al., this issue) to two data sets of elementary school science lessons on the topic of sinking and floating. The video framework allows one to break discourse into "reasoning units" (or coherent segments of reasoning that refer to the same claim or premise) and then to perform a three-dimensional analysis of these units.

The first dimension concerns the quality of reasoning, which is measured by how claims are justified or supported. The system consists of four levels: Level 1 (unsupported claims) is the lowest level and consists of claims only. Level 2 (phenomenological reasoning) contains claims supported by observations or single phenomena. In Level 3, claims are supported with relational or EBR (relational reasoning). Finally, Level 4 consists of claims supported with rule-based generalizations (rule-based reasoning).

The second dimension measures the teacher's contribution to each reasoning unit, by identifying the content for which teacher prompts are provided (for premises, claims, and claim justifications, termed "backings").

The third dimension addresses the conceptual level of reasoning about the phenomena being discussed and is measured with three levels ranging from naive conceptions (Level 1—Naive), to reasoning based on everyday experience (Level 2—Prescientific), and reasoning referring to scientific variables and dimensions (Level 3—Scientific).

As previously noted, little research has addressed developmental change in the quality of students' scientific reasoning and science understanding on the microgenetic level of classroom-based discourse over the course of a science curriculum. Therefore, the questions we addressed were, Will elementary students' level of reasoning be promoted by inquiry-based science curricula? If so, what is the teacher's role in promoting scientific reasoning? What teacher prompts are associated with higher level reasoning in discourse, and do these associations change during the course of the curriculum, that is, from first to last lesson? Finally, what is the relationship between students' conceptual understanding of the phenomena under study and their level of reasoning?

To investigate these questions, we analyzed transcripts of teacher–student classroom discourse from two different data sets, each drawn from an elementary school curriculum on floating and sinking (four lessons each; see Hardy et al., 2006). The curriculum was an instructional unit based on constructivist principles and designed to impart the notions of density (i.e., a qualitative differentiation of weight and density) and buoyancy to students. The unit proceeded from an initial question (why does a large ship of iron float?) through a series of activities addressing the notions of material kind and density, to an understanding of the relation between water displacement and buoyancy force; in the last lesson, the class returned to the initial question and brought the newly acquired concepts to bear on it. The student–teacher discourse used in our analyses was originally recorded in the context of two different studies: one on the effects of instructional scaffolding (Data Set 1; four classrooms) and one on the effects of instruction on the nature of science (Data Set 2; two classrooms). The classrooms were comparable with respect to age group, topic, curriculum, and expected conceptual learning progress, so that our research questions may be productively addressed with each of the data sets. By including two different data sets, it was also possible to test for the generality of the findings independently of specific context factors with respect to improvements in reasoning quality over the course of an instructional unit. Furthermore, because the specific experimental variations used in the two studies have been shown to be effective on the level of individual conceptual change (Grygier, 2008; Hardy et al., 2006), we anticipated these treatment effects to be reflected on the level of classroom discourse. That is, we expected the experimental variation of Data Set 1 (teacher's scaffolding in classroom with high instructional support) and the experimental variation of Data Set 2 (a preceding nature-of-science curriculum) to have positive effects on the level of reasoning over the course of the lessons.

METHOD

Data Sources

The data were derived from transcriptions of science lessons from six elementary school classrooms that had participated in two studies on science instruction. The classrooms were from two settings. Classroom discourse in Data Set 1 was from third-grade classrooms; discourse from Data Set 2 was from fourth-grade classrooms.

Data Set 1: Scaffolding Study

Data Set 1 was derived from transcriptions of four 90-min lessons for each of four third-grade classrooms that were part of a larger study on the role of instructional support for conceptual understanding of floating and sinking. Two of the classrooms received the curriculum under high instructional support and two under low instructional support. An additional two classrooms that were part of the larger study were excluded from our analyses due to minor effects of the experimental variation in the larger study. The students in the four classrooms were comparable with regard to mean levels of conceptual understanding prior to instruction and general characteristics of school and classroom environments (see Hardy et al., 2006, for an extensive description of the sample, curriculum, and results on conceptual understanding in pre- and posttests). The transcriptions for Data Set 1 were from four of the eight lessons in the unit—Lessons 1, 3, 6, and 8—to capture an expected learning progression from the first to last lesson, with two intermediate lessons that included generally comparable lesson content across the high and low instructional support conditions (see Table 1, and Hardy et al., 2006, for a description of the entire curriculum). The experimental variation applied to the four classrooms involved different degrees of instructional support within the learning environments on floating and sinking, concerning the sequencing of content and the use of cognitively structuring statements

TABLE 1
Topics Covered in Curriculum on "Floating and Sinking"

Lesson	*Topics (Instructional Activities): Data Set 1*	*Topic (Instructional Activities): Data Set 2*
1	Topic: Why does a large ship of iron float; material kind (formulation of initial hypotheses; work with solid objects; ordering of objects according to floating and sinking in water)	Topic: Material kind (formulation of initial hypotheses; work with solid objects; ordering of objects according to floating and sinking in water)
2	Topic: Material kind (work with solid objects, addressing misconceptions; ordering of objects according to material)	Topic: Density (work with objects of same volume and different mass; differentiation of mass and volume)
3	Topic: Density (work with objects of same volume and different mass; differentiation of mass and volume)	Topic: Water Displacement (work with solid and hollow objects, varying in volume and mass)
4	Topic: Water Displacement (work with solid and hollow objects, varying in volume and mass)	Topic: Relation of Displaced Water and Density of Objects; Comparison of Densities (work with floating and sinking solid and hollow objects)
5	Topic: Buoyancy Force (work with large hollow objects and own body, experiencing buoyancy force in the swimming pool)	
6	Topic: Water Displacement and Buoyancy Force (work with solid and hollow objects on buoyancy force and relation to water displacement)	
7	Topic: Relation of Water Displacement and Buoyancy Force (work with modeling clay (modeling of boats) on the relation of water displacement and buoyancy force)	
8	Topic: Relation of Buoyancy Force and Gravity (representation of forces by children pulling on each other's hands; work with large hollow objects)	

by the teacher. In the high instructional support group, topics were segmented into smaller units (e.g., material kind, density, water displacement, and buoyancy force) in a predetermined order from more basic to more complex, and the teacher scaffolded the lesson by contrasting student hypotheses and conceptions, addressing naive conceptions, and introducing a hypothesis or observation that the students themselves had not considered.

