Journal of Educational Psychology
2006, Vol. 98, No. 2, 307–326

0022-0663/06/$12.00 DOI: 10.1037/0022-0663.98.2.307

Effects of Instructional Support Within Constructivist Learning Environments for Elementary School Students' Understanding of "Floating and Sinking"

Ilonca Hardy
Max Planck Institute for Human Development

Angela Jonen and Kornelia Möller
University of Münster

Elsbeth Stern
Max Planck Institute for Human Development

In a repeated measures design (pretest, posttest, 1-year follow-up) with 161 3rd-grade students, the authors compared 2 curricula on floating and sinking within constructivist learning environments, varying in instructional support. The 2 curricula differed in the sequencing of content and the teacher's cognitively structuring statements. At the posttest, both instructed groups showed significant gains on a test on understanding the concepts of density and buoyancy force as compared to a baseline group without instruction. One year later, the group of high instructional support was superior to the group of low instructional support on the reduction of misconceptions and the adoption of scientific explanations. Thus, instructional support within constructivist learning environments fostered elementary schoolchildren's conceptual change in the domain of physics.

Keywords: conceptual change, constructivist learning environments, elementary schoolchildren, physics concepts, structural support

CONCEPTUAL CHANGE IN SCIENCE LEARNING: THE CASE OF FLOATING AND SINKING

It is well documented that many phenomena in the physical world invite children to construct their own explanations spontaneously (Carey, 2000). However, at the same time, many of the resulting conceptualizations are partially or entirely inadequate from a scientific point of view, constituting what is referred to as misconceptions (for a recent overview, see Chi, 2005). This is, for instance, the case with children's explanations of why some objects float when immersed into water, while others sink. Interviews conducted by Smith and colleagues (C. Smith, Carey, & Wiser, 1985; C. Smith, Maclin, Grosslight, & Davis, 1997) and Möller (1999) with elementary as well as secondary school students showed that children's explanations often focus on one dimension only: They refer either to the mass of an object ("things that are light will float"), to its volume ("large things will sink"), or to its form ("everything with holes in it will sink"). Moreover, in many explanations, air is seen as an active force that pulls objects upward and water is seen as a force that sucks them downward. These explanations of floating and sinking are incommensurable with scientific explanations (based on the concepts of density and buoyancy force) since, rather than consider the relationship between object and surrounding fluid, the students focus on single properties of objects. Generally, the process of restructuring initial conceptions in order to build conceptions appropriate from a scientific point of view has been framed as conceptual change (e.g., J. Smith, diSessa, & Roschelle, 1993; Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001). In this article, we want to contribute to the question of how elementary school students' conceptual change with regard to floating and sinking may be effectively realized in early science education. We present a study on the long-term effects of two learning environments that allow discovery learning by hands-on experience with authentic material, but which differ in the degree of instructional support. It was investigated whether the sequencing of instructional content and frequency of cognitively structuring statements by the teacher will facilitate students' conceptual change in the long run.

Since scientific explanations of floating and sinking, initiated by the famous principle of Archimedes, were developed by physicists

Ilonca Hardy, Center for Educational Research, Max Planck Institute for Human Development, Berlin, Germany; Angela Jonen and Kornelia Möller, Institute for Early Science Education, Westfälische Wilhelms-Universität Münster, Münster, Germany; and Elsbeth Stern, Center for Educational Research, Max Planck Institute for Human Development, Berlin.

Angela Jonen is now at the Institute for Science Education, Bayerische Julius-Maximilians-Universität Würzburg, Würzburg, Germany.

This research was supported by the German Research Foundation (DFG) as part of the priority program "The Quality of Schools" (MO 942/1-1; STE 539/9-1). We thank Hella Beister and Rainer Roth for data coding and management as well as Jessica Stehling, Ulrike Austermann, Eva Blumberg, and Daniela Aufderhar for their assistance during the classroom lessons. We also thank Kevin Miller and Fritz Staub for their comments on an earlier version of this article. Finally, we thank the participating teachers and students in Münster, Germany, for their enthusiasm and support of this research.

Correspondence concerning this article should be addressed to Ilonca Hardy, Center for Educational Research, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: hardy@mpib-berlin.mpg.de

in a process of scientific discovery and justification, students cannot be expected to discover these principles spontaneously in their everyday life. Usually, however, students' participation in conventional science classes is not sufficient for their restructuring of conceptual understanding. Rather, students' misconceptions have proven to be quite resistant to change through instruction (Wandersee, Mintzes, & Novak, 1994). Misconceptions have been conceived of as fragmented, loosely connected pieces of knowledge (diSessa, 1988) or as a coherent pattern of notions that form consistent explanatory systems within domains (e.g., Vosniadou & Brewer, 1992) or even across domains (Chi, 2005). For example, Vosniadou (1994) argues that concepts are embedded in larger theoretical structures constraining them, involving knowledge revision at the specific theory level and the framework theory level. Conceptual change thus involves the restructuring of underlying concepts, which can take place in various ways depending on the particular aspect of the concept that is relevant for the respective explanatory context. For instance, an understanding of the concept of density requires the simultaneous consideration of the two dimensions of mass and volume, where mass needs to be conceived of as a continuous and measurable characteristic of the material world—a conception that young children often lack (C. Smith et al., 1985). In order to apply fully the concept of density in the context of floating and sinking of objects, a further conceptual shift needs to occur. Rather than its attribution to the characteristics of single objects, the relationship between object and surrounding fluid needs to be considered causal for objects' floating and sinking. Both the comparison of densities (of object and fluid) and the comparison of forces (of gravity and buoyancy) thus require the simultaneous consideration and integration of concepts.

Frequently, the topics of density and buoyancy force are introduced only in secondary school, based on the argument that students need to be able to grasp the formal aspects of the involved formulas such as proportions. Although particularly young children tend to adopt an undifferentiated concept of weight and density (C. Smith et al., 1985), even some secondary school students approach instruction with this type of commonsense theory (C. Smith et al., 1997). A study by Kohn (1993) showed similarities even between preschoolers and adults with regard to their inadequate strategies for judging an object's floating or sinking. Nevertheless, there is evidence that conceptual advancement may be reached by curricula addressing these concepts explicitly. Most of the curricula developed for early secondary school focus on conceptual change with regard to a differentiation of students' conceptualization of matter. For example, C. Smith et al. (1997) showed that employing matrices of squares with dots to illustrate the densities of different materials during an instructional unit helped many students learn to differentiate between weight and density and achieve an integrated understanding of density (see also C. Smith, Snir, & Grosslight, 1992). Lehrer, Schauble, Strom, and Pliggie (2001) had fifth graders model material kind in linear graphs after introducing them to the mathematical concepts of volume measure and similarity. Generally, the judgment of volume as a prerequisite to understand the relation of volume and mass does not seem to present difficulties to students of elementary school age (Halford, Brown, & Thompson, 1986); rather, within a curriculum on floating and sinking, students need to be given opportunities to discover that the amount of water displaced is dependent on an object's volume and not its mass. This understanding of water displacement may then be further differentiated, allowing for the comparison of densities and the determination of buoyancy. In our curriculum, we therefore focus on a basic understanding of water displacement, which then can support the construction of more complex scientific explanations of floating and sinking. Commonly, children are introduced in elementary school to a basic concept of material kind, that is, the realization that solid objects of the same material behave the same way when immersed in water. If, however, children were also introduced to the explanations for the behavior of different materials in water, thus receiving the opportunity to revise misconceptions early on, there is good reason to expect that they will be able to profit more from the formulas of density and buoyancy force treated in secondary school.

CONCEPTUAL CHANGE WITHIN A CONSTRUCTIVIST PERSPECTIVE

To be able to successfully restructure their concepts, students need to engage in a process of knowledge integration (Davis, 2003; Linn, 1995); that is, they need to link newly constructed scientific ideas to their current concepts, refining or dismissing them in the case of incommensurability. Conceptual change research has shown that this process of integration is by no means a sudden shift from a naive explanation to a scientifically adequate view but rather a gradual process of restructuring interrelated concepts, in the course of which misconceptions may reoccur depending on the particular context (e.g., Caravita, 2001; Tyson, Venville, Harrison, & Treagust, 1997; Vosniadou et al., 2001). Why and under what conditions does successful conceptual change occur? According to the pioneering work of Posner, Strike, Hewson, and Gertzog (1982), an essential condition of conceptual change is for learners to become dissatisfied with the conceptions they have, thus experiencing the need for new explanatory mechanisms. Central to this notion is the induction of cognitive conflict as a motor of conceptual change, starting with the initial conceptions a child may hold (Chinn & Brewer, 1993). For instance, in order to challenge the misconception that all floating objects contain air, one may confront children with a hollow, yet sealed object that sinks. Such kinds of confrontations are crucial for challenging plausible but inappropriate explanations, and as a consequence, students may become receptive to new ideas.

There is agreement among psychologists and educational scientists that the adoption of appropriate scientific explanations is a constructive process requiring the active cognitive engagement of individuals rather than a process of direct transmission. As a consequence, learning environments in science and mathematics that honor constructivist processes of knowledge acquisition have at their core the creation of opportunities for students' active sensemaking, allowing them to discover new principles and explanations of the topic at hand (Bransford, Brown, & Cocking, 1999; Cobb, 1994; Greeno et al., 1998). In a typical learning environment of this kind, a group of learners is confronted with complex, challenging, and authentic tasks and materials that encourage activities such as experimentation, exploration, and group generation of hypotheses and explanations. It is expected that students are able to self-regulate their learning to a degree that allows them to make sense of the material and physical resources according to their individual or common prior knowledge and capabilities. The social negotiation and common work on a task is thought of as a vehicle for the sharing of opinions, observations, and preexisting beliefs, adding to the anchoring of new knowledge.

Several studies suggest that students participating in a learning environment constructed according to the principles of constructivist learning achieve a higher degree of conceptual understanding in science and mathematics than students in environments of direct instruction (Christianson & Fisher, 1999; Staub & Stern, 2002; Tynjälä, 1999). However, many open questions remain as to how constructivist ideas of learning can be effectively realized at school. In an insightful, recent article, Mayer (2004) points out that the translation of principles of constructivist learning into learning environments has largely followed the simple formula "constructivism = hands-on activity" (p. 17). As a consequence, it may happen that, during an open-ended and thus highly self-directed learning activity, students do not actually reflect on the relevant concepts, even if the material provided is closely relevant to the phenomenon to be studied. If a setting is too complex and the tasks leave students too much freedom as to their selection and progression through them, students may be actively involved; nevertheless, they cannot arrive at the scientific conclusions intended because their sensemaking is not integrated into a conceptual framework. It is also at the point of addressing misconceptions that the question arises whether purely activity-oriented learning environments will be cognitively challenging enough for all students to enable them to proceed to the validation, correction, or final rejection of these deeply held misconceptions. Under the considerations outlined above, it seems likely that discovery learning environments can contribute to conceptual change; crucially, however, learning environments based on constructivist principles need to be designed in a way that supports students' cognitive activity apart from behavioral activity. That is, behavioral activity may be a vehicle for supporting cognitive activity, but it need not necessarily translate into the cognitive activity necessary for the successful conceptual integration of insights into prior understanding. An important issue for research, therefore, concerns the question of how aspects of learning environments may be successfully structured. Moreover, teachers willing to apply principles of constructivist learning in their classrooms must achieve more clarity about the conditions in which their feedback and prompts can help students to further conceptual understanding. We intend to shed light on these questions.

