

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

Ilonca Hardy

University of Frankfurt, Germany

Birgit Kloetzer

University of Munich, Germany

Kornelia Möller

University of Muenster, Germany

Beate Sodian

University of Munich, Germany

Paper to be presented at the Annual Meeting of the American Educational Research
Association, New York, March 2008

Evidence-based Reasoning in Science Discourse

Students' ability to participate in scientific discourse and to use empirical evidence appropriately as a source for the backing of arguments has been repeatedly stated as a core goal of science education (American Association for the Advancement of Science, 1993; Bybee, 2002; National Research Council, 1996). For example, the National Research Council's *National Science Education Standards* (1996) set forth five "essential features" of scientific inquiry, which include the expectation that students formulate explanations from collected evidence and evaluate these explanations with regard to plausible alternatives. As such, students' participation in scientific inquiry involves a fundamental understanding of the „nature of science“ as an iterative process of collecting evidence based on theories and hypotheses (e.g. Carey & Smith, 1993). Scientific inquiry as an inductive process then is based on transformations from data to evidence, from evidence to patterns, and from patterns explanations (Duschl, 2003). Importantly, scientific inquiry does not merely involve the collection of empirical evidence but rather is dependent upon evidence for the construction, confirmation, or refutation of theories and models (Driver, Newton, & Osborne, 2000).

Analyzing evidence-based classroom discourse

Toulmin's (1958) framework for the analysis of arguments has served as a basis to many instruments for the analysis of classroom discourse in science (e.g., Jimenez-Alexandre, Rodriguez, & Duschl, 2000; McNeill, Lizotte, Krajcik, & Marx, 2006; Simon, Erduran, & Osborne, 2006). Essentially, an argument consists of a claim and supporting evidence, or backing, allowing a fundamental distinction between supported and supported claims in classroom discourse (see Furtak, Hardy, Beinbrech, Shavelson, & Shemwell, 2008). Whereas this distinction is especially useful for the characterization of classroom discourse patterns as well as for the analysis of co-constructed arguments with claims and counter-claims, a further differentiation may be made with regard to the quality of backing that students provide for their claims. Here, a framework originally developed by Driver, Leach, Millar, and Scott,

(1996) and refined for use in elementary school by Tytler and Peterson (2005) is especially important, distinguishing three levels of reasoning about science phenomena. On a first level, explanations about phenomena are equivalent to observations of the phenomenon itself. On a second level, explanations about phenomena are constructed based on (the observation of) correlations or causal relationships, thus relating patterns of observation. On a third level, explanations involve reference to models or theories, constructed on the basis of evidence. In our own framework (see Furtak et al., 2008), we combined the approaches of argument structure based on Toulmin's distinctions with the three levels of reasoning developed by Tytler and Peterson (2005) to arrive at a classification of instructional discourse which can be rated in terms of different levels of argumentative structure, moving from unsupported claims to claims supported with single data, relational evidence, or models and rules.

Research on argumentation in science discourse

In research on argumentation patterns in science classrooms it has been frequently found that episodes of high-level reasoning are rare; unfortunately, discourse is more often characterized by student claims that are not supported by evidence (z.B. Jimenez-Alexandre et al. 2000; Osborne, Erduran, & Simon, 2004). Newton and Newton (2000) state that this is especially frequent with younger children and elementary school teachers, while chances to find episodes of high-quality reasoning may be greater in secondary school. Nevertheless, children's ability to apply more sophisticated reasoning seems to develop during the course of elementary school; in fact, the quality of scientific reasoning has been found to increase in interview studies using similar scenarios repeatedly in a longitudinal design (Tytler & Peterson, 2005). In this study, simultaneously, students' conceptual understanding of the investigated phenomena increased. It may thus be speculated that a more coherent understanding of the physical mechanisms underlying the investigated phenomena, such as evaporation, also supported students' use of more sophisticated explanations, such as the comparison of patterns. Similarly, Kawasaki, Herrenkohl and Yeary (2004) were able to raise

elementary school students' levels of reasoning during the course of a curriculum on floating and sinking. The quality and quantity of participation in classroom discussion increased particularly when students were provided with roles addressing the procedural and cognitive steps involved in science investigations, moving them from phenomenon-based reasoning to relation-based reasoning.

Science Curricula in Elementary School

Curricula Promoting Scientific Reasoning

Despite research findings that have shown classroom discourse as largely dominated by unsupported student assertions (Jimenez-Alexandre et al., 2000), experimental research with young children has revealed that even five-year-olds are capable of differentiating between theory and evidence on a basic level (Ruffman, Perner, Olson, & Doherty, 1993; Wilkening & Sodian, 2005). For example, preschool children were able to interpret patterns of covariation in data as well as to generate answers based on simple cause-effect relationships in domains for which they did not hold strong prior beliefs (Koerber, Sodian, Thoermer, & Nett, 2005). Similarly, elementary school children were able to differentiate between one test with conclusive and another with inconclusive results and to explain their choices if they were prompted to formulate a prior hypothesis (Sodian, Zaitchik & Carey, 1991). These findings stand in contrast to earlier results with more demanding tasks which usually documented children's inability to apply scientific reasoning procedures when investigating or explaining science phenomena (Inhelder & Piaget, 1958; Kuhn, Amsel & O'Laughlin, 1988).

A fundamental understanding of the difference between and the coordination of theory and evidence may be seen as a prerequisite for the construction of scientific arguments, since explanations for scientific phenomena need to be constructed in reference to underlying theories and their supporting evidence rather than on an ad-hoc basis. Thus, for students' ability to use appropriate arguments in science discourse, their realization that the

construction, testing, and revision of theories and hypotheses about science phenomena constitute a fundamental part of the scientific endeavor seems to be a pivotal prerequisite. However, it has been found repeatedly that students in early secondary school exhibit an understanding of the nature of science labeled as „knowledge unproblematic,“ where it is assumed that knowledge is constructed by simple mechanisms such as observation (Carey & Smith, 1993). This position is characterized by an insufficient differentiation between theories and hypotheses on the one hand and evidence on the other hand. Furthermore, the iterative and cumulative nature of scientific knowledge is not sufficiently realized by students (Driver et al., 1996; Lederman, 1992; McComas, 1998).

Despite its relevance for scientific reasoning in science lessons, research into children's understanding of the nature of science has largely concentrated on domains and topics not treated explicitly in science curricula (Sodian, Jonen, Thoermer, & Kircher, 2006). Therefore, classroom studies connecting children's understanding of the nature of science and their scientific argumentation are largely lacking from the literature. An exception is a study by Carey et al. (1989) in which a seventh-grade curriculum on nature of science constructs, involving the explicit testing of hypotheses and reflection on evidence and theory-building, was designed based on the role of yeast formation for baking bread. In a comparison of pre- and post interviews, students on average showed an increase of half a level on Carey's Nature of Science Interview (Carey et al., 1989). Similarly, Sodian, Thoermer, Kircher, Grygier, and Günther (2002) achieved an increase in students' average level of understanding of the nature of science, with students achieving mostly on Level 2 (science as search for explanations) in the post-instructional interviews rather than on Level 1 (science as collecting facts). These findings point to the importance of students' explicit reflection on theory-construction and the use of evidence within engaging science investigations for a change in their implicit theories about the nature of science.