In the low instructional support group, topics were not segmented and the order of presentation was not predetermined. In addition, in the low instructional support group, the instructional discourse was student centered, with students themselves reacting to each other's statements, whereas the teacher's role was focused on organizational supervision with a lower frequency of content-related prompts to students' reasoning processes. The instructional time, the teacher, and the curriculum materials were not varied. Table 1 (left column) lists the topics of Lessons 1 to 8 covered in Data Set 1.

Because the sequencing of content was different in the high and low instructional support groups, the content of the four transcribed lessons (Lessons 1, 3, 6, and 8) was not identical. The instructional content for Lessons 1 (introductory lesson) and 8 (final lesson) were largely similar across all four classrooms, but the content of Lessons 3 and 6 were less so. In the high instructional support condition, Lesson 3 concerned relative density, and Lesson 6 concerned buoyancy, whereas these lessons in the low instructional support condition covered the particular concepts introduced by the students themselves.

Data Set 2: Nature of Science Study

Data Set 2 was derived from transcripts of four 90-min lessons on floating and sinking from each of two fourth-grade classrooms that had participated in a study assessing the effects of a preceding curriculum on nature of science concepts. One of the two classrooms (experimental group) had received an extended unit on the nature of science, whereas the other had received a curriculum on the same content with no emphasis on nature of science discussions (control group). The floating and sinking curriculum from which Data Set 2 was derived was a reduced and more focused variant of the high instructional support curriculum of Data Set 1, with four lessons that covered material kind, density, and the comparison of densities (see Table 1, right column). For this curriculum, the teacher attempted to cognitively activate, challenge, and scaffold students' conceptual understanding much as in the high instructional support group of Data Set 1. Thus, each of the two classes in Data Set 2 received the same floating and sinking curriculum but had had different prior instruction.

The prior instruction given to the Data Set 2 experimental group consisted of 14 weekly 90-min lessons on the nature of science, starting with a collection of students' initial conceptions about scientific inquiry and gradually attempting to refine their conceptions about experimentation and the relation between theory and evidence with regard to topics including perception, light and shadow, yeast formation, and others. The lessons for the control group treated the same topics conceptually, without offering explicit opportunities to reflect on the nature of scientific inquiry (see Sodian, Jonen, Thoermer, & Kircher, 2006, for an extensive description of the curriculum). The teacher for the prior instruction and the unit on floating and sinking remained the same within each classroom. The experimental and control classrooms were comparable with regard to social demographic characteristics and cognitive ability of students (see Sodian et al., 2006).

Results of Prior Analyses on Student Conceptual Understanding

As outcomes from performance for Data Sets 1 and 2 have been analyzed previously with regard to intervention effects on student conceptual understanding, we briefly summarize these results here. In both data sets, students' conceptual understanding was measured by a test on floating and sinking that is described in detail in Hardy et al. (2006). Explanations for floating and sinking were classified on one of three levels of understanding—naive conceptions (using a single dimension such as weight, size, shape, or active air), prescientific conceptions (using single qualities such as hollowness or material kind, or referring to a qualitative relation between water and object), and scientific conceptions (referring to density, buoyancy, and their relation to water displacement). In Data Set 1, we used this test to generate pretest, posttest, and 1-year follow-up measures of the effects of the curricular variation. There were group effects such that students in the high instructional support group rejected significantly more naïve conceptions on the posttest than the low instructional support group and showed superior integrated conceptual understanding over the long run (see Hardy et al., 2006).

Although students in Data Set 2 improved their nature of science understanding between a pre- and posttest on the role of theory and evidence in the scientific process (as assessed by the Nature of Science Interview; Carey et al., 1989) and improved in the application of experimental control strategies (as assessed by an experimental manipulation task; see Sodian et al., 2006) relative to the control group, there were no significant differences between the experimental and control group with regard to mean level of conceptual understanding.

Coding Scheme for Discourse Analyses

Classroom discourse and teacher prompts for the present analyses were taken from whole-class discussions that occurred toward the end of each 90-min lesson after students had worked independently on experiments in the classroom. In these discussions, the teacher asked students to report and reflect on their previously performed activities and insights to assess conceptual difficulties and progress. We chose this period because it provides rich opportunities for students to refer to empirical evidence in support of their assertions about the material and phenomena.

The specific coding manual for this study was developed on the basis of transcripts of science lessons from elementary and secondary school from a variety of data sets. To facilitate coding, each transcript was supplemented with a "story line," summarizing the main content of the discourse. The story line for each transcript was written collaboratively by two coders for all of the transcripts.

Reasoning Units and Scoring

Transcripts were first segmented into "reasoning units" and then were scored for claims, claim justifications (backings), conceptual level, and teacher prompt.

Reasoning units. The coding system (described in detail by Furtak et al., this issue) breaks discourse into units of analysis called "reasoning units." Each reasoning unit includes all the verbal exchange concerned with a given claim (i.e., a student or teacher assertion about a specific topic). Transcripts are segmented into reasoning units by identifying changes in either premises or claims, or both.

Backings/claim justifications. Claim justifications (also referred to as backings) were coded for type of support given for a claim (e.g., use of data, use of evidence, or use of rules in support of initial claims) and assigned a reasoning level. Level 1 indicated no support for a claim. Level 2 (phenomenological) referred to a single observation or property of the phenomenon. Level 3 (relational) related two or more properties or dimensions. Level 4 (rule based) referred to a generalized relation or principle. Table 2 provides an overview of the types of backings and examples. Note that reasoning by analogy was coded at Level 3 (evidence

TABLE 2
Coding of Quality of Reasoning in Classroom Discourse

Reasoning Level	*Definition*	*Examples*
Unsupported	No reasoning: Claim is not backed up	The rock sinks
Phenomenological	Data-based reasoning: Claim is backed up by reference to single property/observation	The rock sinks because it's heavy
Relational	Evidence-based reasoning: claim is backed up by reference to a contextualized relationship between two properties, a property and a consequence of that property, or a specific finding	The rock sinks because its mass is greater than its displaced volume
Rule based	Inductive/deductive rule-based reasoning: Claim is backed up by a generalized relationship, principle, or law	The rock sinks because all objects with a density greater than the density of water will sink

based) because two data points were related to support a claim, whereas reasoning by induction or deduction was scored as Level 4 (rule based) reasoning regardless of whether the rule was described explicitly (e.g., in mathematical terms).