THE ROLE OF INSTRUCTIONAL SUPPORT IN CONSTRUCTIVIST LEARNING ENVIRONMENTS

Which are the dimensions relevant to instructional support within constructivist learning environments? Recently, the construct of scaffolding, originally framed by Wood, Bruner, and Ross (1976) and Vygotsky (1978), has been reexamined in its definition within complex science learning environments (Davis & Miyake, 2004; Hogan & Pressley, 1997). Reiser (2004) proposed the two mechanisms of *structuring* and *problematizing* as essential for the scaffolding of science tasks: While structuring works to reduce the complexity of a task, for example, by decomposition of a complex task and reduction of the degrees of freedom, problematizing subject matter works to relate students' work to a disciplinary framework, for example, by eliciting elaborations or by addressing discrepancies and disagreements in science discourse. Pea (2004) refers to the processes of *channeling* and *focusing* students' attention on relevant aspects of a task on the one hand, and to *modeling* more advanced solutions on the other hand, as critical aspects of scaffolding. The two key elements of instructional support, framed as scaffolding, thus seem to be (1) the structuring of tasks to allow students to remain focused on important aspects and (2) the support of students' reflection on their insights within a larger context of scientific reasoning.

Teacher prompts may be one way in which students can be supported to reflect on their scientific understanding. For example, Davis and Linn (2000) found that using prompts encouraging reflection significantly increased eighth graders' conceptual understanding of thermodynamics and light in a computer-supported learning environment (see also Davis, 2003). Scaffolding within an entire classroom may involve the teacher providing refocusing statements, challenges to assumptions or interpretations made by students, or advance organizers directing students' attention to certain aspects of a phenomenon to be investigated subsequently (Hogan & Pressley, 1997).

In sum, instructional support should help students to move along a sequence of conceptual development during which they can actively discover possible inadequacies of initial conceptions as well as convincing new explanations. As Reiser (2004) points out, it is a delicate task to create an optimum level of challenge through the provided support so that learners continue to engage actively in the learning process. In order to create this level of support, a teacher needs to be tuned to the needs of the learners in advance, anticipating their learning trajectories in the appropriate design of learning tasks. However, despite the need for individually tuned support of student reasoning processes, scaffolding environments can be achieved for an entire class if the social processes of prompting and modeling scientific reasoning support and reflect on individual sensemaking of the topic at hand (Hogan & Pressley, 1997).

THE CURRENT STUDY: EXPERIMENTAL VARIATION OF INSTRUCTIONAL SUPPORT ON THE TOPIC OF "FLOATING AND SINKING"

In our study, we investigated the effects of different degrees of instructional support within constructivist learning environments with respect to elementary school students' conceptual understanding of floating and sinking. We employed a repeated-measures design (pretest, posttest, and 1-year follow-up measure of conceptual understanding of floating and sinking), with two groups based on an experimental variation of instructional support (three third-grade classrooms each) and one baseline group which received the tests only (two third-grade classrooms). Our operationalization of instructional support was based on two central aspects of scaffolding, supporting (1) students' focusing on relevant aspects of the tasks and (2) students' reflection on their insights and scientific reasoning processes. It needs to be pointed out again that both learning environments designed in this study were based on constructivist principles of learning, with a high degree of students' active experimentation with relevant material in both experimental groups, allowing them to discover relationships between physical quantities and scientific principles by independent group work and to discuss hypotheses and discoveries in sessions with the entire class. Between the two experimental groups of high instructional support (HIS) versus low instructional support (LIS), the instructional time, the teacher, and the material for experimentation were not varied.

Variation of instructional support with regard to the task environment concerned the sequencing of instructional content into units which followed an anticipated path of conceptual development with regard to floating and sinking. While in both learning environments the investigation of the question of "Why does a large ship of iron float in water?" introduced the topic of floating and sinking, this topic was segmented into smaller units only in the group of HIS. That is, a complex learning environment allowing for experimentation with regard to different aspects of floating and sinking—such as material kind, density, water displacement, or buoyancy force on different learning sites—was used in both groups; however, it was prestructured into a sequence of experimental activities moving from more basic concepts such as material kind to the integrated concepts of the comparison of densities and relation of buoyancy force and gravity in the group of HIS. As Vosniadou et al. (2001) suggest, a central aspect of environments conducive to conceptual change is the consideration of the order in which concepts are acquired, based on the logical structuring of the involved domain. In the case of floating and sinking, the concept of material kind as an explanation of everyday life is fundamental to understanding more advanced concepts such as the comparison of densities. Moreover, many initial misconceptions may be rejected on the basis of evidence concerning the behavior of objects of the same material but different mass and volume in water. We therefore included lessons allowing these types of conclusions early on in the group of HIS, then advanced to the scientific concepts of density, water displacement, and buoyancy force, with an attempt to integrate these concepts in the last lesson. The group of LIS, in contrast, was free to use any experimental setting extensively within the context of researching the question about the iron ship. Here, students could work on activities with various foci such as density or water displacement, researching a certain self-generated issue of investigation; it is important to note, however, a complex question was not segmented in advance into smaller conceptual units.

With regard to the support of students' reflective processes, discussions involving the entire class were scaffolded to a higher degree by the teacher in the group of HIS. That is, the teacher intended to activate students cognitively with statements relating and contrasting ideas or hypotheses proposed by different students, by once more referring to earlier misconceptions, or by introducing a hypothesis or observation that the students themselves had not considered. In contrast, in the group of LIS, group discussions were much more student centered, with students themselves reacting to each other's statements and challenging each other's hypotheses. The role of the teacher was closer to that of organizational supervision, with a lower frequency of content-related prompts to students' reasoning processes.

HYPOTHESES

We hypothesized that through participation in a learning environment based on constructivist learning theories, third graders will show achievement growth in their understanding of floating and sinking that can be characterized as conceptual change. This means that, after the instructional unit, learners will show evidence of having rejected misconceptions and adopted explanations compatible with the scientific concepts of density and buoyancy force. Students are expected to make decisions about the floating and sinking of objects based on their differentiated conceptualizations with regard to density as a feature of substances, allowing the comparison of densities, and with regard to buoyancy force as resulting in a relational system of object and displaced water.

Due to the test-taking experience and informal exposure to the topic of floating and sinking, we expected some gains in the adoption of so-called explanations of everyday life, which are based on concepts of material kind and hollowness, even in the baseline group. However, neither those students' misconceptions nor their scientific explanations were expected to change significantly in comparison to students who attended the learning environments.

Our major hypothesis concerned the difference between the two groups of instructional support. We expected the group of HIS to outperform the group of LIS in both aspects of conceptual change, that is, by the decrease of misconceptions as well as by an increase of appropriate scientific explanations. Moreover, we expected the superiority of the group of HIS to be reflected by an integrated conceptual understanding at the individual level, where students will simultaneously hold misconceptions and appropriate conceptions to a lower extent than in the group of LIS. As a consequence of a more integrated understanding in the group of HIS, we expected long-term effects in the 1-year follow-up measure for the group of HIS, indicating fundamental conceptual change with a lower degree of reemergence of misconceptions.

METHOD

Design and Procedure

In a 3 (time) × 3 (group) repeated measures design, we investigated the effects of instructional support (group of HIS, group of LIS, baseline group with repeated test administration only) on third graders' conceptual understanding of floating and sinking. The intervention was based on two eight-lesson curricula which were preceded by a pretest and followed by a posttest one week after the end of instruction as well as a follow-up test one year after the instruction had taken place. In addition to the repeated Test on Floating and Sinking, we employed a transfer test on the application of principles of floating and sinking within three days after the administration of the posttest. At the posttest, the teachers of all classes were instructed not to revisit the topic of floating and sinking during the upcoming school year. One teacher of a class of LIS reported to have revisited the topic shortly after the posttest in the course of a class excursion on a ferry. The students were not told that there would be a 1-year follow-up test; only the teachers were informed shortly before the date.

In order to ensure the implementation of our instructional variation, all of the instruction was realized by Angela Jonen, who was educated as an elementary school teacher. Angela Jonen had not taught in any of the three schools before. The entire instructional unit was taught in 2-week intervals, alternating classes with HIS and LIS in one school at a time. At each school, the two classes were randomly assigned to either the condition of HIS or LIS. All of the instruction was videotaped.

Participants

A total of 161 third graders in eight intact classrooms participated in this study. The original sample size of 185 was reduced only by students who missed one of the tests or important parts of the curriculum in third grade, or else were not any more part of the same class in fourth grade at the

1-year follow-up. All classes came from elementary schools located in middle-class neighborhoods of a midsize town in Germany. The total of eight participating classes (six experimental classrooms, two baseline classrooms) were selected because they were composed of students with mixed social background (according to the school district), they showed a high variance in students' cognitive capabilities, and students were expected to have experienced discovery learning as well as traditional techniques in their instruction. That is, students were used to some forms of open, hands-on experimentation in their instruction but had not been taught exclusively by this method. We observed a total of three science lessons instructed by the respective classroom teachers before the beginning of the study to ensure that the type of instruction that students were used to was indeed comparable. The observational instrument focused on the teacher's employment of a variety of organizational forms (such as group work, experimental learning sites, and traditional instruction) and the treatment of complex science topics that students could investigate in a self-directed way.

The group of HIS comprised 66 students (26 girls, 40 boys; mean age in third grade of 9.02 years) from three classes, and the group of LIS comprised 59 students (28 girls, 31 boys; mean age in third grade of 8.95 years) from three classes. The baseline group was composed of 36 students (23 girls, 13 boys; mean age in third grade of 9.33 years) from two classrooms of the same school.

Curriculum

We developed two instructional units on the topic "Why does a large ship of iron float?" which were both based on constructivist principles of learning but varied in the degree of instructional support provided to students. Between the two instructional units we did not vary the material (Styrofoam, wood, metal, wax, modeling clay, etc.) that was available to students either as chunks of varying size and form or as everyday objects (metal plates, pins, wooden buttons, candles, nails, etc.) in order to enable their investigation of the topic of floating and sinking. Apart from everyday objects, the material for experiments included specifically designed objects such as cubes of the same volume but different material, objects of the same mass but different volume, and hollow and solid objects of the same mass and material. These objects were constructed to confront typically held misconceptions such as the concentration on the mass of objects. In the two instructional environments, we also employed the same experimental learning sites, which consisted of experiments on specific aspects related to floating and sinking such as water displacement. These sites were prepared in advance and were designed as independent student hands-on activities that addressed specific research questions, including directions for students' handling of material and hints to draw students' attention to specific features and outcomes of their investigations. The experimental learning sites concerned the topics of material kind, density, water displacement, and pressure or buoyancy force. Both instructional units comprised eight 90-min lessons. The fifth lesson in the sequence was the same in both instructional units; this lesson concerned the topic of water pressure and took place in a public swimming pool in order to enable students to experience water pressure on large objects and on their own bodies.