Curricula Promoting Conceptual Change

According to results by Tytler and Peterson (2005), there may be a relationship between students' level of reasoning and their conceptual understanding of the respective science phenomena. This seems plausible as students who display only a limited understanding of the concepts and mechanisms involved in producing a certain phenomenon may be constrained in their consideration of respective patterns of evidence to support their initial claims. Inquiry-based science curricula, which involve students in the investigation of complex science phenomena, may then also promote students' level of scientific reasoning as students' evolving conceptual understanding may serve as a vehicle for the more sophisticated construction of scientific explanations.

The literature on conceptual change has long been concerned with individual processes of knowledge construction with regard to science phenomena, especially investigating the role of students' naive conceptions on their successful construction of adequate scientific explanations. Due to informal experiences in their everyday life, children construct initial concepts that they apply for the interpretation of phenomena in the world. In many cases, these naive conceptions are not compatible with scientific models and thus need to be revised, or fundamentally restructured, into new, scientifically valid ideas (diSessa, 2006; Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001). Rather than viewing conceptual development as an abrupt change from naive to scientific conceptions, this process of restructuring is seen as gradual, encompassing phases with intermediate models, or phases where naive concepts are held in parallel with more sophisticated ideas.

Studies investigating the effects of traditional science instruction have largely shown that students' naive conceptions are quite resistant to change, frequently lasting beyond or re-emerging after the instructional units (Wandersee, Mintzes & Novak, 1994). Linn (2006) suggests that instruction should expose students to criteria by which they can differentiate between multiple competing ideas. In this way, naive conceptions may be integrated with

more sophisticated ones, or, in case of incommensurability, be discarded as inappropriate (see also Tytler, 1998). In the end of an instructional unit, students thus may ideally reveal integrated, scientific knowledge or else, fragmented knowledge still in need of fundamental restructuring. Our research group found a pattern of conceptual development corresponding to these assumptions with regard to elementary school students' understanding of floating and sinking after a two-week instructional unit (Hardy, Jonen, Möller, & Stern, 2006).

The Teacher's Role within Science Learning Environments

The importance of students' active construction of new scientific explanations and insights within learning environments is not generally disputed, as it follows the basic tenets of a constructivist orientation to knowledge (e.g., Bransford, Brown, & Cocking, 1999). However, the way that this active student involvement may be realized in instruction is still debated (e.g., Kirschner, Sweller, & Clark, 2006). For example, Mayer (2004) pointed out that within constructivist learning environments one needs to differentiate fundamentally between the dimensions of cognitive and behavioral engagement of students. While hands-on activities may be associated with cognitive activity, not every behaviorally observable activity such as the conduction of experiments will necessarily involve relevant student cognition. Importantly then, students' cognitive processes need to be supported by adequate teacher actions, or scaffolding.

The construct of scaffolding, originally framed by Wood, Bruner, and Ross (1976) and Vygotsky (1978), has been re-examined in its definition within complex science learning environments (Davis & Miyake, 2004; Furtak, 2006; Hogan & Pressley, 1997; Puntambekar & Hübscher, 2005). In a science context, scaffolding involves teacher actions to focus students' attention on the relevant aspects of a task, as well as actions that help students put ideas into a larger scientific context, thus modeling more advanced thinking and understanding (Pea, 2004; Reiser, 2004). With regard to fostering a culture of scientific argumentation, the teacher's support of the construction of evidence-based arguments

becomes especially relevant. In scientific discussions, teachers thus need to provide opportunities for students to use scientifically valid argumentation patterns, by asking students to provide explanations for their claims. Teachers prompting or modeling the use of evidence to support theories and claims may then be able to establish the use of evidence as a necessary element of scientific discourse. Similarly, Duschl and Gitomer (1997) describe assessment conversations that engage students in discussion of topics involving the use of evidence. Here, teacher prompts in terms of scaffolding instructional discourse include criteria such as focusing on relationships, consistency, or use of examples as a means to promote discourse concerned with reasoning processes rather than facts.

The topic of floating and sinking

The topic of floating and sinking, involving a fundamental understanding of the concepts of density, water displacement, and buoyancy, lends itself to the investigation of sophisticated student reasoning and conceptual development. As a challenging instructional topic, it is frequently taught only in secondary school. However, also most elementary school curricula and learning progressions involve reference to investigations of the behavior of different objects in water, attempting explanations of floating and sinking on a level of „material kind.“ Conceptually, explanations of objects’ floating and sinking may be regarded as quite challenging, as the integration of several concepts is called for in Archimedes’ principle. Rather than attributing the causes of floating and sinking to the characteristics of single objects, the relationship between object and surrounding fluid needs to be considered. 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.

Typically, elementary school students will focus on one dimension only when explaining the floating and sinking of objects: 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;“ see Smith, Carey, & Wiser, 1985; Möller, 1999). Moreover, in many explanations, air is seen as an active force that pulls objects upwards and water is seen as a force that sucks them downwards (Tytler & Peterson, 2004). Although particularly young children tend to adopt an undifferentiated concept of weight and density (Smith et al., 1985), even some secondary school students approach instruction with this type of commonsense theory (Smith, Maclin, Grosslight, & Davis, 1997). Most of the curricula developed for early secondary school focus on conceptual change with regard to a differentiation of students’ conceptualization of matter, frequently based on the use of visual representations (e.g., Lehrer, Schauble, Strom, & Pliggie, 2001; Smith, Snir, & Grosslight, 1992; Smith et al., 1997). 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.

Research Questions

In the following study, we addressed three main research questions to explore the effects of science curricula in elementary school on reasoning patterns in classroom discourse:

- (1) Can elementary students’ level of reasoning be promoted by extensive science curricula?
- (2) What is the relationship between students’ level of reasoning and their conceptual understanding?

- (3) Which teacher moves are associated with high-level reasoning in discourse? Do these associations change during the course of the curriculum, i.e., from first to last lesson?

In order to investigate these questions, we analyzed transcribed teacher-student discourse of two different data sets, both involving an elementary school curriculum on floating and sinking (four lessons each). Although both data sets are based on an additional experimental variation (one with regard to effects of instruction on nature of science, and one with regard to effects of scaffolding), they are comparable on the level of age group, instructed topic, and curriculum / learning progression so that our research questions may be productively addressed with both data sets. Thus, in addition to the overall research questions, the experimental variations used in data sets 1 and 2 enabled us to address the three research questions within the context of a curriculum on the nature of science (data set 1) and within the context of differential instructional support in the learning environment (data set 2).