Conceptual understanding. In addition to level of reasoning, we scored each reasoning unit for level of conceptual understanding of the phenomenon using a coding scheme developed in a previous study of elementary school science instruction (Hardy et al., 2006). Responses were classified as naive conceptions (Level 1), prescientific conceptions (Level 2), or scientific conceptions (Level 3). In reasoning units with several conceptual codes, the highest level of conceptual understanding was taken as an indicator of students' and teachers' (collaborative) conceptual understanding. When there was no conceptual content in a reasoning unit, no code for conceptual understanding was assigned. A description of coding levels and examples is shown in Table 3.

Teacher Prompts

Teacher contributions to the classroom discourse were scored for each reasoning unit. These scores marked teacher contributions as providing or prompting for premises, claims, or backings.

Each data set had one primary coder and (for reliability checks) a secondary coder for 25% of the transcripts. The percentage agreement of two independent coders for: (a) identifying reasoning units; (b) identifying backings; (c) scoring backings as "data," "evidence," or "rule"; (d) coding conceptual level; and (e) coding teacher contribution varied between 80% and 95%, yielding Cohen's kappa of .71 to .75. The codes of the primary coder for each data set were used for analyses.

The final sample consisted of 1,011 reasoning units for Data Set 1 and 393 reasoning units for Data Set 2. Of these, 476/47% (Data Set 1) and 137/35% (Data Set 2) were also assigned a code of conceptual understanding.

TABLE 3
Coding of Conceptual Understanding

Levels of Conceptual Understanding	*Definition*	*Examples*
Level 1: Naive conceptions	Conceptions that are incommensurable with scientific explanations, commonly focusing on single properties of objects (e.g., mass, volume, shape) or air an active force	The wooden button sinks because it has holes; it floats because the board is spread out; it sinks because it is so heavy
Level 2: Prescientific conceptions	Conceptions which hold in many situations of everyday life without considering the causal mechanisms of floating and sinking, such as mentioning the role of the water (as the object being lighter or heavier than water), the concept of material kind, and the hollowness of objects	It floats because it is wooden; it sinks because the water weighs less; it floats because it is hollow
Level 3: Scientific conceptions	Conceptions considering one or more of the physical quantities of density, water pressure, or buoyancy force as in the comparison of densities (of object and fluid) and the comparison of forces (of gravity and buoyancy)	It floats because the water pushes the board up strong enough; the wooden board floats because the displaced water weighs more than the board

RESULTS

Development of Levels of Reasoning

The distribution of reasoning level for different types of claim justifications/backings across all the reasoning units in each set of lessons for Data Sets 1 and 2 are shown in Figures 1 and 2. A high percentage of reasoning units in all of the groups was coded as Level 1 (i.e., unsupported claims and premises). In Data Set 2 (Figure 2), an increase in level of reasoning from first to last lesson was observed, with increasing percentages for Level 3 (premise, claim, and backing by evidence), and even some Level 4 (premise, claim, and backing by rule) reasoning in later lessons. For example, in the fourth lesson, 10% of the reasoning units in the experimental group (nature of science) are coded on Level 4, whereas this level was only observed in 1.5% of the reasoning units in the first lesson. In contrast, in Data Set 1, such a progression was not observed over the course of the curriculum. In both groups, more than 70% of all reasoning units involved unsupported statements, even in Lesson 8. Changes in the average level of reasoning across the four lessons was analyzed with an analysis of variance (ANOVA), with group (High vs. Low Instructional Support in Data Set 1; With/Without nature of science in Data Set 2) as a between factor and lesson as a within variable, separately for each data set.

In Data Set 1, there was a significant effect for lesson, $F(3, 1011) = 5.04$, $p < .01$, $\eta^2 = .015$, qualified by a Lesson $\times$ Group interaction, $F(3, 1011) = 2.90$, $p < .05$, $\eta^2 = .009$. Post hoc comparisons for each group showed significant changes in the average level of reasoning across lessons for the high instructional support group, $F(3, 594) = 8.73$, $p < .05$,

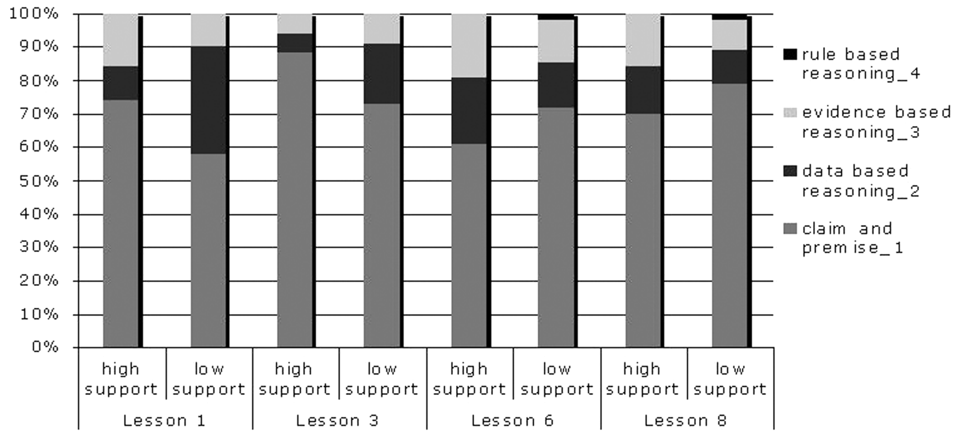

FIGURE 1 Levels of quality of reasoning by group and lesson for Data Set 1.

$\eta^2 = .042$, but nonsignificant changes for the low instructional support group. However, post hoc comparisons, carried out separately for each group, revealed no significant change from Lessons 1 to 8, even in the high instructional support group. These results indicate that for both high and low instructional scaffolding there was little gain in overall reasoning level.