In the instructional unit of LIS, all of the objects and chunks of different material as well as the experimental learning sites were available to students throughout the entire instructional unit (except for the fifth lesson in the swimming pool). Students could use the material as they wanted to investigate their own research questions, which they derived from answering the overarching question of why a large ship of iron floats in water, in a self-directed way. They could also work on the experimental learning sites, which were designed for students to get ideas about how to adapt their own experiments, to draw attention to phenomena not yet investigated or perceived by the students, and to instigate discussion between students on explanatory mechanisms for floating and sinking. During students' experimental work, the teacher gave individual support to students concerning the process of investigation. She also provided written feedback on students' learning diaries in which they recorded their investigations, explanations, and further questions. It was up to the students to modify their research agenda depending on the insights they had gained during experimental work and group discussions. Students presented results of their investigations in small groups and in discussions with the entire class. The discussions were student centered; that is, students themselves would introduce new explanations, hypotheses, and problems encountered during their research, add to each other's interpretations, and challenge each other's hypotheses. The teacher enabled the process of discussion organizationally but only rarely contributed to discussions in terms of the investigated content.

In the instructional unit of HIS, the availability of objects of different material and experimental learning sites was determined by the specific subtopic to be investigated in the particular lesson. In this way, the complex research question of why a large ship of iron floats was partitioned into subordinate research questions, all of which were assumed to build toward an integrated understanding of floating and sinking. Lessons proceeded from the concepts of material kind; the differentiation of light and heavy material (or density), water displacement, and buoyancy force to the integration of these aspects in the relation between buoyancy force and gravity and the comparison between densities. Table 1 summarizes the sequencing of content in the HIS environment. While students worked on the experimental learning sites, the teacher would provide individual support as needed. During discussions with the entire class, the teacher frequently structured interaction on a content level, that is, attempted to cognitively activate students in making connections to their own knowledge base. For example, the teacher highlighted contradictory statements; asked for causal explanations, validations, or applications; or gave summaries of results. Table 2 summarizes the central characteristics of the LIS and HIS environments with respect to variation in students' work with the experimental learning sites and class discussions. In addition, the Appendix contains the coding scheme used to rate teacher and student discourse during class discussions, which served as an implementation check for differences in the two learning environments. This scheme provides insight into the exchanges that took place, including examples of teacher and student comments. We expected a higher frequency of teacher structuring statements in the group of HIS and a higher frequency of student contributions to other students' statements in the group of LIS.

Pilot Study

We conducted pilot studies in a total of four classrooms in order to optimize the implementation of the two curricula. Since we opted for the same experienced elementary school teacher, Angela Jonen, to manage both instructional variations, we needed to ensure that she could practice the variation of her teaching style. Her style of teaching is grounded in a constructivist view of learning, where she tends to hold back herself as much as possible during learning activities and group discussions in order to enable students' active construction of meaning. She thus tended to a style of teaching more in line with the curriculum of LIS, and she needed to get experience especially in the degree to which she was to provide cognitively structuring statements in the group of HIS (e.g., in terms of prompting for misconceptions or contrasting students' answers in group discussions). Optimization of the two curricula further concerned the optimal arrangement and wording of tasks for the experimental activities performed in both instructional groups. That is, since the instructional variation consisted only of the additional sequencing of tasks in the group of HIS, it was necessary to ensure for both curricula that students could work with the experimental learning sites on their own. Observations conducted by research assistants and one of the primary researchers in the pilot classrooms confirmed the differing implementations according to the degree of instructional support.

Table 1
Sequencing of Content in the High Instructional Support Environment

Lesson	Topic	Learning sites and activities
1	"Why does a large ship of iron float?"	Form initial hypotheses; work with solid objects; order objects according to floating and sinking in water
2	"What floats, what sinks?" (material kind)	Work with solid objects, address misconceptions; order objects according to material
3	"Why does wax float but metal sink?" (density)	Work with objects of same volume and different mass; differentiate mass and volume
4	"What happens to water when an object is immersed into it?" (water displacement)	Work with solid and hollow objects, varying in volume and mass
5	Excursion to the swimming pool: "What does the water do to objects?" (buoyancy force)	Work with large hollow objects and own body, experiencing buoyancy force
6	"What does the water do to objects?" (water displacement and buoyancy force)	Work with solid and hollow objects on buoyancy force and relation to water displacement
7	Relation of water displacement and buoyancy force	Work with modeling clay (modeling of boats) on the relation of water displacement and buoyancy force
8	Relation of buoyancy force and gravity: "Why does a large ship of iron float?"	Represent forces by children pulling on each other's hands; work with large hollow objects

Instruments

In order to address our hypothesis on the effects of instructional support, four instruments were administered. First, and most important, the Test on Floating and Sinking was administered to students as the pre-, post-, and follow-up test. Second, in order to obtain evidence of the construct validity of the Test on Floating and Sinking as a measure of students' conceptual understanding of density and buoyancy force, we administered a transfer test on floating and sinking shortly after the posttest. Third, in order to ensure that the instructional variation of HIS and LIS was actually implemented in the intended way, videotaped lessons were rated by trained persons not informed about the experimental variation with regard to teacher activities (implementation check). Fourth, the instructional discourse in each class was categorized according to criteria such as structuring statements by the teacher (implementation check).

Test on Floating and Sinking

Test Content

A test of students' explanations with regard to floating and sinking, comprised 36 items, was employed as pretest, posttest, and follow-up test. Three items were free-response items, while 33 were multiple-choice items that varied in the number of answers to be confirmed. Seventeen items were true-false items; that is, students were asked to judge the correctness of a single statement by checking it. Sixteen items were multiple-choice items with four to seven possible answers, and the participants were asked to confirm those they considered to be true. The test was developed on the basis of elementary school students' spontaneous explanations derived from interview data, including formulations regarding the concepts of material kind, density, and buoyancy force in child-appropriate language, which were adapted in several cycles of pilot testing. The use of child-appropriate language also means that we did not employ the unfamiliar terms of mass and volume; instead, we used the colloquial terms of weight and size for these concepts. In addition, the meaning of the new term of water displacement was introduced briefly before the test. The ordering of items in the test was determined by content considerations. Free-response items were presented in the beginning of the test to allow for maximum variation of student language; multiple-choice items were presented in a mixed order to ensure variation of item contexts. Additionally, the item formats requiring specific instructions were placed in the first half of the test so the teacher could address the entire class during explanations. The items were constructed to test two broad areas of understanding.

Table 2
List of Characteristics of High Instructional Support and Low Instructional Support Environments

Characteristic	High instructional support	Low instructional support
	Experimental learning sites	
Learning material	Differently shaped and sized material and everyday objects	Differently shaped and sized material and everyday objects
Content	Sequenced into eight consecutive units; availability only of learning sites concerning one subtopic	Availability of all experimental learning sites and possibility to design own experiments
Teacher	Managerial and contentwise structuring statements	Managerial and contentwise structuring statements
Students	Document their insights on handouts providing research questions and structuring comments for each learning site	Document their experiments and insights in a learning diary, using provided handouts as needed
Social forms of learning	Individual or group experimentation at learning sites with time limit for each learning site	Individual or group experimentation at learning sites with no time limit within the lesson; spontaneous group discussions
	Class discussions	
Students	Mainly react to teacher statements; free to react to other students	Free to initiate topics; need to call on other students
Teacher	Teacher-directed with contentwise structuring comments; topic is focused on content of each lesson	Teacher keeps conversation going
Time	About 50% of entire unit	About 50% of entire unit

1) Water displacement. A group of multiple-choice items (N = 9), separated into three contexts, focused on the topic of water displacement as a core concept from which other explanations of floating and sinking can be derived. In these items, the misconception of dependency of amount of water displacement on mass, not volume, is tested. Figure 1 shows a sample item confronting students with information about an object's mass and volume to determine the water displaced by it.

The three items of the second context showed four glasses, each of which contained a ball of the same volume but different mass than the others. It required students to draw lines for the respective water levels, given the water level for one ball in the first glass. The three items of the third context questioned students' judgment about which of two balls of different volume and mass would displace more water, with the mass of the objects displayed qualitatively by a balance beam.

2) Explanations of floating and sinking. The second group of items focused on students' explanations of floating and sinking. This group was further divided into two areas.

2a) Explanations of specific situations. This group of items focused on students' explanations of specific objects' behavior when immersed into water. The objects, which were always shown to students and depicted in the test, were chosen to be extreme cases with regard to volume, mass, or form. For each item, it was made clear that objects were first entirely immersed into water before releasing them. This group included seven multiple-choice items and the three free-response items. The multiple-choice items incorporate typical misconceptions of elementary school students (such as reference to an object's mass, volume, form, or air as an active force) as well as explanations derived from everyday life (such as reference to an object's material or its hollowness, or mentioning of the role of water in floating) and scientific concepts (such as water displacement, density comparisons, and buoyancy force). In order to avoid confronting students with unfamiliar vocabulary, all explanations were tested carefully to be understandable to elementary school students, yet at the same time scientifically appropriate in the relationships expressed. Words such as *lighter/heavier than* or *pushes up,* which were used repeatedly in instruction, were offered in a correct and an incorrect version within one item. For each of the situations, the students first had to decide whether the object would sink or float in water; then they had to check the explanation(s) judged as most appropriate. The seven situations to be judged by students concerned the floating or sinking of a wooden button, a flat piece of metal, a flat piece of Styrofoam, a piece of metal wire, a pin, a wooden block, and an iron ship.

The two items shown in Figure 2 exemplify the construction of the multiple-choice items. The item referring to the wooden button includes prominent misconceptions regarding the mass and form of the object (Answers 2 and 3), an explanation of everyday life (Answer 5), and the scientifically correct explanations with the concepts of density (Answer 6) and buoyancy force (Answer 1), along with one worded in the opposite direction (Answer 4). The scientific explanation based on a comparison of the object's mass with its displaced water (Answer 6) refers to the total amount of water that may be displaced by an object due to its volume, thus allowing a comparison of mean densities of object and water. Therefore, an object floats if the amount of water that can be displaced weighs more than the object itself, or the density of the water is greater than the density of the object. In contrast, the actual amount of water that is displaced by a floating object weighs the same as the object, a relationship referred to as the principle of Archimedes. Students confronted with this item first need to make a choice about the object's behavior in water and then need to check the explanations they find most convincing. Since the conceptual change literature shows that the integration of new explanations into existing frameworks is a continuous process, the possibility of multiple answers was given to students. The item referring to the piece of metal wire is constructed according to similar principles. Two scientific explanations refer to density, one worded in a correct way (Answer 5) and one worded in the opposite direction and therefore incorrect (Answer 3).

The three free-response items referred to questions explaining what happens if (1) a large tree trunk, (2) a wooden board with holes, and (3) a ship of iron are immersed into water. Again, students had the opportunity of giving multiple explanations for the objects' floating or sinking in water.