We hypothesized that a curriculum on the nature of science preceding an instructional unit on floating and sinking would have positive effects on the level of reasoning evidenced in classroom discussions since students would be attuned to the role of evidence for supporting conjectures and theorizing. In a similar vein, we expected that a curriculum in which the teacher supports students' conceptual development by scaffolding discourse would also show positive effects on the scientific reasoning evidenced in these classrooms.

Method

Data Sources

Data Set 1

Data set 1 consists of each four transcribed 90-minute lessons of an elementary school curriculum on floating and sinking derived from the instructional approach described in more detail in data set 2 (see also Hardy et al., 2006). This instructional unit on floating and sinking was implemented in two fourth-grade classrooms after an experimental variation had been

realized based on instruction on the „nature of science“ in one of the classrooms. The two classrooms were comparable with regard to social demographic characteristics and the mean intelligence of the students (CFT 20). The experimental variation was based on 14 weekly 90-minute lessons on the nature of science, starting with a collection of students' initial conceptions about scientific inquiry and gradually attempting to refine their conceptions about experimentation and the relation between theory and evidence with regard to the topics of perception, light and shadow, yeast formation, and others. In the control group, an equivalent number of lessons treated the same topics conceptually, yet without offering explicit opportunities to reflect on the nature of scientific inquiry (see Sodian et al., 2006, for an extensive description of the curriculum). Both instructional units as well as the unit on floating and sinking were realized by the same teacher in both classrooms.

Data Set 2

Data set 2 is based on each four transcribed 90-minute lessons of a total of four third-grade classrooms (two experimental classrooms and two control classrooms). For the purposes of the analyses presented here, lessons 1, 3, 6, and 8 of the eight-lesson instructional unit were transcribed. Due to a recording failure of one the cameras used during lesson 8 in one of the experimental classrooms, transcripts of this lesson were not yet available for the purposes of this analysis. The experimental variation applied in this study involved the employment of different degrees of instructional support within the learning environments on floating and sinking, concerning 1) the sequencing of content and 2) the use of cognitively structuring statements by the teacher (see Hardy et al., 2006, for an extensive description of the curriculum). 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 on material kind, density, water displacement, and buoyancy force in the group of high support (experimental group). That is, a complex learning environment allowing for experimentation with regard to different aspects of floating and

sinking was used in both groups; however, it was pre-structured 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 high support. In addition to the sequencing of content, the instructional discourse was scaffolded by the teacher to a greater degree in the experimental group. This scaffolding was intended to provide conditions for conceptual change, as the teacher contrasted student hypotheses and conceptions, addressed misconceptions, or introduced a hypothesis or observation which the students themselves had not considered. In contrast, in the control group, the instructional discourse was student-centered, with students themselves reacting to each other's statements, while the teacher's role was focused on organizational supervision, with a lower frequency of content-related prompts to students' reasoning processes. Between the two experimental groups, the instructional time, the teacher, and the material for experimentation were not varied.

As the experimental variation involved a variation in the sequencing of content, the content of the four transcribed lessons (lesson 1, 3, 6, and 8) of the four classrooms varied in the curriculum of high and of low support. While the instructional content of lessons 1 (introductory lesson) and 8 (final lesson on integration of all concepts related to floating and sinking) were largely similar across all four classrooms, the content of lessons 3 and 6 varied in the two curricula. That is, in the curriculum of high support, the instructional content of lesson 3 concerned the topic of relative density, and the content of lesson 6 concerned the topic of buoyancy, while the content of those lessons in the curriculum of low support varied according to the particular concepts introduced by the students themselves.

In contrast to the curriculum employed in data set 2, the curriculum employed in data set 1 was based on a reduced, and more focused variant of the experimental curriculum of data set 2, involving a learning progression from material kind and density to the comparison of densities, with the teacher attempting to cognitively activate, challenge, and scaffold

students' conceptual understanding much as in the group of high support of data set 2. While the instructional unit implemented in data set 2 involved a total of eight 90-minute lessons on the topic of floating and sinking, this curriculum was reduced to four 90-minute lessons in data set 1. It needs to be mentioned that the two classes of data set 1 were taught with exactly the same curriculum on floating and sinking, and thus covered the same topics in the same sequence in both classes, while the classes involved in data set 2 experienced different versions of this curriculum, yet having available to them the same experimental material and resources.

Coding Scheme

For the analysis of instructional discourse with regard to students' evidence-based reasoning, their evolving levels of conceptual understanding, and the role of the teacher in supporting student reasoning, we employed the coding scheme described by Furtak et al. (2008). We developed this coding scheme on the basis of transcripts of science lessons from elementary and secondary school from diverse data sets. The application of the coding scheme allows us to segment the instructional discourse into reasoning units, each of which is assigned a level of reasoning (based on a classification of statements into elements of reasoning as premises, claims, data, evidence, or rule, respectively), a corresponding level of conceptual understanding, and different codes for teacher contributions (providing or prompting for different elements of reasoning), see Furtak et al., 2008. For the coding of the conceptual level within reasoning units, we employed a coding scheme based on a distinction between misconceptions (level 1), prescientific conceptions (level 2), and scientific conceptions (level 3; see Hardy et al., 2006 and section on results of prior analyses below).

For the purposes of the present analyses, in each of the transcribed lessons the post-experimental phases were identified, i.e. that phase during a lesson which occurred after the students had been given the opportunity to work on small experimental sites independently. These phases in all of the lessons of data sets 1 and 2 involved whole-class discussions in

which the teacher would ask students to report and reflect on their previously performed activities and insights. Post-experimental phases are of special importance for the analysis of scientific reasoning as they may provide rich opportunities for the use of empirical evidence in support of assertions.

After an initial agreement on the application of the codes, the two data sets were coded on different research sites. On each site, the story line for each transcript was written collaboratively by two coders for all of the transcripts. Reliability analyses were then performed for the coding of both data sets, with 25% of the total transcripts coded by two coders. For data sets 1 and 2, the percent agreement of two independent coders on 1) the coding of premises and claims, 2) the coding of backings, 3) the coding of backings in to data, evidence, or rule, 4) the conceptual level, and 5) the teacher contribution ranged between 80% and 95%. For each data set, the codes of the primary coder were used for all analyses.

Results of prior analyses on student conceptual understanding

As both of the curricula employed within data set 1 and within data set 2 have been analyzed previously with regard to effects on student conceptual understanding, these results will be briefly summarized here. In data set 1, effects of the curriculum on the nature of science were found with regard to a significant increase from pre- to posttest in the experimental group's understanding of the role of theory and evidence in the scientific process (as assessed by the Nature of Science Interview; Carey et al., 1989) and their application of experimental control strategies (as assessed by an experimental manipulation task), see Sodian et al., 2006. In a follow-up test one year after the instruction, the experimental group still showed a superior understanding of experimental strategies, whereas there were no significant differences with regard to the level of understanding achieved in the interviews. With regard to students' conceptual understanding of floating and sinking achieved after the additional participation in the curriculum on floating and sinking analyzed here, there were no significant differences between the experimental and the control group

with regard to means in conceptual understanding as assessed by a pre- and posttest on floating and sinking.