Analysis for Data Set 2 revealed a significant main effect for lesson, $F(3, 393) = 3.51$, $p < .05$, $\eta^2 = .026$, and no interactions. Using follow-up tests to assess differences in the levels of reasoning between the lessons for each group, a significant increase was found for the group with a nature of science curriculum comparing Lesson 1 to Lessons 3 and 4 ($p < .05$), as well as for the control group, with Lesson 1 versus 2 and 1 versus 4 ($p < .05$). An ANOVA comparing initial levels of reasoning for the two groups revealed that the experimental group started with a greater mean level of reasoning in Lesson 1, indicating a higher initial level of reasoning in the group with nature of science tuition. At the same time, differences between the

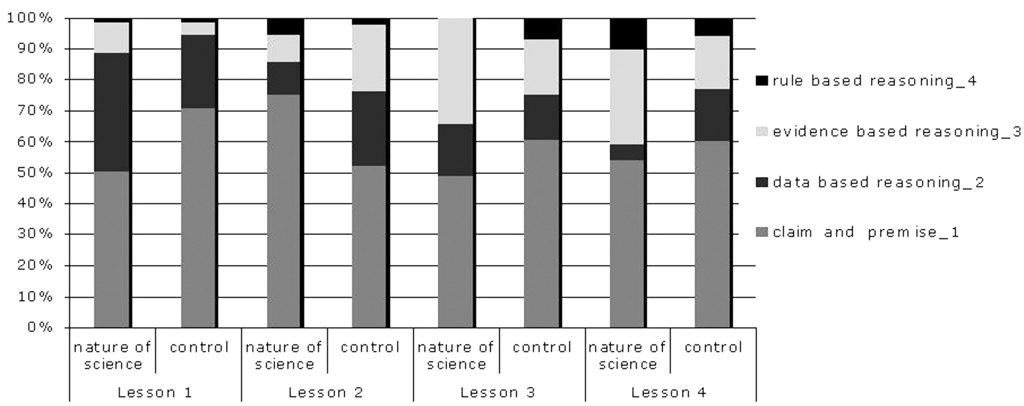

FIGURE 2 Levels of quality of reasoning by group and lesson for Data Set 2.

groups were not significant in a comparison of Lesson 4, indicating that the nature of science effect was not maintained over the course of the curriculum.

Development of Levels of Conceptual Understanding

Descriptive results for the development of conceptual levels across the four lessons show that an increasing number of reasoning units was coded at Levels 2 and 3 in the later lessons than in earlier ones across conditions and data sets. The frequency of Level 3 (scientific concepts) responses increased from close to 0 in the first lesson to about 20% in the last lesson (which contained a return to the initial question and wrap-up). Students' increasingly sophisticated conceptual understanding confirms the learning progression intended by the curricula on "floating and sinking." In addition, it is evident that a large number of reasoning units were not rated at all on a conceptual level. These units largely involved single unsupported claims without conceptual content related to the explanation of floating and sinking, such as student observations of certain facts during experimental activities. As Figures 3 and 4 illustrate, the percentage of uncoded units in this study, similar to the levels of reasoning, seems to be dependent on the particular topic of the lesson.

An ANOVA on the development of conceptual understanding across the four lessons in Data Set 1 showed a significant effect for group, $F(1, 476) = 16.79$, $p < .001$, $\eta^2 = .034$; lesson, $F(3, 476) = 21.49$, $p < .001$, $\eta^2 = .12$; as well as a Lesson × Group interaction, $F(3, 476) = 4.14$, $p < .01$, $\eta^2 = .025$. Follow-up analyses revealed that the high instructional support group improved significantly across the curriculum, $F(3, 254) = 20.25$, $p < .001$, $\eta^2 = .19$, with significant effects for the comparisons of Lesson 1 with Lessons 3, 6, and 8 ($p < .05$). Similarly, the group of low instructional support improved across the four lessons analyzed, $F(3, 222) = 10.26$, $p < .001$, $\eta^2 = .12$, with significant effects for comparisons of Lesson 1 with Lessons 3 and 8 ($p < .05$). Because a higher percentage of reasoning units was rated

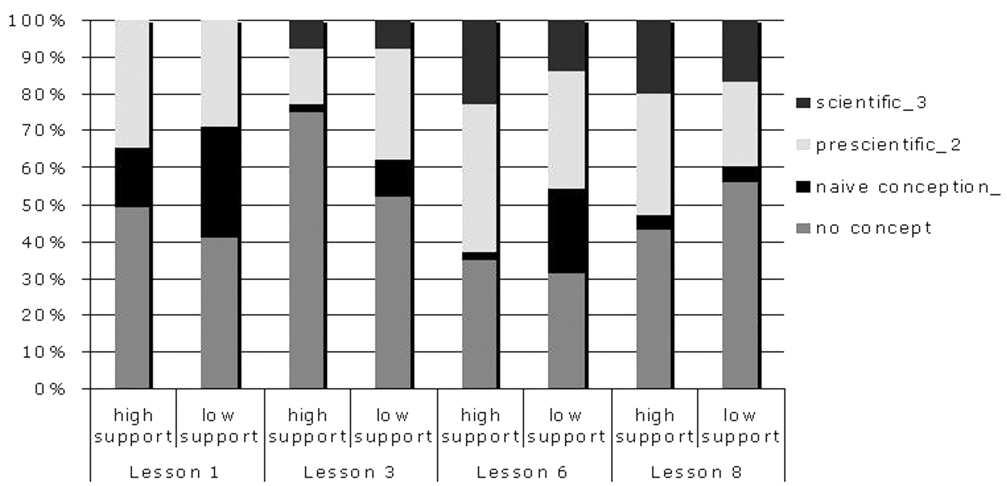

FIGURE 3 Levels of conceptual understanding by group and lesson for Data Set 1.

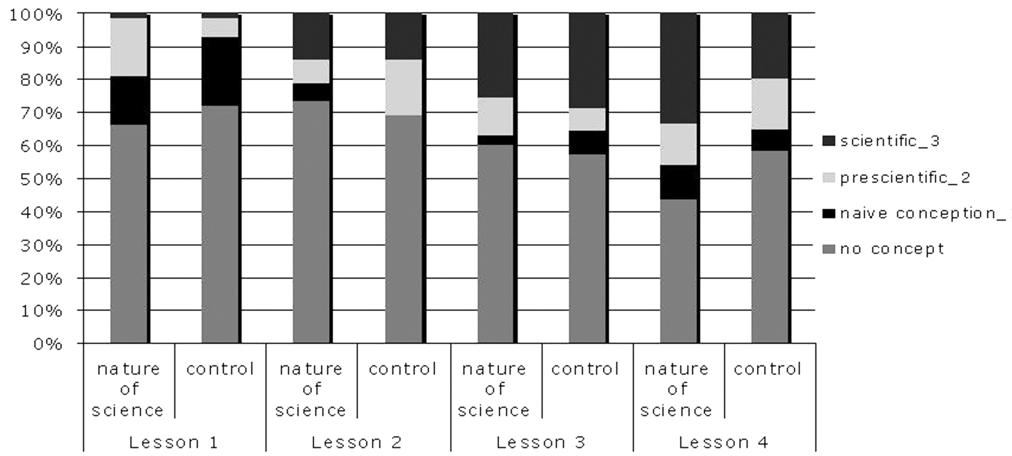

FIGURE 4 Levels of conceptual understanding by group and lesson for Data Set 2.