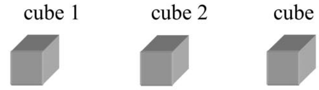
2b) Generalized concepts. A third group of items consisted of a total of 17 true-false statements with respect to explanatory mechanisms of floating and sinking or principles generalized beyond specific objects. For example, students were asked to check whether statements such as "Heavy objects cannot float in water," "Things without air cannot float in water," "Some types of material always float in water," "Sinking objects are being sucked down by the water," "Water pushes more against large than small objects," or "Any hollow object will float in water" are correct. This group of items also included five items testing students' generalized understanding of the concept of material kind, asking them to make predictions about floating or sinking of a certain material, given information about its size (see Figure 3).

Here you can see different cubes. Underneath each cube there is written how much it weighs. All of the cubes sink in water.

Which cube displaces the most water?

Check one answer:

cube 1 cube 2 cube 3



20g 80g 50g

❒ cube 1
❒ cube 2
❒ cube 3
❒ All of the cubes displace the same amount of water.

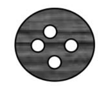
Note. All test items were printed in color.

Figure 1. Sample item of the content area of water displacement.

The following objects are being immersed into water. What will happen?

1. Check the right answer.
2. Then check *all of the correct* explanations.

The *wooden button* will ❒ sink ❒ rise to top



❒ because it is pushed up by the water strongly enough.
❒ because it is so light.
❒ because it has got holes.
❒ because it is pushed down by the water.
❒ because it is made of wood.
❒ because the displaced water weighs more than the button.

The *metal wire* will ❒ sink ❒ rise to top

❒ because it holds on to something.
❒ because it is so long and so thin.
❒ because the displaced water weighs more than the metal wire.
❒ because it is made of metal.
❒ because the displaced water weighs less than the metal wire.
❒ because it is so light.
❒ because it is not pushed up strongly enough by the water.

Figure 2. Sample item of the content area of explanations of specific situations of floating and sinking.

Test Administration

All tests were administered by Angela Jonen in the respective classrooms. All students of a class took the tests at the same time. The students were presented with a test booklet. The test administrator read out loud sample items of different test formats (multiple-choice items, true-false items, and free-response items) and explained how to proceed in answering them. The administrator demonstrated the material or the situation described in the sample items, with all of the objects described in the test located at the front desk. Students were free to look at and pick up all the items described in the test throughout test taking as they wished. In the pretest, the term *water displacement* was explained to students by showing a cube, immersing it into water, and asking whether there will still be water in the place where the cube was now. Then, the term *displaced water* was introduced to specify the amount of water which was replaced by the cube. Thus, no explanations of the physical principles underlying water displacement were given. Also, the teacher demonstrated how an object was entirely immersed in water in order to clarify the situations in which students should decide on an objects' floating or sinking when immersed in water. There were no demonstrations of objects' actual floating or sinking, nor were the students allowed to try out objects' behavior in water. Students were free to work on the items at their own pace. Each student took the test with the same order of items. Altogether, it took approximately 60 minutes to administer the entire test. Although the students took the test very seriously, its administration took place in a relaxed atmosphere.

Consider each object, then check the correct answer.

This large iron nail sinks.

This small iron nail
❒ sinks ❒ floats

Figure 3. Sample item on the concept of material kind.

Coding scheme for free-response items. We scored answers on free-response items on three broad levels of conceptual understanding, with the highest level separated into two sublevels. Each student answer could be scored on multiple levels.

1) Misconceptions. This level of explanation was scored if students answered by a one-dimensional focus on, for example, mass, volume, or shape, or if they considered air an active force. Also, students who indicated with a question mark that they did not have an explanation were included in this category. Examples of student answers to the question of what happens if a wooden board with holes is immersed into water include "sinks because it has holes and it is really flat," "sinks because the water will come through the holes and lies on the board and then the board will be heavier," and "floats because the weight of the board is spread out."

2) Explanations of everyday life. This level of explanation includes answers 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 as explanations of floating and sinking. Examples of student answers are "floats because it is wooden," "floats because water weighs more," and "floats because this board with holes is made of wood. And it doesn't matter whether the wooden board is with or without holes."

3a) Scientific explanations based on one concept (Scientific Explanations 1). Answers that consider one of the physical quantities of density, water pressure, or buoyancy force were scored on this level of conceptual understanding. Examples of student answers include "floats because the water wants back its place" and "floats because the water pushes the board up."

3b) Scientific explanations based on more than one concept (Scientific Explanations 2). Answers that consider two of the physical quantities of density, water pressure, or buoyancy force simultaneously, or in which all three aspects are integrated into one explanation, were scored on this highest level of explanation. Examples of student answers are "floats because the holes don't matter and because the displaced water weighs more than the wooden board," "floats because the displaced water weighs more than the board, it is pushed up," "floats because the board displaces as much water as it weighs," and "floats because it is lighter than the same amount of water."

While Levels 1, 2, and 3b correspond to the explanations offered by multiple-choice items, there is no level corresponding to Level 3a in multiple-choice items. Since this lower level of scientific explanations turned out to be a frequent answering level of students, we decided to split the level of scientific explanations into the two sublevels 3a and 3b. Interrater reliabilities on each of the four levels of rating were $> .95$, with two trained persons each rating all of the free-response items.

Scoring

We scored the test with a total score indicating students' overall answering patterns on the test and separate scores for different levels of conceptual understanding of floating and sinking and water displacement. These scores were computed separately for multiple-choice and free-response items.

1) Separate levels of conceptual understanding. With the separate scores of conceptual understanding, we can evaluate to what extent students gave up misconceptions and adopted everyday life or scientific explanations independently of each other. Since conceptual understanding at each level may vary depending on the selection or production of respective answers, we computed separate scores for multiple-choice (MC) and free-response (FR) items. This decision is justified by the low to medium correlations between ratings of free-response items with respective conceptual levels of multiple-choice items, ranging from .13 to .38 in the pretest, from .30 to .42 in the posttest, and from .25 to .39 in the follow-up test. Apart from two pretest correlations, all correlations are significant ($p < .001$).

For the content area of "water displacement," we computed a sum score of correct answers (Water Displacement, MC), with a maximum of 9 points and Cronbach's alpha (pretest, posttest, follow-up test) = .79, .89, .88, respectively. Here, we did not differentiate between different levels of conceptual understanding, since a correct answer on these multiple-choice questions means that a student correctly considered volume instead of mass (incorrect answer) as the determining factor for water displacement, thus equivalent to the level of scientific explanations.

Answers on multiple-choice and true-false items of the content area of "explanations of floating and sinking" were grouped together as three sum scores reflecting different levels of conceptual understanding (Misconceptions, MC; Explanations of Everyday Life, MC; Scientific Explanations, MC). Here, answers on multiple-choice items of "explanations of specific situations" were assigned to one of the three scores of conceptual understanding, depending on which level of conceptual understanding they expressed. For example, Answers 2 and 3 on the sample item in Figure 2 would be assigned to the score of misconceptions. Answers on the group of "generalized concepts" were also assigned to one of the three scores of conceptual understanding, with the items testing material kind as part of the score of Explanations of Everyday Life. Reliability analyses of scales showed satisfactory results, with Cronbach's alpha (pretest, posttest, follow-up test)/Misconceptions = .59, .62, .75, respectively; Cronbach's alpha (pretest, posttest, follow-up test)/Explanations of Everyday Life = .57, .62, .63, respectively; and Cronbach's alpha (pretest, posttest, follow-up test)/Scientific Explanations = .70, .78, .81, respectively. The maximum scores for the three multiple-choice scores were 17 points for Scientific Explanations, 11 points for Explanations of Everyday Life, and 32 points for Misconceptions.

Answers on free-response items were summed up on four levels of understanding, with separate sum scores for Misconceptions (FR), Explanations of Everyday Life (FR), Scientific Explanations 1 (FR), and Scientific Explanations 2 (FR). Scientific Explanations 1 includes answers that address only one physical quantity, while answers scored as Scientific Explanations 2 integrate two or more physical quantities. Note that for the scores of Scientific Explanations and Explanations of Everyday Life we expect an increase due to time and instruction, while for the scores of Misconceptions we expect a decline.

2) Answering pattern on the entire test. Although students' construction of new scientific explanations and their rejection of or adherence to misconceptions may be seen as separate processes, it is important to assess the extent to which students were able to integrate their new explanations into a coherent conceptual framework. Therefore, we computed a score that reflects students' integrated conceptual understanding of floating and sinking, which is based on students' simultaneous correct rejection of misconceptions and adoption of new scientific explanations. Students' response patterns on all multiple-choice items and true-false items of the content areas of water displacement and explanations of floating and sinking as well as the answers on the three free-response items on explanations of floating and sinking were combined to a sum score of Integrated Conceptual Understanding (ICU). On multiple-choice items of explanations of specific situations, it is only if none of the misconceptions and at least one of the scientific answers within one item have been checked that the student scores a point. That is, students scored a point on these items only if they showed evidence of adhering to a scientifically acceptable conceptual framework that does not include the simultaneous inclusion of misconceptions. In the score of ICU, the explanations of everyday life are not being considered since, although they are correct, they may or may not be combined with scientific explanations. Similarly, students scored a point on free-response items if they had produced an explanation on Level 3 (scientific explanations) without mentioning a misconception (Level 1) at the same time. Again, Level 2 explanations were not considered in this scoring.

On the items of water displacement and generalized concepts, students needed to show the correct answering pattern for the entire item context (e.g., the context of material kind) to score one point. Scoring all of the 36 items of the test in this stringent way resulted in a maximum of 17 points.

Cronbach's alpha (pretest, posttest, follow-up test)/ICU (MC) was .48, .82, .81, respectively. Due to the stringent scoring procedure, the variance in the pretest, where students are often confronted with explicit explanations of floating and sinking for the first time, is more restricted than in the posttest and follow-up test.

Transfer Test

In order to test students' ability to employ the scientific principles underlying objects' floating and sinking in fluids for the explanation of situations beyond those covered in instruction, we designed a transfer test with a total of eight items. This tests relies less on verbal expressions than the Test on Floating and Sinking used in the repeated measures design and thus serves as a further indicator of students' conceptual understanding of floating and sinking.

Test Content

1) Application of the concept of density. A total of five multiple-choice items tested students' conceptual understanding of density by focusing on the comparison of densities of one object and two liquids or else of two liquids, thus going beyond the comparison of densities of object and water assessed in the Test on Floating and Sinking. The application of the concept of density may be assessed by items requiring little verbal explication for the answers. Rather, students' conceptual understanding of density may be inferred from their consideration of the correct variables underlying a comparison of densities. For example, one item asked students to determine how much a cube of a given material will weigh if it sinks in oil but floats in water, given the masses of a cube of water and a cube of oil, respectively, of the same size as the object. For the correct answer, students need to know that the number for the mass of the object needs to be smaller than the mass of the cube of water and larger than the mass of the cube of oil (see Figure 4). During the instruction, "a cube of water" had been referred to as the amount of water displaced by a cube of a certain material.