In both data sets, we employed a test on floating and sinking for the assessment of students' conceptual understanding which is described in detail in Hardy et al. (2006). It focuses on conceptual understanding on three levels of understanding—misconceptions (such as weight, size, shape, or active air), prescientific conceptions (such as hollowness, qualitative relation between water and object, material kind), and scientific conceptions (density, buoyancy, and their relation to water displacement). Employing this test as a pretest, posttest and one-year follow-up test for assessing the effects of the curricular variation in data set 2, we found a superiority of the experimental group with a greater degree of instructional support. That is, students in the experimental group rejected significantly more misconceptions on the posttest than the control group, and showed superior integrated conceptual understanding in the long run (see Hardy et al., 2006).

Results

Development of Levels of Reasoning

In a first step, the distribution of the four levels of reasoning for all of the lessons analyzed is presented as percentages with respect to all coded reasoning units. As may be seen in Figures 1 and 2, a high percentage of reasoning units in all of the groups was coded at level 1 (claim and premise). In data set 1 (Figure 1), an increase in the level of reasoning from first to last lesson is observed, with increasing percentages for level 2 (premise, claim and backing by data), level 3 (premise, claim, and backing by evidence), and level 4 (premise, claim, and backing by rule) in later lessons. For example, in the fourth lesson, 10% of the reasoning units in the experimental group (nature of science) were coded on level 4, whereas this level was only observed in 1.5% of the reasoning units in the first lesson. In data set 2, the percentages of levels of reasoning seem to vary more strongly according the particular lesson analyzed, whereas a particular trend for more sophisticated support of claims with data-

evidence-, or rule-based reasoning is not apparent at first sight. Thus, the group of low support seems to use data-based backing already in the first lesson to a higher degree than does the group of high support, while by the last lesson, the percentage of backed statements has actually decreased (from 42% of statements with backing in the first lesson to 22% of statements with backings in the last lesson). In both the groups of high and low support more than 70% of all reasoning units involve unsupported statements even in lesson 8.

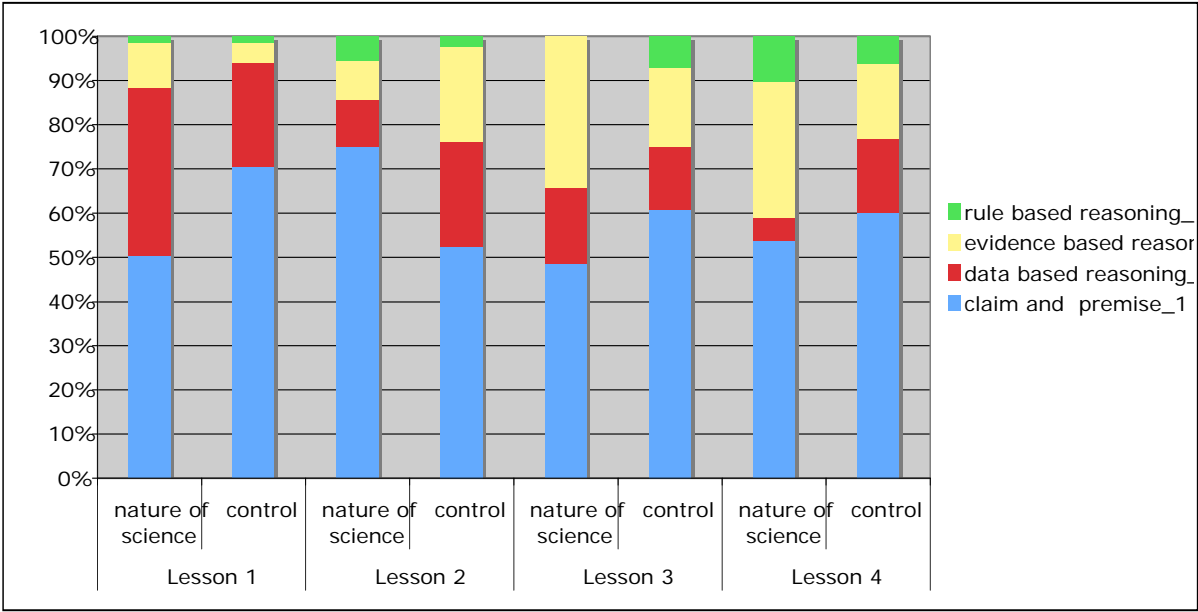


Figure 1. Levels of Quality of Reasoning by Group and Lesson for Data Set 1.

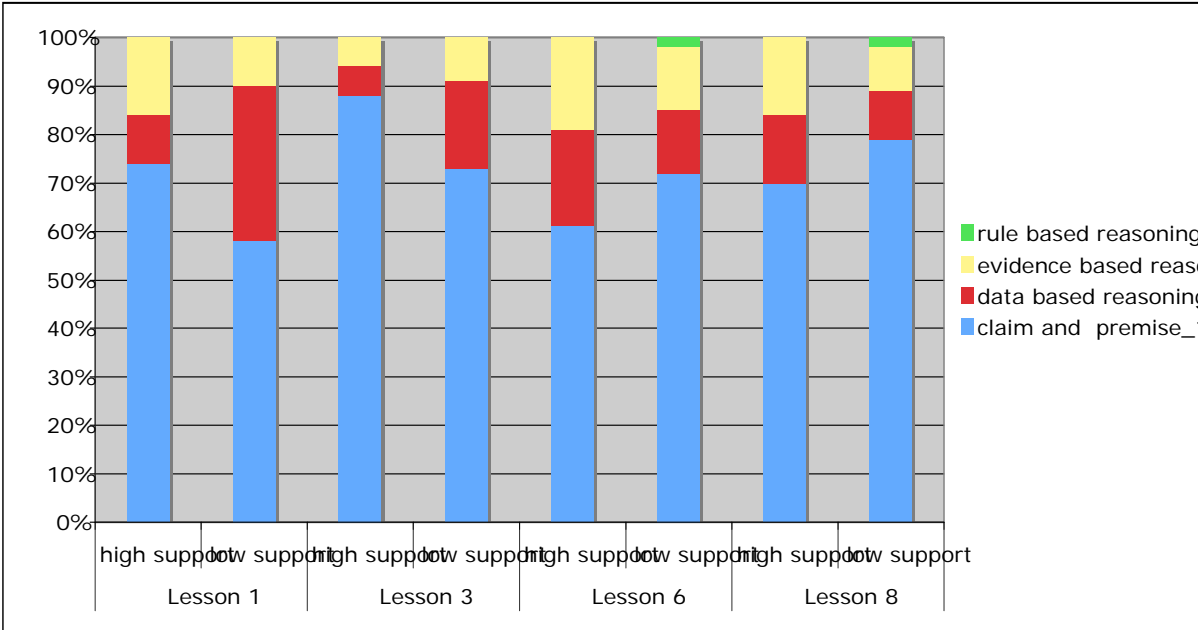


Figure 2. Levels of Quality of Reasoning by Group and Lesson for Data Set 2.