as nonapplicable in the low instructional support group, we performed an additional ANOVA that recoded the units that were not rated conceptually as zero instead of excluding these units from the analyses. When this is done, there is a significant effect of lesson, $F(3, 1011) = 15.93$, $p < .001$, $\eta^2 = .045$, and Lesson × Group, $F(3, 1011) = 5.37$, $p = .001$, $\eta^2 = .016$. Post hoc analyses showed that the average conceptual level of the high instructional support group was higher than the low instructional support group in Lesson 8 ($p < .05$), with M (high support) = 1.29 ($SD = 1.21$) and M (low support) = 1.02 ($SD = 1.22$). Because Lesson 8 is the last lesson of the curriculum, in which all the acquired concepts are taken together to bear on the explanation of the floating of an iron ship, between-group differences in this lesson are especially important for determining the effects of the experimental treatment.

A corresponding ANOVA for Data Set 2 showed a significant increase in the average conceptual level from first to last lesson, $F(3, 137) = 19.13$, $p < .001$, $\eta^2 = .26$, with no group or interaction effects. Follow-up analyses revealed that both the experimental and the control group significantly improved in their conceptual understanding comparing Lesson 1 with Lessons 2, 3, and 4, respectively ($p < .05$).

The Relationship Between Levels of Reasoning and Conceptual Understanding

The relationship between levels of reasoning and conceptual understanding is reported separately for each group. Because the level of reasoning and conceptual level were coded for each reasoning unit, Pearson's r was computed for these variables across all units of analysis. In Data Set 1, the average correlation between conceptual understanding and level of reasoning was low, with $r = .13$ (*ns*) for the high instructional support group and $r = .24$ ($p < .05$) for the low instructional support group. However, the correlation in Lesson 8 was higher and significant (high instructional support: $r = .307$, $p < .05$; low instructional support group: $r = .318$, $p < .05$).

In Data Set 2, the average correlation between level of reasoning and conceptual understanding was $r = .45$ ($p < .05$) for the experimental group and $r = .40$ ($p < .05$) for the control group.

A cross-tabulation of conceptual level and level of reasoning suggests that higher conceptual understanding is a necessary but not sufficient prerequisite for higher reasoning levels. Whereas the third and fourth reasoning levels were more likely to be associated with the higher prescientific and scientific conceptual understanding levels, rather than naive conceptions, higher conceptual understanding was not consistently associated with high reasoning level (reasoning at Levels 1 and 2 in 50% of the units in Data Set 1 and 25% in Data Set 2). Because the third and fourth reasoning levels presume that relations between variables are taken into account, a corresponding conceptual understanding of the relation between concepts seems to be a prerequisite.

Teacher Prompts and Levels of Reasoning

Teacher prompts for premise, claim, and claim justifications/backings are shown in Figures 5 and 6. Analyses performed on differences between these teacher prompts with regard to the reasoning level of the associated reasoning units concentrate on the code of teacher prompts for backing.

As Figures 5 and 6 indicate, the percentages for the respective codes vary extensively according to the type of group (high or low instructional support) in Data Set 1, whereas for Data Set 2, the percentages between the experimental and control group are similar. Because Data Set 1 involved an experimental variation based on teachers' use of cognitive support, or verbal scaffolding of student conceptual understanding, this variation was expected and may actually be regarded a type of implementation check of the intended experimental variation.

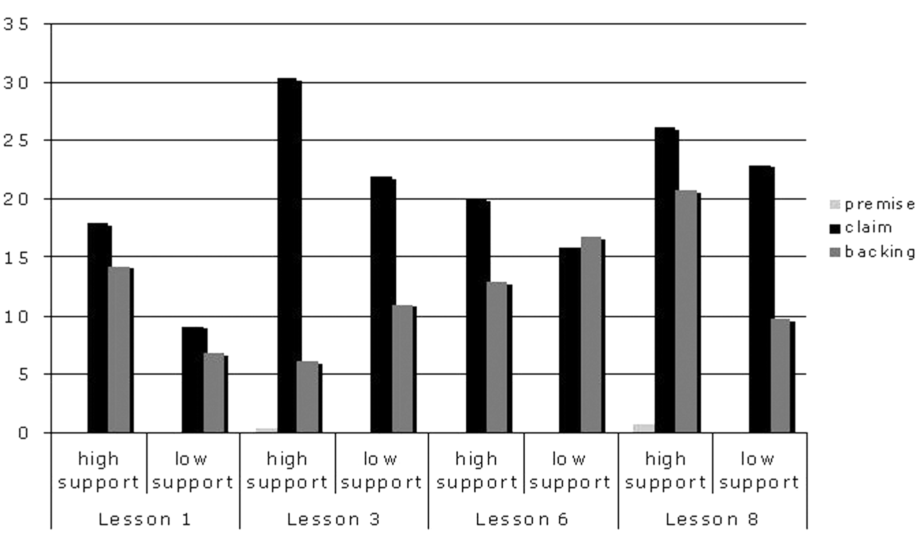

FIGURE 5 Percentages of teacher prompts by group and lesson for Data Set 1.

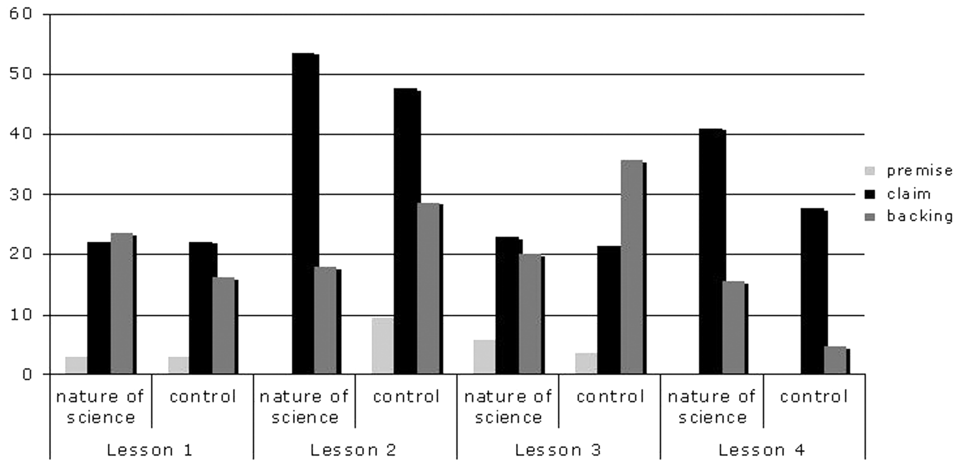

FIGURE 6 Percentages of teacher prompts by group and lesson for Data Set 2.