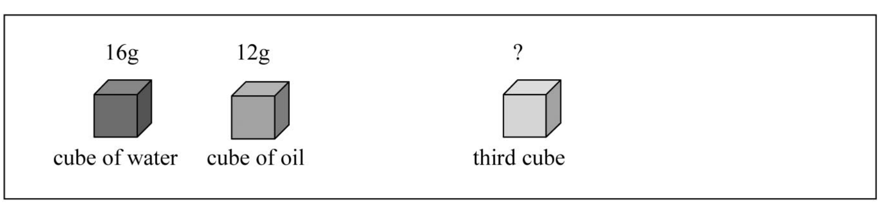
Another item asked students to check which one of four suggestions would help someone intending to "let an egg float in salty water": the addition of more water, the addition of more salt (correct answer), the addition of more of the same salty water, or the use of a smaller egg than the one displayed in the item. Figure 5 displays an item in which students had to decide which of the displayed cubes would sink in water, given the mass and volume of a cube of water. In this item, multiple answers could be checked.

2) Application of the concept of buoyancy force. Three multiple-choice items focused on students' explanations of floating and sinking with a focus on buoyancy force. The situations described in the items went beyond single objects' floating and sinking as covered during the curriculum by either involving the comparison of two floating objects or the comparison of two states of an object. Similar to the group of multiple-choice items on explanations of floating and sinking in the Test on Floating and Sinking, students had to check all explanations they judged as being correct for a specific situation, with possible answers covering misconceptions and scientific explanations. Specifically, one item asked students to decide which of two depicted boats of the same mass but different volume would be able to carry a treasure better and to justify their response. Another item asked students to choose answers which best explained why a piece of clay on a fishing rod feels lighter when immersed in water. The third item asked students to choose answers that best explained why fish in water can rise to the top when they increase their gas bladder (see Figure 6). All three items allowed multiple answers to be checked.

Test Administration

The Transfer Test was administered within three days after the first posttest, that is, within two weeks after the instructional unit. The teacher

You can see three cubes of the same size. They do not weigh the same. The first cube is made of water and the second one is made of oil.



The third cube floats in water, but it sinks in oil.

How much does it weigh?

Check the correct answer.

- ❒ 10g
- ❒ 11g
- ❒ 14g
- ❒ 17g
- ❒ 18g

Figure 4. Sample item from the Transfer Test on the application of density.

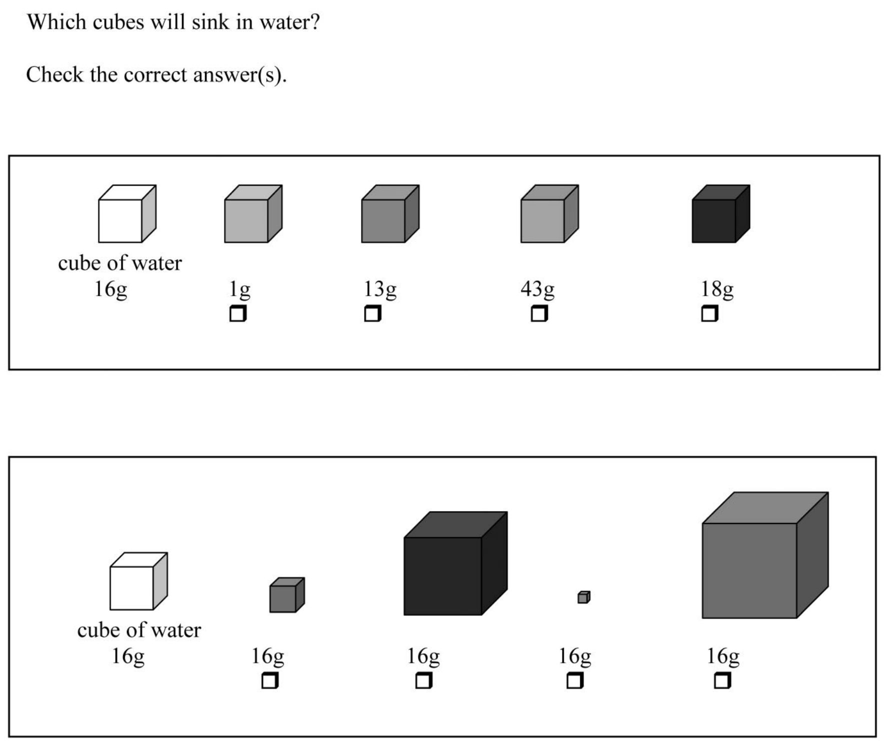


Figure 5. Sample item from the Transfer Test on the application of density.

proceeded the same way as she had done for the Test on Floating and Sinking, explaining all of the situations and item formats and demonstrating all of the experimental situations described in the test booklet using real objects and material (such as oil floating on water). Then students proceeded to work on the test booklets independently in a prescribed sequence of items, taking their individually needed time.

Scoring

We computed separate sum scores for misconceptions and scientific explanations. Since items on the Transfer Test focused on students' application of the concepts of density and buoyancy force in a wider context, no answers involving explanations of everyday life were provided. The sum score for Scientific Explanations (TR) was computed by summing up correct answers of all eight items, with a maximum of 19 points and α = .53. The correlations of this score with the sum score of Scientific Explanations (MC) of the Test on Floating and Sinking are .46 (posttest) and .32 (follow-up test). Similarly, the sum score for Misconceptions (TR) consists of the total of adopted misconceptions provided in the eight items of the Transfer Test, with a maximum of 18 points and α = .54. The correlations of this score with the sum score of Misconceptions (MC) of the Test on Floating and Sinking are .58 (posttest) and .33 (follow-up test). Since half of the items of the Transfer Test were true multiple-choice items with only one correct answer, we did not compute a score of ICU. The medium correlations between scores of both tests for each conceptual level (Misconceptions and Scientific Explanations) suggest that the answers provided in multiple-choice items of the Test on Floating and Sinking may be regarded as a valid indicator of students' conceptual understanding of floating and sinking.

Implementation Checks

Because both instructional units were administered by the same teacher, it had to be ensured that the teacher actually implemented the characteristic dimensions of the instructional variation. We employed two implementation checks to ensure the experimental variation of HIS versus LIS in each of the six experimental classes.

(1) Each of the 48 videotaped 90-min lessons was rated by a total of 12 independent raters as either a lesson of HIS or LIS. As criteria, the raters had received a list of characteristic elements of instruction intended for the two curricula, which were based on the two manipulated domains of the sequencing of content and the teacher's cognitive structuring of classroom discourse.

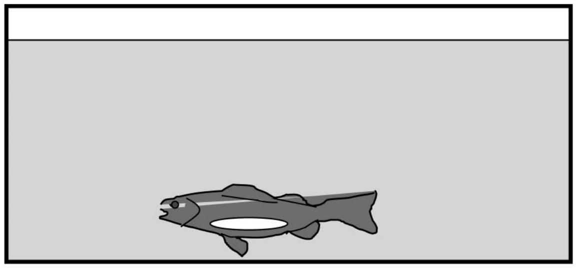
(2) Thirty percent of tapes were randomly chosen to be categorized according to the types of verbal utterances by students and teachers during class discussions, thus excluding independent work at the experimental learning sites. The drawing of videos ensured that all of the lessons in the instructional sequence and all of the classes were represented equally. Verbal utterances were categorized in 10-s sequences, with the possibility of multiple codes per unit. Teacher statements were coded on whether they actively structured conversation on a content level, for example, by addressing misconceptions or requesting student explanations; whether they passively kept the conversation going, for example, by repeating student statements; or whether they concerned managerial issues. Note that none of the statements intended as structural comments concerned explicit explanations on a content level. Rather, they intended to focus students' attention on relevant aspects of investigated phenomena and to model reasoning about scientifically relevant relationships. Student statements were rated as to their active contribution of content, their contribution of content as a reaction to a teacher statement, and their contribution of content as a reaction to another student's statements. The coding scheme was exhaustive regarding teacher and student statements. The Appendix summarizes the specifications of the coding scheme for the rating of teacher and student discourse.

RESULTS

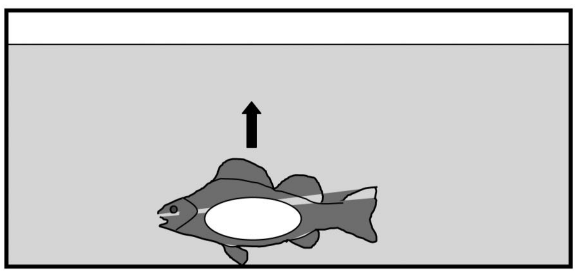
Implementation Checks

For our first implementation of the overall categorization of lessons as either part of the curriculum of HIS or LIS, the interrater

Fish can float in water without moving their fins. Each fish has a gas bladder. Fish can make their gas bladders large or small.



When a fish makes its gas bladder large, it rises to the top.



Why is this so?

Check all the correct answers.

- ❒ The gas pulls the fish to the top.
- ❒ The fish becomes larger and will be pushed up more strongly by the water.
- ❒ The fish becomes larger, and therefore the water will push less strongly.
- ❒ With the larger gas bladder, the fish becomes lighter.
- ❒ With the larger gas bladder, the fish is lighter than the water displaced by it.

Figure 6. Sample item from the Transfer Test on the application of buoyancy force.

reliability was .94 with a master coder. For our second implementation check of ratings of teachers' and students' verbal statements during class discussions, we performed a multivariate analysis of variance (MANOVA) for the entirety of coded statements at 10-s intervals and chi-square tests for single codes. There was a significant effect for group differences at the multivariate level, Wilks's $\lambda = .83$, $F(7, 2821) = 82.24$, $p < .001$, $\eta^2 = .17$, with significant differences between the LIS and HIS environments for all of the coded statements using chi-square tests. Table 3 summarizes descriptive statistics and results of chi-square tests for single codes. It is important to note that teachers in the HIS condition gave cognitively structuring statements twice as frequently as did teachers in the LIS condition (relative to the total number of coded 10-s intervals), while they provided managerial comments about half as often as did teachers in the LIS classrooms. As expected, students in the LIS condition showed higher relative frequencies for reactions to other students' contributions and for independent content contributions than students in the HIS condition, while students in the HIS condition more often reacted to teacher statements.

Overall, our two implementation checks underline that in the six experimental classrooms the intended variation of instructional support with regard to the sequencing of content and the structuring of social interaction was implemented according to the designed curricula.