In both data sets, the development of the average level of reasoning across the four lessons was analyzed with ANOVAs using the factors of group and lesson. In data set 1, there was a significant main effect of lesson ($F(3, 393) = 3.51, p < .05, \eta^2 = .026$) while neither the effects for group or group x lesson were significant. Using follow-up tests to assess differences in the levels of reasoning between the lessons for each group separately, a significant increase was found for the group with a nature of science curriculum comparing lesson 1 to lessons 2, 3, and 4, respectively ($p < .05$) whereas none of the differences were significant for the control group. Comparing initial levels of reasoning for the two groups, however, an ANOVA revealed that the experimental group started with a greater mean level of reasoning in lesson 1, while differences between the groups were not significant in a comparison of lesson 4.

In data set 2, a respective ANOVA showed a significant effect for lesson ($F(3, 1011) = 5.04, p < .01, \eta^2 = .015$) and lesson x group ($F(3, 1011) = 2.90, p < .05, \eta^2 = .009$). Post-hoc comparisons for each group separately showed that the group of high support changed significantly in their average level of reasoning across the four lessons ($F(3, 594) = 8.73, p < .05, \eta^2 = .042$). Comparisons of levels of reasoning in lesson 2 with lessons 6 and 8 were significant ($p < .05$), with a significantly lower level of reasoning in lesson 3 than in lessons 6 or 8, while a comparison of lesson 1 and 8 only showed a tendency of increasing levels. This shows that the initial level of reasoning in these classes was largely comparable to the last lesson, while there was some variation in lessons 3 and 6. These variations were most likely due to the particular topics covered in these lessons. Post-hoc comparisons for the group of low support showed nonsignificant results with regard to differences in reasoning levels between the four lessons ($F(3, 417) = .82, n.s.$).

Development of Levels of Conceptual Understanding

Descriptive results for the development of conceptual levels across the four lessons show that an increasing number of reasoning units was coded at levels 2 and 3 in the later lessons than in earlier ones in both data sets, see Figures 3 and 4. That is, while level 3 (scientific concepts) is virtually absent from the instructional discourse of the first lessons, in the last lessons, this code is apparent to a much higher percentage across all of the groups analyzed. As the topic of floating and sinking is quite challenging for students conceptually, this descriptive result does not come as a surprise. Rather, students' increasingly sophisticated conceptual understanding confirms the learning progression intended by the curricula on floating and sinking. Additionally, it is evident that a large number of reasoning units was not rated at all on a conceptual level. These units largely involved single unsupported claims without conceptual content related to the explanation of floating and sinking, such as student observations of certain facts during experimental activities. As may be seen especially in Figure 4, the percentage of uncoded units in this study, similar to the levels of reasoning, seems to be dependent on the particular topic of the lesson.

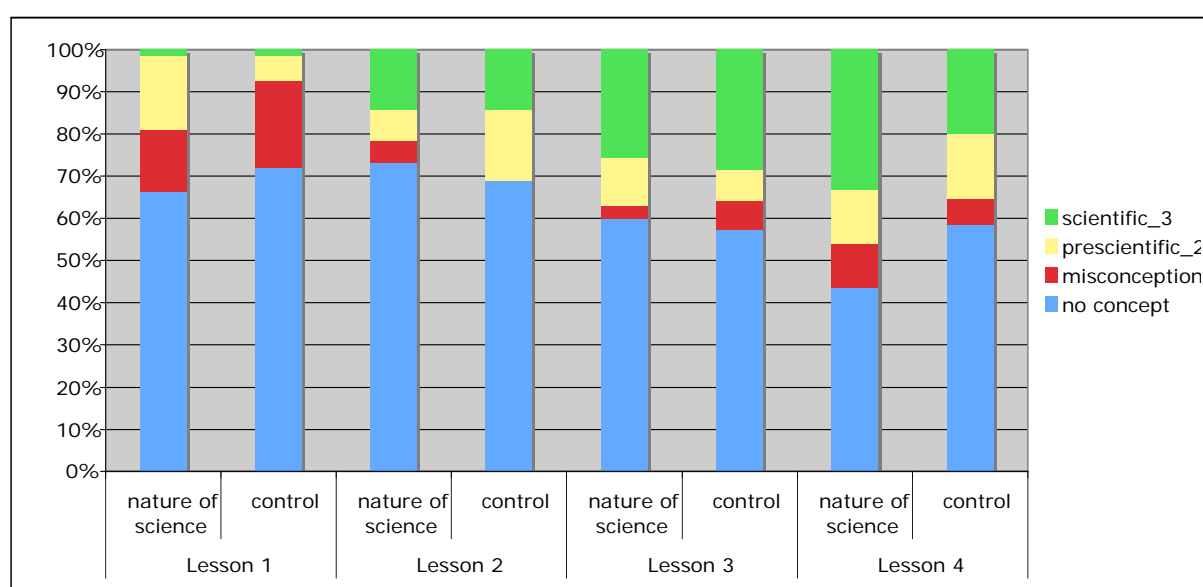


Figure 3. Levels of Conceptual Understanding by Group and Lesson for Data Set 1.