Across both data sets, prompts for claims occur at a higher rate than prompts for backings (mean percentage of prompts for claims was 26%, and for prompts for backing was 15.4%). Prompts for premises were rarely observed. Note that there may be multiple teacher codes within one unit.

Prompts for Claim Justification/Backing

The effects of teacher prompts on levels of reasoning were analyzed using the type of teacher contribution as an independent variable in addition to the variables of group and lesson used in previous analyses. In Data Set 1, there was a significant effect for prompt for backing, $F(1, 1003) = 70.77$, $p < .001$, $\eta^2 = .066$, as well as a three-way interaction of Prompt for Backing × Lesson × Group, $F(3, 1003) = 4.23$, $p < .01$, $\eta^2 = .012$. Follow-up analyses showed interaction effects of Prompt for Backing × Lesson for both the high instructional support, $F(3, 590) = 4.51$, $p < .01$, $\eta^2 = .022$, and low instructional support, $F(3, 413) = 2.69$, $p < .05$, $\eta^2 = .019$, groups, revealing that reasoning units with prompts for backing were associated with higher levels of reasoning.

Analyses of Data Set 2 revealed effects for teacher prompts for backing, $F(1, 385) = 86.80$, $p < .001$, $\eta^2 = .184$, and lesson, $F(3, 385) = 6.35$, $p < .001$, $\eta^2 = .047$, qualified by a Lesson × Prompt for Backing, $F(3, 385) = 4.92$, $p < .01$, $\eta^2 = .037$, interaction. Mean levels of reasoning were significantly higher if teachers prompted for backing than if they did not do so. Follow-up analyses revealed that there was an increase in mean level of reasoning between Lessons 1 and 2, and Lessons 1 and 4 ($p < .05$) when teachers prompted for backings for the experimental group, but not for the control group.

Further exploration looked at the relation between prompts for backing and levels of conceptual understanding. These follow-up analyses were only performed with the larger Data Set 1 as results obtained with Data Set 2 were not interpretable due to a restricted sample size. In Data Set 1, crosstabs of level of reasoning (1 to 4) and teacher prompt for backings (yes/no) in reasoning units with low conceptual understanding (Level 1) showed a significant association for both groups: high instructional support, $\chi^2(2) = 13.54$, $p = .001$; low instructional support, $\chi^2(2) = 7.70$, $p < .05$. In reasoning units with high conceptual understanding, however, we found no effect of teacher prompts in the group of low instructional support, $\chi^2(2) = 2.49$, *ns*, whereas for the group of high instructional support the association between level of reasoning and prompting remains significant, $\chi^2(2) = 14.72$, $p = .001$. For example, 56.3% of reasoning units were at Level 3 in the low instructional support group and 83.3% at Level 3 in the high instructional support group when teachers prompted for backing. When teachers did not prompt for backing, 45.5% of reasoning units in the low instructional support group and 60.4% in the high instructional support group were at Level 1. This suggests that the association between teacher prompts and level of reasoning is found independently of students' conceptual understanding within the respective reasoning units in the group of high instructional support, whereas this association only exists for reasoning units associated with low conceptual understanding in the group of low instructional support. Apparently, then, students in the former group did not rely on teachers prompting them for backings of claims as much as did the students in the experimental group to reason at high levels, especially concerning topics in which their conceptual understanding was sufficiently high.

DISCUSSION

The present study presents an exploration of use of the EBR Video Framework (see Furtak et al., this issue) to generate an in-depth analysis of classroom discourse on science topics and to explore the relations among EBR, conceptual understanding, and teacher prompts.

Evidence-Based Reasoning in Elementary School?

One basic question addressed in this study was whether and to what extent elementary school students' classroom discourse can be considered scientific discourse at all, characterized by theoretical claims that are explicitly supported or refuted by (empirical) evidence. Previous analyses of classroom discourse in secondary school (e.g., Jimenez-Aleixandre, Rodriguez, & Duschl, 2000) have found whole-class discussions to be dominated by claims that are unsupported by (empirical) evidence. The present study yielded similar results, as the majority (about 60%) of the reasoning units were scored on Level 1 for claim justifications (backings), that is, no evidence was produced to support or refute a claim. The percentage of reasoning units with rule-based reasoning was below 10% in all classrooms under study. If EBR was observed at all, it was predominantly based on single observations or data points.

As the introduction describes, the developmental literature on scientific reasoning indicates that elementary school children are capable of EBR when assessed individually in structured tasks and given a high extent of contextual support (Zimmerman, 2007). The quality of reasoning that is necessary to distinguish a confounded from a controlled experiment (Bullock & Ziegler, 1999) or a conclusive from an inconclusive test (Sodian et al., 1991) is at least relational, that is, would correspond to Level 3 reasoning in the EBR Video Framework. Thus, it appears that there is a discrepancy between the competencies observed in individual assessments of cognitive development and those observed in real-world classroom discourse.

Many factors may contribute to these differences in performance. One important source of difficulty is the conceptually challenging content domain used in this study, in contrast with the knowledge-lean tasks often used in individual assessments. Real-world scientific reasoning in science education takes place in the context of students' acquiring an understanding of complex and challenging natural phenomena. The present findings indicate that the level of scientific reasoning in elementary school students is low under such conditions and that the inquiry-based curricula used in the present study, although effective in raising the level of conceptual understanding, were not demonstrably effective in promoting the quality of EBR, even in the last lessons of the curriculum.