Students' Achievement Gains on the Test on Floating and Sinking

In a first step, differences between the two experimental groups and the baseline group were examined with regard to students' overall answering pattern on the Test on Floating and Sinking on the pretest, posttest, and follow-up test. In order to explain and further analyze overall differences in answering patterns, we then proceeded with analyses of separate conceptual levels of under-

Table 3
Results of Chi-Square Tests of Teacher and Student Statements at 10-s Intervals

Variable	Statement type	Relative frequencies HIS (N = 1,567)	LIS (N = 1,262)	χ^2	p	d
Teacher	Structuring	.63	.28	337.83	.000	.76
	Passive	.12	.08	10.61	.001	.14
	Managerial	.15	.29	85.43	.000	.35
Student	Active	.06	.15	55.0	.000	.37
	Reaction to student	.06	.20	120.02	.000	.23
	Reaction to teacher	.48	.33	71.11	.000	.31
	Other	.04	.05	9.36	.002	.05

Note. HIS = high instructional support; LIS = low instructional support.

standing across the three points of measurement. In the following sections, we first report on tests to identify possible pretest differences between groups. Then the results of overall tests of mean comparison are reported, followed by tests on differences between points of measurement (pretest to posttest; pretest to follow-up test). We first address differences between instructional groups and then between instructional groups and the baseline group. If appropriate, results of *t* test pertaining to gains within groups are also reported. The chosen α level is .05.

Integrated Conceptual Understanding of Floating and Sinking

A univariate analysis of variance ensured that there were no pretest differences between the groups of HIS, LIS, and the baseline group (BS) on the score of ICU, $F(2, 158) = 0.24$, *ns*. To test the effects of instructional support on students' conceptual understanding of floating and sinking, we performed a 3 (time) $\times$ 3 (group) ANOVA with the score of ICU. Figure 7 illustrates mean differences across points of measurement for each group. There were significant effects of time, $F(2, 158) = 104.25$, $p < .001$, $\eta^2 = .40$; Time $\times$ Group, $F(4, 316) = 10.09$, $p < .001$, $\eta^2 = .11$; and group, $F(2, 158) = 12.46$, $p < .001$, $\eta^2 = .14$.

Follow-up ANOVAS with scores of differences between points of measurement (pretest to posttest; pretest to follow-up test) revealed that the two instructional groups did not significantly differ from each other in their gains from pretest to posttest; however, they significantly differed from the baseline group (effect of instruction: $F(2, 158) = 15.27$, $\eta^2 = .16$, $p < .001$; planned contrast HIS–BS: $p < .001$, $d = 1.24$; planned contrast LIS–BS: $p < .001$, $d = 1.11$). In a comparison of gains from pretest to follow-up test, there was a significant effect for group, $F(2, 158) = 10.49$, $p < .001$; $\eta^2 = .12$, with a significant difference between gains of the two instructional groups ($p < .05$, $d = .38$), as well as between the two instructional groups and the baseline group (planned contrast HIS–BS: $p < .001$, $d = .98$; planned contrast LIS–BS: $p < .05$, $d = .61$).

T tests of mean values for each group revealed that while the group of HIS did not change significantly from posttest to follow-up test, there was a significant decline of mean scores in the group of LIS, $t(58) = 2.29$, $p < .05$, $d = .20$. *T* tests further showed that the baseline group also gained significantly from pretest to posttest, $t(35) = -2.65$, $p < .05$, $d = .54$, and in the long run in the follow-up measure, $t(35) = -3.30$, $p < .01$, $d = .71$. However, the gains from pretest to follow-up test for the group of HIS and the group of LIS were larger (HIS: $t(66) = -10.61$, $p < .001$, $d = 1.75$; LIS: $t(59) = -7.94$, $p < .001$, $d = 1.32$). In sum, analyses of the score of ICU showed a significant advantage of the group of HIS as compared to the group of LIS in the follow-up test, while both experimental groups showed significantly higher means than the baseline groups both at the posttest and the follow-up test.

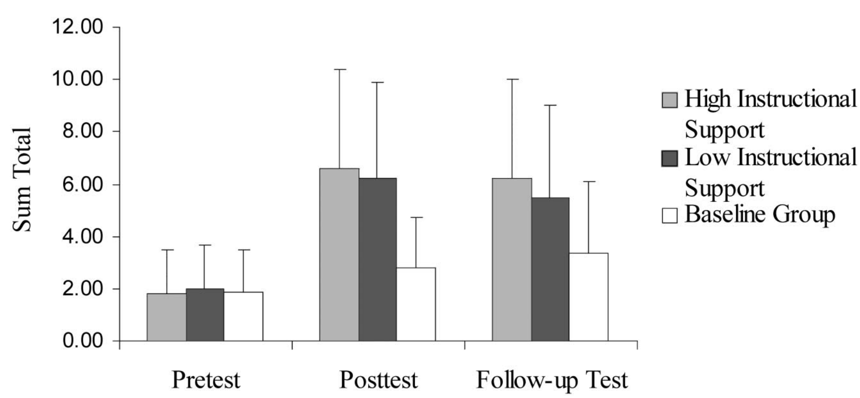


Figure 7. Means of score of Integrated Conceptual Understanding (ICU) by time and group.

Analyses of Separate Conceptual Levels

For further investigation of the composition of answering patterns combined within the score of ICU, we performed analyses with eight separate scores of conceptual understanding separately for multiple-choice and free-response items.

Univariate ANOVA for pretest differences between groups showed no significant differences for the scores of Water Displacement, Misconceptions (FR, MC), Explanations of Everyday Life (FR, MC), Scientific Explanations 1 (FR), and Scientific Explanations 2 (FR). There was a significant difference in mean pretest scores of Scientific Explanations (MC), $F(2, 158) = 3.36$, $p < .05$, $\eta^2 = .04$. Scheffé post hoc comparisons showed a difference between the groups of HIS and LIS ($p = .051$), with a greater mean of the group of HIS, M (HIS) = 7.20, M (LIS) = 5.73.

A 3 (time) × 3 (group) MANOVA with the dependent measures of Water Displacement, Misconceptions (MC, FR), Explanations of Everyday Life (MC, FR), Scientific Explanations (MC), Scientific Explanations 1 (FR), and Scientific Explanations 2 (FR) resulted in significant effects of time, Wilks's $\lambda = .21$, $F(16, 143) = 31.22$, $p < .001$, $\eta^2 = .78$; time × group, Wilks's $\lambda = .44$, $F(32, 286) = 4.58$, $p < .001$, $\eta^2 = .34$; and group, Wilks's $\lambda = .57$, $F(16, 302) = 5{,}90$, $p < .001$, $\eta^2 = .24$. Tables 4 and 5 list means and standard deviations for the multiple-choice items and free-response items, respectively, for the three experimental groups at three points of measurement.

Univariate follow-up analyses with each of the dependent measures were performed. The respective statistics are listed in Tables 6 and 7. There were significant gains across points of measurement for all dependent variables, and there were significant interactions of Time × Group for all dependent variables except Explanations of Everyday Life (MC) and Water Displacement (MC). Table 8 lists results of ANOVAs for differences between points of measurement (pretest–posttest and pretest–follow-up test), which were performed to determine the difference in gains between the three groups relevant to our hypotheses.

Differences From Pretest to Posttest

As may be seen from Table 8, the two instructional groups show significant differences only in their gains on Scientific Explanations 1 (FR). The group of HIS outperformed the baseline group in all measures except for the scores of Explanations of Everyday Life (MC, FR). The group of LIS showed a similar pattern; however, they also outperformed the baseline group with regard to the production of Explanations of Everyday Life (FR), while they did not differ from them in the production of Scientific Explanations 1 (FR). The differences in the reduction of misconceptions and the adoption of scientific explanations between the two experimental groups and the baseline group can be considered very large, with Cohen's d ranging from .89 to 1.71.

T tests for single groups showed that the baseline group did not change significantly with regard to mean scores of Water Displacement, Misconceptions (MC), Explanations of Everyday Life (FR), and Scientific Explanations (FR); however, there was a significant decrease from pretest to posttest on free-response answers rated as Misconceptions (FR), $t(35) = 2.84$, $p < .01$, $d = .36$, as well as an increase on Explanations of Everyday Life (MC), $t(35) = -3.4$, $p <$

Table 4
Means (and Standard Deviations) of Separate Levels of Conceptual Understanding (Multiple-Choice Items) by Time and Group

		Pretest			Posttest			Follow-Up test		
Content	Conceptual level	HIS (N = 66)	LIS (N = 59)	BS (N = 36)	HIS	LIS	BS	HIS	LIS	BS
Water Displacement	Scientific Explanations (MC)	3.46 (2.35)	3.63 (2.10)	3.47 (2.56)	5.24 (3.15)	5.41 (3.20)	4.03 (2.69)	5.65 (2.91)	4.83 (3.20)	4.58 (3.17)
Explanations of Floating and Sinking	Misconceptions (MC)	14.21 (3.70)	12.93 (3.94)	13.08 (3.94)	8.48 (5.05)	8.85 (5.04)	13.25 (4.28)	8.62 (4.43)	9.14 (4.36)	12.08 (4.20)
	Explanations of Everyday Life (MC)	7.18 (2.30)	6.51 (1.87)	6.72 (2.39)	9.21 (1.70)	8.54 (2.06)	8.03 (1.80)	8.07 (2.25)	7.69 (1.98)	8.25 (1.86)
	Scientific Explanations (MC)	7.20 (3.21)	5.73 (3.50)	7.11 (3.23)	11.47 (3.88)	10.41 (3.65)	8.36 (3.31)	10.29 (4.46)	9.37 (3.91)	8.58 (3.10)

Note. HIS = high instructional support; LIS = low instructional support; BS = baseline group; MC = multiple choice.

Table 5
Means (and Standard Deviations) of Separate Levels of Conceptual Understanding (Free-Response Items) by Time and Group

Conceptual level	Pretest HIS (N = 66)	Pretest LIS (N = 66)	Pretest BS (N = 66)	Posttest HIS	Posttest LIS	Posttest BS	Follow-up test HIS	Follow-up test LIS	Follow-up test BS
Misconceptions (FR)	2.62 (1.29)	2.32 (1.06)	2.58 (1.23)	.76 (1.04)	.86 (1.01)	2.14 (1.20)	.79 (.89)	1.37 (1.16)	1.92 (.97)
Explanations of Everyday Life (FR)	1.04 (1.33)	1.32 (1.06)	1.11 (1.33)	1.71 (1.19)	2.46 (1.61)	1.19 (1.08)	1.35 (1.13)	1.49 (1.33)	1.72 (1.22)
Scientific Explanations 1 (FR)	.26 (.55)	.27 (.54)	.08 (.26)	.85 (1.03)	.44 (.58)	.28 (.60)	.82 (1.04)	.86 (.94)	.31 (.61)
Scientific Explanations 2 (FR)	.05 (.26)	.02 (.12)	.00 (.00)	1.21 (1.44)	.92 (1.06)	.00 (.00)	1.08 (1.18)	.66 (.81)	.00 (.00)

Note. HIS = high instructional support; LIS = low instructinal support; BS = baseline group; FR = free response.

.01, d = .63, and on Scientific Explanations (MC), $t(35) = -2.43$, $p < .05$, d = .38. Thus, the experience of test taking apparently led to conceptual advances even in the baseline group.