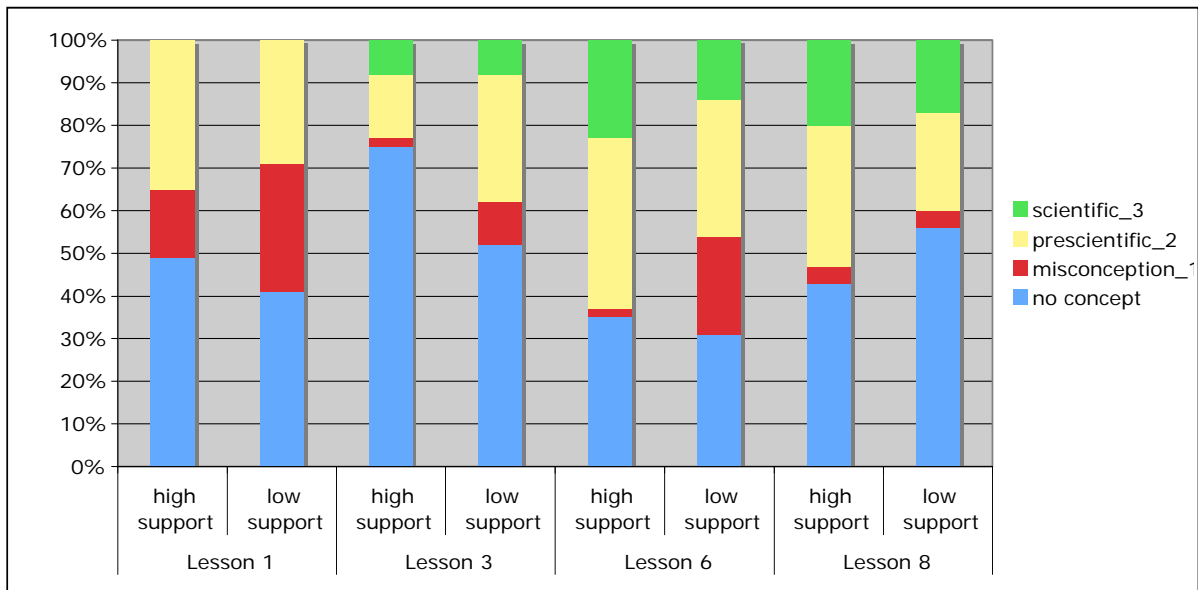


Figure 4. Levels of Conceptual Understanding by Group and Lesson for Data Set 1.

Analyzing the development of conceptual understanding across the four lessons using ANOVAs, in data set 1, there is a significant increase in the average conceptual level from first to last lesson ($F(3, 137) = 19.13, p < .001, \eta^2 = .26$), while the main effect for group and the effect for group x lesson are not significant. Follow-up analyses revealed that both the experimental and the control group significantly improved in their conceptual understanding comparing lesson 1 with lessons 2, 3, and 4, respectively ($p < .05$).

In data set 2, a respective ANOVA showed a significant effect for group ($F(1, 476) = 16.79, p < .001, \eta^2 = .034$), lesson ($F(3, 476) = 21.49, p < .001, \eta^2 = .12$), and lesson x group ($F(3, 476) = 4.14, p < .01, \eta^2 = .025$). Follow-up analyses revealed that the group of high support improved significantly across the curriculum ($F(3, 254) = 20.25, p < .001, \eta^2 = .19$), with significant effects for the comparisons of lesson 1 with lessons 3, 6, and 8 ($p < .05$). Similarly, the group of low support improved across the four lessons analyzed ($F(3, 222) = 10.26, p < .001, \eta^2 = .12$), with significant effects for comparisons of lesson 1 with lessons 3 and 8 ($p < .05$). As a higher percentage of reasoning units were rated as non-applicable in the group of low support, an additional ANOVA considered the units that were not rated

conceptually at a lower level, using a code of 0 instead of excluding these units from the analyses. If this is done, there is a significant effect of lesson ($F(3, 1011) = 15.93, p < .001, \eta^2 = .045$) and lesson \times group ($F(3, 1011) = 5.37, p = .001, \eta^2 = .016$). Post-hoc analyses showed that the average conceptual level of the group of high support is significantly higher than the level evidenced with group of low support in lesson 8 ($p = .05$), with M (high support) = 1.29 ($SD = 1.21$) and M (low support) = 1.02 ($SD = 1.22$).

The Relationship between Levels of Reasoning and Conceptual Understanding

The relationship between levels of reasoning and conceptual understanding is reported separately for the experimental and control groups of each data set as the experimental variations may be assumed to produce differences in the strength of association between the two variables. In data set 1, the average correlation between level of reasoning and conceptual understanding is $r = .45$ ($p < .05$) for the experimental group and $r = .40$ ($p < .05$) for the control group. Looking at correlations separately for each lesson, however, differences between the groups emerge. Here, the experimental group shows an initially low correlation between the codes with $r = .14$ (n.s.) in lesson 1 and an increased association by lesson 4 ($r = .42, p = .05$), whereas the control group reveals the opposite pattern, with $r = .59$ ($p < .05$) in lesson 1 and $r = .14$ (n.s.) in lesson 4.

In data set 2, the average correlation between conceptual understanding and level of reasoning is much lower, with $r = .13$ (n.s.) for the group of high support and $r = .24$ ($p < .05$) for the group of low support. Looking at the correlations separately for each lesson, however, a similar trend as in the experimental group of data set 1 is evidenced. Here, the group of high support starts with a correlation of $r = -.034$ (n.s.) in lesson 1 reaching a correlation of medium strength in lesson 8 ($r = .307, p < .05$), whereas the group of low support starts with a correlation of $r = -.26$ (n.s.), to end with a correlation of $r = .318$ ($p < .05$) in lesson 8.

The relationship between teacher contributions and levels of reasoning

With regard to the coded teacher contributions, we opted to concentrate on those teacher moves that were assumed to be most closely related to students' levels of reasoning. Thus, descriptively, percentages of the codes of teacher prompts for premise, claim, and backing as well as teacher provides premise, claim, and backing, are reported. Analyses performed on differences between these teacher moves with regard to the reasoning level of the associated reasoning units concentrate on the codes of teacher prompts for claim or backing and teacher provides claim or backing. As may be seen from Figures 5 and 6, the percentages for the respective codes vary extensively according to the type of group (high or low support) in data set 2, whereas for data set 1, the percentages between the experimental and control group are similar. Since data set 2 involved an experimental variation based on teachers' use of cognitive support, or verbal scaffolding of student conceptual understanding, this variation was expected and may actually be regarded a type of implementation check of the intended experimental variation. Yet, the question of whether the association between student conceptual understanding, reasoning level, and teacher prompts will vary during the course of the curriculum is still of interest.

At a first look, it is also evident that the teachers provide premises and claims at a much higher rate than they provide backings (across both data sets, an average of 6.25% of reasoning units was provided by teachers with backings, whereas an average of 25% of units was provided with claims). With regard to the prompts employed by the teachers, prompts for claims occur at a higher rate than prompts for backings (with an average percentage of prompts for backings of 15.4, and an average of 26% for prompts for claim), while prompts for premises are rarely observed. Note that a teacher may both prompt for and provide several elements of reasoning within one unit.