Development of Scientific Reasoning and Conceptual Understanding

There was a low to moderate correlation between conceptual level and reasoning level in the present samples. Partial independence between reasoning quality and conceptual understanding should be expected because it is possible, of course, to reason about false theories in a sophisticated way, just as it may be possible to grasp a scientifically adequate explanation without giving any evidence for it. In the data sets analyzed for this article, we observed only one of these two types of dissociations: High-level reasoning (including backing with evidence or rules) was associated with higher level conceptual understanding in terms of prescientific or scientific conceptions; low-level reasoning, however (including unsupported claims or claims supported with data only), occurred at all levels of conceptual understanding. Thus, a more mature understanding of the scientific concepts of density or buoyancy did not guarantee that reasoning also used evidence in support of claims. It is possible, of course, that students had in such cases failed to grasp the evidence for a scientifically adequate explanation and only partially attained conceptual understanding. It is also possible that students with a full conceptual understanding may have failed to support a conceptually sophisticated claim with appropriate evidence because students (and teachers) interpreted a justification as self-evident in the context of the ongoing discourse. We return to the effects of context in the upcoming sections.

It should also be noted that the correlations between conceptual understanding and level of reasoning varied across lessons and between the two data sets. The larger correlation for Data Set 2 may reflect the fact that the curriculum for Data Set 2 was shorter and more focused, whereas Data Set 1 included a wider range of concepts within a larger time frame.

In contrast to the average reasoning levels, the level of conceptual understanding increased from the first to the last lesson for both data sets, and there was little effect of the experimental variations (instructional quality or prior curriculum). The frequency of Level 2 and Level 3 conceptions was greater during the later lessons of the curricula than in the beginning, when Level 1 conceptions were most dominant. When the large number of reasoning units that could not be coded at a conceptual level is taken into account in the analyses, the experimental high instructional support group in Data Set 1 shows higher conceptual understanding, a finding that was expected given that the students in this group showed significantly higher levels of conceptual understanding in written assessments both directly after the end of the instructional program and at a follow-up test 1 year later. Nevertheless, the association between the achievement test data and the quality of conceptual understanding in instructional discourse as assessed with the discourse-based instrument appears to be only moderate, even though the coding schemes differentiating between the three basic levels of understanding were largely the same in both analyses.

Can inquiry-based curricula be effective in raising elementary school students' level of reasoning? Findings from Data Set 2 indicate that the preceding nature-of-science curriculum in the experimental group did have an effect on the average level of reasoning. Fifty percent of the reasoning units in the first "sinking and floating" curriculum lesson were evidence based, compared to 30% in the control group. It is likely that this was an effect of the prior curriculum because the two classrooms were carefully paralleled on cognitive ability and socioeconomic status and showed comparable conceptual understanding of floating and sinking in the written pretest. However, despite higher proportions of claim justifications, reasoning was still relatively low level (Level 2 phenomenological level, i.e., single observations or data), and it did not persist over the course of the floating and sinking instructional unit. Thus, it appears that students in the nature-of-science curriculum acquired some understanding of the importance of providing justifications for claims but that they failed to apply this understanding at higher levels of reasoning when challenged with difficult physical phenomena. Further research is necessary to determine whether high-quality reasoning can be transferred from one domain to another and can be fruitfully applied to the acquisition of conceptual knowledge at the elementary school level (see Kuhn, 2010, for a review of studies on argument skills in science and nonscience domains).

The Role of Teacher Prompts in Assessment Conversations

Although no overall effects of instruction on the level of reasoning were observed, there were clear effects on a microanalytic level of teacher prompts for evidence-based justifications of claims on students' reasoning levels. How did the teachers support the quality of reasoning in whole-class discussions? As expected, teacher prompts for claim justifications significantly raised the level of reasoning in associated reasoning units in both data sets, as students provided the requested backing (data, evidence, or rule). Follow-up analyses with Data Set 1 showed that teacher prompts were especially effective in the high instructional support group where the reasoning level was raised in reasoning units associated with both high and low conceptual understanding. This result thus underscores the importance of teacher prompting in establishing a classroom culture of scientific discourse within a scaffolded learning environment. Teacher scaffolds for the appropriate use of empirical evidence in class discussions thus seem to be a pivotal element in students' construction of scientific arguments. At the same time, adequate teacher prompts of the kind proposed by Duschl and Gitomer (1997) will provide teachers with opportunities to monitor and assess students' ability to use and coordinate evidence in ongoing scientific discourse.

The Evidence-Based Reasoning Video Framework as an Assessment Tool

An additional methodological aim of the present study was to take a first step toward a validation of the EBR Video Framework as an assessment tool. One concern in developing the coding system was that it might not be sensitive enough to capture reasoning competencies at the level of elementary school science. Given the low frequency of EBR observed in the present data sets, it was not possible to obtain reliable effects of instruction across time or between different instructional implementations. Thus, the present study is preliminary with respect to the sensitivity of the video framework to instructional variations. Note, however, that the video framework did capture the effects of specific teacher prompts on student reasoning. It is quite possible, and supported by the developmental literature, that only direct and specific interventions have effects on reasoning levels in elementary students, whereas more general contextual variations as they were implemented in the present designs do not.

Another concern was that the video framework may measure variations due to students' general cognitive abilities rather than their specific use of evidence in discourse. If this was the case, then we would expect a high correlation between reasoning level and conceptual level, because the coding for conceptual understanding distinguished between Level 1 concepts (naive conceptions) that frequently only consider single variables or observations and more sophisticated concepts on Level 2 (prescientific conceptions) and Level 3 (scientific conceptions) that are based on relations between constructs. Similarly, the coding for reasoning levels distinguished between single observations, relational, and rule-based reasoning that increased in cognitive complexity. The finding that the correlation between reasoning level and conceptual level was only low to moderate indicates that the two dimensions were separable in elementary school students' discourse and thus supports the validity of the instrument.