Differences From Pretest to Follow-Up Test

There were significant differences with regard to the decline of Misconceptions (MC, FR) and the gain in Scientific Explanations 2 (FR) between the two instructional groups. At the follow-up test, the group of HIS showed significantly fewer misconceptions both in multiple-choice and free-response items than did the group of LIS, and students produced significantly more answers rated as Scientific Explanations 2. Generally, differences between instructional groups with regard to misconceptions seem to have increased over time, while differences with regard to the production of scientific explanations seem to have shifted from a focus on only one quantity (Scientific Explanations 1) at the posttest to a focus on two or more quantities (Scientific Explanations 2) at the follow-up test.

T tests within groups revealed that the group of LIS showed a significant increase from posttest to follow-up measure on the scores of Misconceptions (FR), $t(58) = -3.26$, $p < .01$, d = .47; thus, students regained many of the misconceptions they had already resolved a year earlier at the posttest. The group of HIS, in contrast, did not change significantly from posttest to follow-up test with regard to the production or adoption of misconceptions. With regard to answers rated as Scientific Explanations 2, the group of LIS showed significant gains from pretest to follow-up test, $t(58) = -6.15$, $p < .001$, d = 1.36, as did the group of HIS, $t(65) = -7.20$, $p < .001$, d = 1.41. The baseline group did not produce any answers rated as Scientific Explanations 2 at any point of measurement. The groups of LIS and HIS showed a significant increase in answers rated as Scientific Explanations 1 (FR) (LIS: $t(58) = -6.67$, $p < .01$, d = .79; HIS: $t(65) = -3.78$, $p < .001$, d = .69), while the baseline group did not significantly improve from pretest to follow-up test.

In comparisons of pretest and follow-up test scores, the two instructional groups and the baseline group show similar gains on scores of Explanations of Everyday Life (FR, MC), with Cohen's d ranging from .14 to .31. On the score of Scientific Explanations 1 (FR), the groups of HIS and LIS still outperform the baseline group. On the score of Water Displacement (MC), only the group of HIS shows greater gains than the baseline group. *T* tests for single groups showed that from pretest to follow-up test both instructional groups and the baseline group gained significantly on Explanations of Everyday Life (MC) (HIS: $t(65) = -2.50$, $p < .05$, d = .39; LIS: $t(65) = 4.47$, $p < .001$, d = .62; BS: $t(35) = -3.65$, $p < .001$, d = .73). On free-response items of Explanations of Everyday Life (FR), however, gains are significant from pretest to posttest for only the two instructional groups (HIS: $t(65) = -3.01$, $p < .001$, d = .53; LIS: $t(58) = -4.76$, $p < .001$, d = .86),

Table 6
Analyses of Variance for Levels of Conceptual Understanding (Multiple-Choice Items)

Source	df	Water Displacement F	η^2	p	Misconceptions F	η^2	p	Explanations of Everyday Life F	η^2	p	Scientific Explanations F	η^2	p
Between subjects													
Group	2	1.31	.016	.27	6.93	.081	.001	2.61	.032	.077	4.35	.052	.014
Error	158												
Within subjects													
Time	2	32.28	.128	.000	51.14	.245	.000	42.47	.212	.000	65.53	.293	.000
Time × Group	4	2.31	.028	.058	10.35	.116	.000	1.94	.024	.103	5.06	.060	.001
Error	316												

Table 7
Analyses of Variance for Levels of Conceptual Understanding (Free-Response Items)

Source	*df*	Misconceptions *F*	η^2	*p*	Explanations of Everyday Life *F*	η^2	*p*	Scientific Explanations 1 *F*	η^2	*p*	Scientific Explanations 2 *F*	η^2	*p*
Between subjects													
Group	2	12.97	.141	.000	3.72	.045	.026	9.82	.111	.000	17.77	.184	.000
Error	158												
Within subjects													
Time	2	86.32	.35	.000	12.23	.072	.000	14.58	.084	.000	42.06	.210	.000
Time × Group	4	9.95	.112	.000	5.40	.064	.000	2.53	.031	.041	10.24	.115	.000
Error	316												

while the baseline group also showed gains in a comparison of pretest and follow-up test, $t(35) = 3.24$, $p < .001$, $d = .48$. With regard to Scientific Explanations 1 (FR), *t* tests for single groups showed that while both instructional groups gained from pretest to follow-up test (HIS: $t(65) = 10.21$, $p < .001$, $d = .69$; LIS: $t(58) = 7.44$, $p < .001$, $d = .80$), the baseline group did not show significant changes on this variable.

To sum up analyses of scores of separate conceptual levels, there was a clear advantage of a high degree of instructional support with regard to students' more frequent production of scientific explanations with respect to one quantity in the posttest. The long-term effects of instructional support were evident with regard to both misconceptions and scientific explanations. In the follow-up test, students of the group of HIS significantly differed from students of the group of LIS on the less frequent selection and production of misconceptions and the more frequent production of scientific explanations with respect to two or more scientific concepts. While from posttest to follow-up test the group of LIS significantly increased their production of misconceptions, the means of the group of HIS did not change. Additionally, there is a long-term impact of instruction on the score of Water Displacement, where the group of HIS outperformed both the group of LIS and the baseline group. It is interesting that, over the long run, the two instructional groups did not differ from the baseline group on both the production and selection of Explanations of Everyday Life. The differences between the groups witnessed in the overall answering pattern on the score of ICU are apparently produced by differences with regard to Misconceptions (MC, FR) and Scientific Explanations 2 (FR) in the follow-up test, while in the posttest the superiority on Scientific Explanations 1 (FR) in the group of HIS did not contribute enough variance to affect the overall score of ICU.

Group Differences on the Transfer Test

For the Transfer Test, separate sum scores of Misconceptions and Scientific Explanations were computed. Means and standard deviations for the three groups were *M* (HIS) = 7.03 (*SD* = 3.22), *M* (LIS) = 7.35 (*SD* = 3.26), and *M* (BS) = 10.42 (*SD* = 2.72) for the score of Misconceptions and *M* (HIS) = 13.32 (*SD* = 2.96), *M* (LIS) = 12.00 (*SD* = 2.70), and *M* = 8.78 (*SD* = 2.35) for the score of Scientific Explanations. A MANOVA with the factor of group (LIS, HIS, BS) and the scores of Misconceptions and Scientific Explanations showed a significant effect of group, Wilks's $\lambda = .68$, $F(4, 312) = 16.31$, $p < .001$, $\eta^2 = .17$. Univariate ANOVAs revealed a significant group effect for both dependent variables, with $F(2, 157) = 32.13$, $p < .001$, $\eta^2 = .29$ for Scientific Explanations and $F(2, 157) = 15.08$, $p < .001$, $\eta^2 =$

Table 8
P Values and Cohen's d for Planned Contrasts With Mean Gains Between Points of Measurement

	Pretest–posttest						Pretest–follow-up test					
	LIS	HIS	HIS	BS	LIS	BS	LIS	HIS	HIS	BS	LIS	BS
Conceptual level	*p*	*d*	*p*	*d*	*p*	*d*	*p*	*d*	*p*	*d*	*p*	*d*
Water Displacement (MC)	.98	.01	.01	.45	.01	.46	.08	.32	.09	.34	.89	.03
Misconceptions (MC)	.07	.29	.01	1.15	.01	.89	.03	.39	.01	.97	.01	.58
Misconceptions (FR)	.07	.21	.01	1.28	.01	.89	.01	.59	.01	.90	.35	.22
Explanations of Everyday Life (MC)	.99	.01	.11	.33	.12	.33	.55	.11	.24	.23	.54	.14
Explanations of Everyday Life (FR)	.12	.26	.09	.45	.01	.80	.64	.07	.36	.22	.19	.31
Scientific Explanations (MC)	.48	.12	.01	.93	.01	1.07	.48	.13	.07	.38	.02	.51
Scientific Explanations 1 (FR)	.02	.42	.05	.43	.90	.03	.87	.02	.14	.36	.11	.40
Scientific Explanations 2 (FR)	.18	.23	.01	1.65	.01	1.71	.02	.40	.01	1.76	.01	1.60

Note. HIS = high instructional support; LIS = low instructional support; BS = baseline group; MC = multiple choice; FR = free response.

.16 for Misconceptions. Planned contrasts confirmed a significant difference between the instructional groups of HIS and LIS and the baseline group (HIS–BS: $p < .001$, $d = 1.71$; LIS–BS: $p < .001$, $d = 1.28$) and significantly higher scores for the HIS group compared to the LIS group ($p < .01$, $d = .47$) for the score of Scientific Explanations. For the score of Misconceptions, the two instructional groups differed significantly from the baseline group but not from each other (HIS–BS: $p < .001$, $d = 1.14$; LIS–BS: $p < .001$, $d = 1.03$).

DISCUSSION

The major concern of this study was to contribute to the question of how best to support elementary school students' conceptual understanding of complex science topics within classroom environments based on constructivist ideas of learning. Our learning environments allowed for discovery learning by hands-on experience with authentic material and encouraged sensemaking through peer interaction and classroom discussion, yet differed in the degree of instructional support. Under the condition of LIS, students could themselves decide when and how long to make use of experimental material within a given time frame, and the classroom discussions were student centered. Under the condition of HIS, students' experimental activities were sequenced into activities addressing basic concepts such as material kind to addressing more complex concepts such as the relationship between buoyancy force and water displacement. During classroom discussions, the teacher repeatedly requested clarifications and justifications of student ideas, addressed misconceptions by empirical evidence, or fostered processes of generalization (Reiser, 2004).

In accordance with conjectures put forward by Mayer (2004), our results showed that elementary school students' conceptual understanding of floating and sinking can be optimized by instructional support provided through the sequencing of instructional content and the frequency of cognitively structuring statements by the teacher. In the long run, students in the group of HIS developed a more coherent understanding of why some objects float in water while others sink than did students in the group of LIS. With regard to the score of ICU, students in the group of HIS could keep their level of conceptual understanding from posttest to follow-up test, while students in the group of LIS showed a significant decline. Among students in the group of LIS, misconceptions reemerged to a significant extent between posttest and follow-up test, while mean values in the group of HIS did not change. Similar patterns were also found for the scores of Water Displacement and Scientific Explanations 2, where two of three physical quantities—density, water displacement, or buoyancy force—must be taken into account simultaneously. A comparison of gains in producing scientific explanations from pretest to follow-up test revealed no differences between both instructional groups for Scientific Explanations 1, but larger gains in the group of HIS for Scientific Explanations 2. The superiority of students in the group of HIS with regard to the integration of physical concepts in scientific explanations is also indicated by their higher scores on the Transfer Test where the concepts of density and buoyancy force needed to be applied in new contexts.