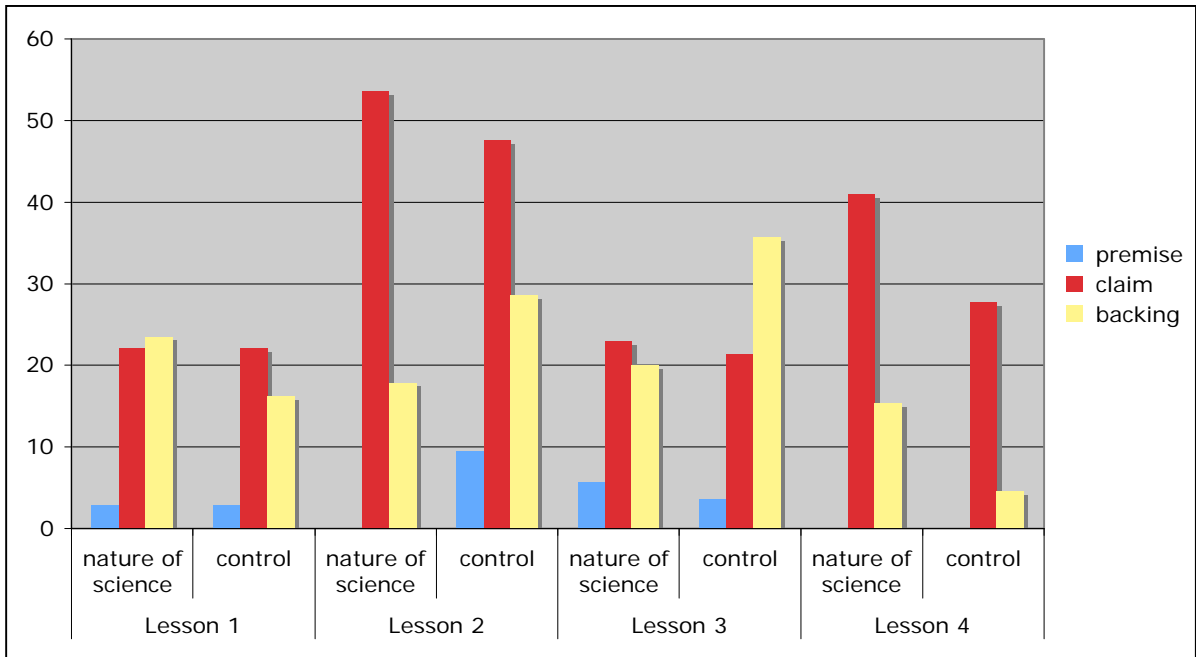


Figure 5. Percentages of Teacher Prompts by Group and Lesson for Data Set 1.

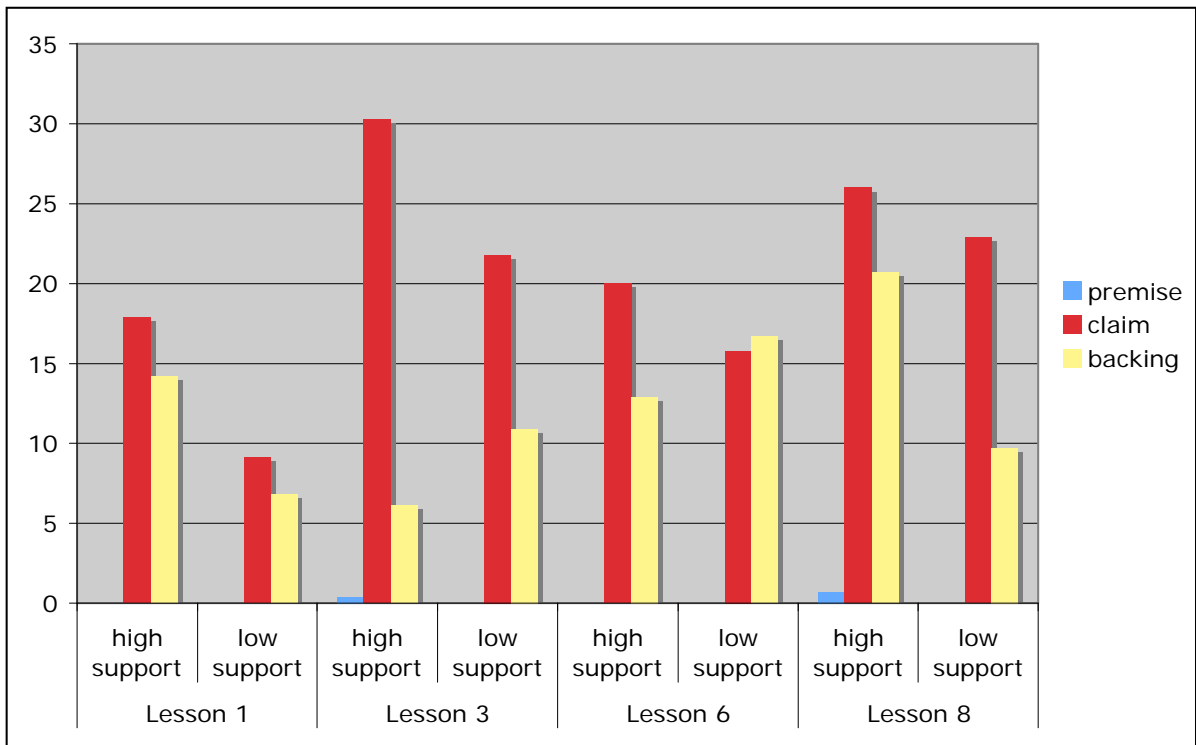


Figure 6. Percentages of Teacher Prompts by Group and Lesson for Data Set 2.

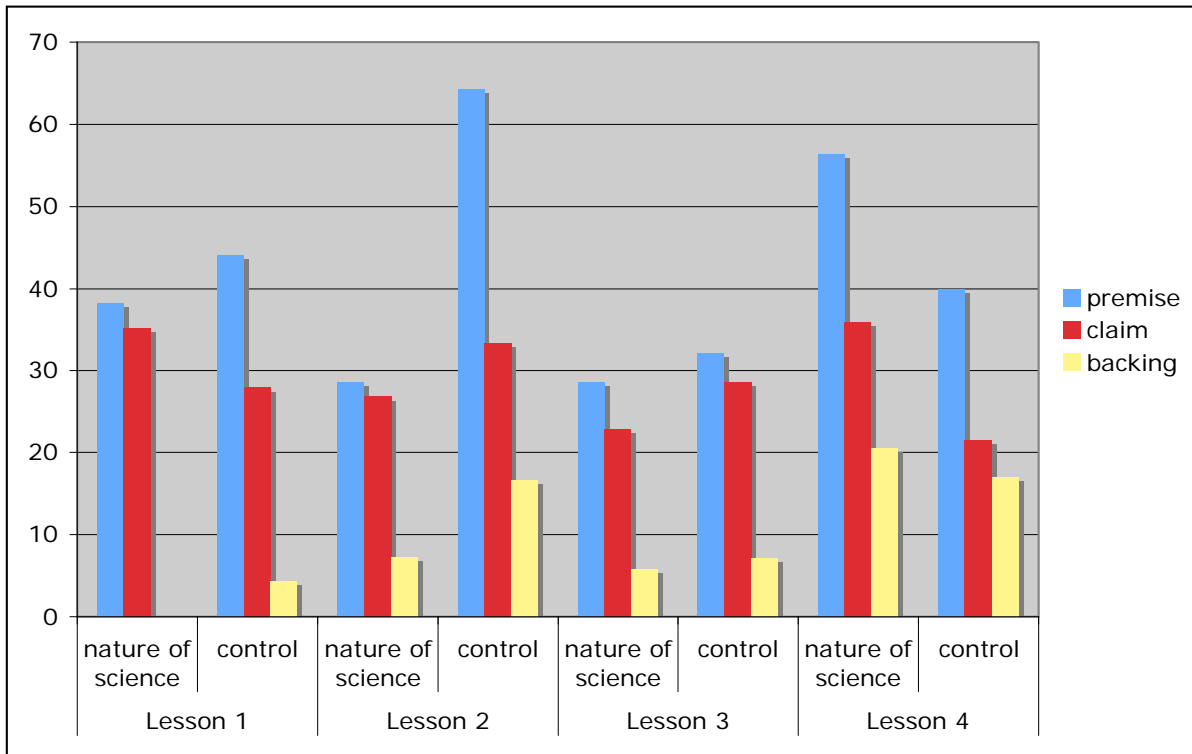


Figure 7. Percentages of Teacher Provides by Group and Lesson for Data Set 1.