In sum, the present study indicates that the EBR Video Framework can be used productively toward an in-depth understanding of scientific reasoning in elementary school. Certainly, the use of a coding procedure for analyzing classroom discourse is associated with specific gains and limitations. Although the use of this video framework makes possible the quantification of classroom discourse, and thus a comparison of developmental trends and patterns over time and across experimental conditions of relatively large samples, it also means that the contribution of individual students to classroom discourse is not acknowledged the same way that it could be done in a qualitative analysis of discourse. That is, the classification of one unit of analysis as reasoning unit considered the collectively emerging discourse in a classroom rather than individual contributions. Nevertheless, by coding the teacher prompts and claims separately, a meaningful relation between teacher actions and student responses within reasoning units could be established. By laying a focus on the function of statements within the conversational flow, single statements were interpreted depending on preceding arguments. That is, a statement may have been considered a claim in one part of a discussion where no reference to the same premise (i.e., content or topic) had been made previously, whereas the same statement may have been coded as a form of backing, if it was made in support of or reference to a previously stated claim. Although we certainly cannot ensure that our functional interpretation of discourse is akin to individual students' interpretations, the reliable establishment of reasoning units was a useful means of segmenting discourse across different speakers. A variable necessarily missing from a coding scheme aiming at quantification and comparison, however, is the contextual embedding of utterances within the conversational flow. In this study, we did not focus on the evolving nature of discourse related to similar conceptual content or instructional contexts as it has been done in previous analyses of classroom argumentation (e.g., Osborne, Erduran, & Simon, 2004). For example, it would be very interesting to select sequences of discourse related to one topic, such as density or displacement, and compare how these topics were discussed depending on the goal of the lesson and their contextual embedding across different lessons.

The variations in level of reasoning for different lessons especially in Data Set 1 point to the important role of instructional progression and learning goals for levels of reasoning. Further analyses may thus take the EBR Video Framework as a starting point to select instances for in-depth comparisons of reasoning processes in discourse.

ACKNOWLEDGMENTS

This project was funded by the German Research Association DFG (grant numbers MO 942/2-1; SO 213/24-1; STE 539/13-1).

REFERENCES

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Bullock, M., Sodian, B., & Koerber, S. (2009). Doing experiments and understanding science. Development of scientific reasoning from childhood to adulthood. In W. Schneider & M. Bullock (Eds.), *Human development from early childhood to early adulthood: Findings from a 20-year longitudinal study* (pp. 173–197). Mahwah, NJ: Erlbaum.

Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert, & W. Schneider (Eds.), *Individual development from 3 to 12. Findings from the Munich Longitudinal Study* (pp. 38–54). Cambridge, UK: Cambridge University Press.

Bybee, R. W. (2002). Scientific literacy: Mythos oder Realität? In W. Gräber (Ed.), *Scientific literacy. Der Beitrag der Naturwissenschaften zur allgemeinen Bildung* (pp. 21–43). Opladen, Germany: Leske und Budrich.

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

Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist, 28*, 235–251.

Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology, 94*, 327–343.

diSessa, A. (2006). A history of conceptual change research. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 265–281). Cambridge, UK: Cambridge University Press.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Bristol, UK: Open University Press.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*, 287–312.

Dunbar, K. (1994). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.). *Mechanisms of insight* (pp. 365–395). Cambridge, MA: Harvard University Press.

Duschl, R. A. (2003). Assessment of inquiry. In J. M. Atkin & J. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41–59). Arlington, VA: NSTA Press.

Duschl, R. A. (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., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment, 4*, 37–73.

Erduran, S., & Jimenez-Aleixandre, M. (Eds.). (2007). *Argumentation in science education: Perspectives from classroom-based research.* Dordrecht, the Netherlands: Springer.

Grygier, P. (2008). *Wissenschaftsverständnis von Grundschülern im Sachunterricht* [Nature of science in elementary school classrooms]. Bad Heilbrunn, Germany: Klinkhardt.

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006) Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking." *Journal of Educational Psychology, 98*, 307–326.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). "Doing the Lesson" or "Doing Science": Argument in high school genetics. *Science Education, 84*, 757–792.

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children. Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology, 64*, 141–152.

Kuhn, D. (2010). Teaching and learning science as argument. *Science Education, 94*, 1–15.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.

Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how). In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology. Volume 2: Cognition, perception, and language* (pp. 953–994). Hoboken, NJ: Wiley.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60*(4, Serial No. 245).

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

Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A Review of the Research. *Journal of Research in Science Teaching, 29*, 331–359.

McComas, W. F. (1998). *The nature of science in science education*. Dordrecht, the Netherlands: Kluwer.

Möller, K. (1999). Konstruktivistisch orientierte Lehr-Lernprozessforschung im naturwissenschaftlich-technischen Bereich des Sachunterrichts [Research of constructivist learning and teaching processes in science and technical domains of science instruction]. In W. Köhnlein (Ed.), *Vielperspektivisches Denken im Sachunterricht* (pp. 125–191). Bad Heilbrunn, Germany: Klinkhardt.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council. (2001). *National science education standards*. Washington, DC: National Academy Press.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41*, 994–1020.

Sampson, V., & Clark, D. (2006). Assessment of argument in science education: A critical review of literature. In S. Barab, K. Hay, & D. Hickey (Eds.), *Proceedings of the 7th International Conference of the Learning Sciences* (pp. 655–661). Bloomington, IN: International Society of the Learning Sciences.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49*, 31–57.

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education, 28*, 235–260.

Simon, S., Osborne, J., & Erduran, S. (2003). Systematic teacher development to enhance the use of argumentation in school science activities. In J. Wallace & J. Loughran (Eds.), *Leadership and professional development in science education: New possibilities for enhancing teacher learning* (pp. 198–217). London: Routledge Falmer.

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition, 21*, 177–237.

Smith, C., Maclin, D., Houghton, C., & Hennessey, M. G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition & Instruction, 18*, 349–422.

Sodian, B., Jonen, A., Thoermer, C., & Kircher, E. (2006). Die Natur der Naturwissenschaften verstehen—Implementierung wissenschaftstheoretischen Unterrichts in der Grundschule [Understanding the nature of science—Implementing instruction on the nature of science in elementary school]. In M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms*. Münster: Waxmann.

Sodian, B., Thoermer, C., Kircher, E., Grygier, P., & Günther, J. (2002). Vermittlung von Wissenschaftsverständnis in der Grundschule [Instruction on the nature of science in elementary school]. *Zeitschrift für Pädagogik*, 45. Beiheft, 192–206.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62*, 753–766.

Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and independence. *Cognitive Development, 23*, 488–511.

Thoermer, C., & Sodian, B. (2002). Science undergraduates' and graduates' epistemologies of science: The notion of interpretive frameworks. *New Ideas in Psychology, 26*, 263–283.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Tytler, R., & Peterson, S. (2004). From "try it and see" to strategic exploration: Characterizing young children's scientific reasoning. *Journal of Research in Science Teaching, 41*, 94–118.

von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching, 45*, 101–131.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction, 15*, 317–419.

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