Is it conceivable that our test scores really reflect a more coherent conceptual understanding in students of the HIS group? Several arguments may be put forward in support of this claim. First, the correlations between our separate scores of conceptual understanding on the Test of Floating and Sinking and the Transfer Test, assessing the quality of students' explanations for contexts not treated during the curriculum, suggest that we assessed a similar construct in both tests. Second, the variety of students' explanations on free-response items for explanations scored as Scientific Explanations 1 and 2 shows that students did not memorize fragments of explanations provided during instruction. Had students actually memorized rules, the superiority of the group of HIS in the follow-up test with regard to Scientific Explanations 2 would have been highly unlikely. In addition, our implementation check shows that although the teacher focused students' attention on relevant aspects of investigated phenomena under the condition of HIS, she never explicated scientific rules herself. Thus, the learners of the group of HIS did not get better opportunities for rote-leaning teacher statements than students in the group of LIS. Third, conceptual understanding was assessed not only in terms of the production of scientific explanations but also in terms of a decline in the production of misconceptions. The group of HIS was clearly superior in this aspect.

The relatively small differences between both instructional groups in the posttest and the clear superiority of the group of HIS one year later suggest that different kinds of conceptual change occurred during the period in which the curriculum was applied and during the following year when no instruction on floating and sinking was encountered. By running experiments and by interacting with their peers, learners from both groups restructured their knowledge about water displacement, extended their repertoire of explanations, and learned that some of their previously constructed explanations were inappropriate. However, while students in the group of HIS integrated the knowledge acquired during the curriculum into a coherent conceptual system where inappropriate explanations would be replaced by scientific ones, learners in the group of LIS may not have contrasted their existing misconceptions with new explanations to a sufficient degree (see Davis, 2003; Linn, 1995). Thus, a reemergence of misconceptions was observed in the follow-up test, although students still remembered the scientific explanations adopted during the curriculum, as the high scores on the multiple-choice items and the production of Scientific Explanations 1 indicate. In other words, under the condition of LIS learners may not have realized the incompatibility of different explanations to the same degree as did learners under the condition of HIS. When working on the follow-up test, the explanations developed during the curriculum competed with the earlier misconceptions, thus leading to an inconsistent response pattern. This pattern of reoccurring misconceptions is commonly found in research on conceptual change, indicating that previously adopted explanations were not well integrated into existing conceptual frameworks (e.g., Vosniadou et al., 2001).

In the current design, the sequencing of content and the structuring of classroom discussions are confounded; therefore, the single contribution of each dimension remains as yet unknown. Whether there are additive effects of both dimensions, or whether either dimension is sufficient for inducing conceptual change, must be clarified by further studies with experimental variations of structuring classroom practice. It is also plausible that the interaction between sequenced content and cognitively structuring teacher statements produced the effectiveness of the HIS environment for conceptual change. That is, the sequencing of content probably led to classroom discussions that were focused on specific aspects of investigations experienced by all students of a class. Therefore, the teacher may have been able to provide cognitively structuring statements in response to individual student conjectures and explanations which in effect also targeted the level of conceptual understanding of most students in the class. In contrast, the teacher's structuring statements in the group of LIS, occurring with significantly lower frequency, needed to address individual students' ideas to a greater degree, due to the variety of experiments and content on which students had worked during prior independent work. It is likely that these teacher statements were not as effective in furthering the conceptual understanding of students who had worked on different instructional material prior to the discussion.

In an extension of Hogan and Pressley (1997), we suggest that the provision of scaffolding need not be an intervention based on individual tutoring but may be implemented within an entire classroom; the sequencing of content, contributing to the similarity of students' conceptual background, will likely add to the effectiveness of teacher-provided scaffolding. In our study, elementary school students with little prior domain-specific knowledge profited from participation in a learning environment with instructional support. Further research needs to address the extent to which learners of different age groups and cognitive preconditions such as greater metacognitive control will profit from scaffolded learning environments to a similar degree.

CONCLUSIONS

Why should children as young as nine years of age learn about the concepts of density and buoyancy force? Data from our baseline group showed that the adoption of answers based on explanations of everyday life is a conceptual advancement during elementary school that does not need to be supported by instruction to a great extent. In contrast, conceptual advances with regard to the reduction of misconceptions and the adoption and construction of scientific explanations do need to be supported by instruction. Directing young children's attention toward basic physics explanations may lay the foundation for their later understanding at a more formal level. In addition, participation in learning environments of challenging science content will likely support the development of students' scientific reasoning skills such as their experimental skills, their use of empirical data to support theorizing, and their ability to reason about patterns of covariation (e.g., Bullock & Ziegler, 1999; Toth, Klahr, & Chen, 2000).

In our curriculum, children were directed to discover that if objects are immersed into water, water pushes against the object. This understanding may support the understanding that also forces always come in pairs—equal and opposite forces of action and reaction. It is well documented that even after having participated in physics instruction in secondary school students rarely understand the concept of force as a characteristic of a system, but rather treat it as a characteristic of single objects (McCloskey, 1983). It needs to be left to future longitudinal studies to show that an early understanding of science concepts in elementary school will help students to profit better from learning environments provided in secondary education.

References

Bransford, J., Brown, A., & Cocking, R. (Eds.). (1999). *How people learn.* Washington, DC: National Academy Press.

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

Caravita, S. (2001). A re-framed conceptual change theory? *Learning & Instruction, 11,* 421–429.

Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology, 21,* 13–19.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14,* 161–199.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1–49.

Christianson, R. G., & Fisher, K. M. (1999). Comparison of student learning about diffusion and osmosis in constructivist and traditional classrooms. *International Journal of Science Education, 21,* 687–698.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23,* 13–20.

Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *Journal of the Learning Sciences, 12,* 91–142.

Davis, E. A., & Linn, M. C. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education, 22,* 819–837.

Davis, E., & Miyake, N. (2004). Explorations of scaffolding in complex classroom systems. *The Journal of the Learning Sciences, 13,* 265–272.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age: The Jean Piaget symposium series* (pp. 49–70). Hillsdale, NJ: Erlbaum.

Greeno, J., & the Middle School Mathematics Through Applications Project Group. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53,* 5–26.

Halford, G. S., Brown, C. A., & Thompson, R. M. (1986). Children's concepts of volume and flotation. *Developmental Psychology, 22,* 218–222.

Hogan, K., & Pressley, M. (1997). Scaffolding scientific competencies within classroom communities of inquiry. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues* (pp. 74–107). Louiseville, Quebec, Canada: Brookline Books.

Kohn, A. S. (1993). Preschoolers' reasoning about density: Will it float? *Child Development, 64,* 1637–1650.

Lehrer, R., Schauble, L., Strom, D., & Pliggie, M. (2001). Similarity of form and substance. Modeling material kind. In M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 39–74). Mahwah, NJ: Erlbaum.

Linn, M. C. (1995). Designing computer learning environments for engineering and computer science: The scaffolded knowledge integration framework. *Journal of Science Education and Technology, 4,* 103–126.

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59,* 14–19.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299–324). Hillsdale, NJ: Erlbaum.

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* [Thinking in multiple perspectives in science education] (pp. 125–191). Bad Heilbrunn, Germany: Klinkhardt.

Pea, R. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences, 13,* 423–451.

Posner, G., Strike, K., Hewson, P., & Gertzog, W. (1982). Accomodation of scientific conception: Toward a theory of conceptual change. *Science Education, 66,* 211–227.

Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences, 13,* 273–304.

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., Grosslight, L., & Davis, H. (1997). Teaching for understanding. A study of students' preinstruction theories of matter and a comparison of the effectiveness of two approaches to teaching about matter and density. *Cognition and Instruction, 15,* 317–393.

Smith, C., Snir, J., & Grosslight, L. (1992). Using conceptual models to facilitate conceptual change: The case of weight–density differentiation. *Cognition and Instruction, 9,* 221–283.

Smith, J., diSessa, A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3,* 115–163.

Staub, F., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, 94,* 344–355.

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

Tynjälä, P. (1999). Towards expert knowledge? A comparison between a contructivist and a traditional learning environment in university. *International Journal of Educational Research, 31,* 357–442.

Tyson, L. M., Venville, G. J., Harrison, A. G., & Treagust, D. F. (1997). A multidimensional framework for interpreting conceptual change events in the classroom. *Science Education, 81,* 387–404.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4,* 45–69.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24,* 535–585.

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

Vygotsky, L. (1978). *Mind and society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wandersee, J., Mintzes, J., & Novak, J. (1994). Research on alternative conceptions in science. In D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 177–210). New York: Macmillan.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 17,* 89–100.

(*Appendix follows*)

Appendix

Coding Scheme for Teacher (T) and Student (S) Statements During Class Discussions

Source	Type	Classification	Examples
Teacher	Structuring	Comments on observations, descriptions of experiments, or requests to repeat results	What happened exactly? What did you feel when you immersed something into water? Can you please demonstrate this?
		Requests explanations and justifications	How come that the water rises? Why does the water push against your hand?
		Points to transfer or application	Have you observed this also in a different experiment? Where else did you notice this?
		Fosters reasoning processes of generalization	Is this correct: The water pushes all objects upwards? You said that the water raises with a stone, with a piece of clay, with a pot, with a ball; does the water always rise? How else could you express this: the larger an object, the more . . .?
		Addresses misconceptions by empirical confrontation or by calling attention to conflicting results	You said that all light things will float (*immerses a metal pin into water*). What about a ship that rests in the harbor?
		Underlines important comments	What Kathrin just said was really important. Yes, the water does push against it.
		Focuses attention on specific aspects	Look really closely at what happens when I immerse this pot. What happens to the rubber band? Look closely.
		Elaborates on concepts	You said that the water is pushed up or that the water is pushed away. We also say that the water is displaced.
		Helps with other means (drawing, gesture, demonstration)	Show it with your hands. This is important; we will talk about this in a minute; now first finish your thought. Why don't you show it with arrows on the board?
		Reinforces student comments	Well observed. Good idea.
	Passive	Poses an open question	Did anyone find out something new? Who would like to report what you have observed?
		Turns over question to another student/students	Max said it may be because of the air. What do the others think about this?
		Checks comprehension	You said it's because of its form; what do you mean by this?
		Repeats	You mean that it is because of the size, not the weight (*of the object*).
		Requests repetition	Could you say this again? Could you show this again so that everyone can see?
		Encourages to rethink	Who has got another idea? All of your ideas are important.
		Responds to student question	S. Are they all the same size? T: Yes, they are all the same size.
	Managerial	Discipline	Please be quiet. Now listen and stop the noise.
		Task	You are supposed to weigh the cubes and sort them according to their weight.
		Organizational	We'd best take this container and mark it. Please put your folders under your seats.
Student	Active	Offers own ideas	How come that this piece of wood is lighter than the other one? How can this be that metal will sink but a ship floats? It's made of iron, too. I discovered something; this container displaces the same amount of water that fits into it.
		Answers open teacher question	T: Who would like to share their ideas? S: I made a ship of clay and then I tried out whether it will sink in water with dish soap.
	Reaction to student	Responds to student comment with own ideas	S1: I think it's because of the salty water. S2: No, that can't be because ships will also float in rivers.
		Reinforces student comments	What Teresa said about the space I think is really good.
	Reaction to teacher		T: What did you find out? S: I found out that the water rises.
	Other	Organizational	I need this container and a pen.
		Observational	He is loading his boat. This doesn't fit into the container.
		Other	Be quiet.

Received February 1, 2005
Revision received January 31, 2006
Accepted February 2, 2006 ■