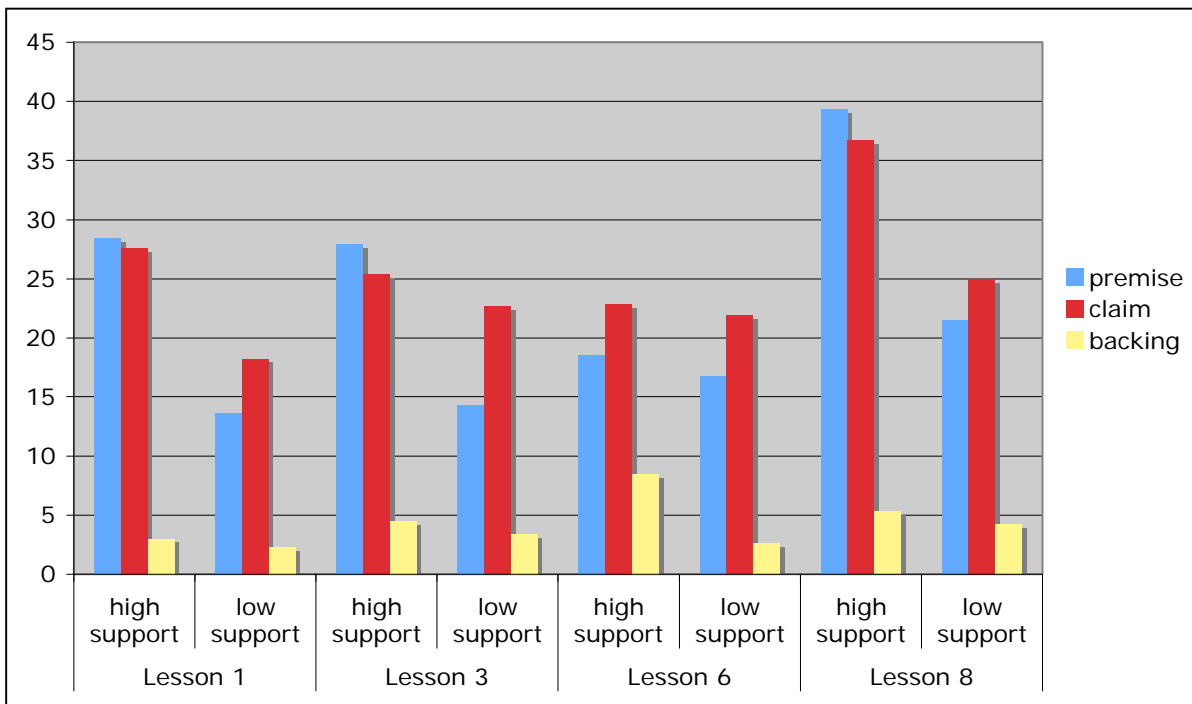


Figure 8. Percentages of Teacher Provides by Group and Lesson for Data Set 2.

Code of teacher prompts for backing.

The effects of teacher contributions on levels of reasoning were analyzed using the type of teacher contribution (yes / no) as independent variable in addition to the variables of group and lesson used in previous analyses. In data set 1, we found a significant effect for the code of teacher prompts for backing ($F(1, 385) = 86.80, p < .001, \eta^2 = .184$), lesson ($F(3, 385) = 6.35, p < .001, \eta^2 = .047$) and lesson x prompt for backing ($F(3, 385) = 4.92, p < .01, \eta^2 = .037$). Means of level of reasoning were significantly higher if teachers prompted for backing than if they did not do so. Follow-up analyses revealed that for the experimental group, there was a significant increase in mean level of reasoning between lessons 1 and 2, and lesson 1 and 4 ($p < .05$) if teachers prompted for backing, while for the control group results were not significant.

In data set 2, there was a significant effect for prompt for backing ($F(1, 1003) = 70.77, p < .001, \eta^2 = .066$) as well as a three-way interaction of prompt for backing x lesson x group ($F(3, 1003) = 4.23, p < .01, \eta^2 = .012$). Follow-up analyses showed that in the group of high support, there was a significant interaction effect of prompt for backing x lesson ($F(3, 590) = 4.51, p < .01, \eta^2 = .022$), revealing that across the four lessons, reasoning units with prompts for backing were associated with higher levels of reasoning albeit some variation across the four lessons. Similarly, in the group of low support, there was a significant interaction of prompt for backing x lesson ($F(3, 413) = 2.69, p < .05, \eta^2 = .019$), with consistently higher levels of reasoning when teachers provided prompts for backing.

Code of teacher prompts for claim.

With regard to the code of teacher prompts for claim, in data set 1 an ANOVA with prompt for claim, lesson, and group showed a significant effect for the presence of teacher prompts for claim, with significantly higher levels of reasoning if teachers did not prompt for claims ($F(1, 385) = 14.54, p < .001, \eta^2 = .036$) across groups and lessons. In data set 2, there were no significant main or interaction effects for a respective ANOVA.

Code of teacher provides backing.

An ANOVA performed with the factors of code of teacher provides backing (yes, no), lesson, and group showed a significant main effect for the presence of teacher-provided backing across groups and lessons, with significantly higher means if backings were provided in data set 1 ($F(1,302) = 23.09, p < .001, \eta^2 = .071$). Similarly, in data set 2, there was a significant main effect for teacher prompt for claim ($F(1, 1003) = 48.18, p < .001, \eta^2 = .046$), while none of the other effects approached significance, with higher mean levels of reasoning associated with reasoning units in which teacher provided backings.

Code of teacher provides claim.

Finally, analyses of effects with regard to the code of teacher provides claim in data set 1 revealed significant effects for lesson ($F(3,385) = 5.6, p = .001, \eta^2 = .042$), teacher provides claim ($F(1,385)=9.50, p < .01, \eta^2 = .024$), as well as significant two-way interaction group x teacher provides claim ($F(1,385) = 4.71, p < .05, \eta^2 = .012$), lesson x teacher provides claim ($F(3, 385) = 2.73, p < .05, \eta^2 = .021$), and group x teacher provides claim ($F(3, 385) = 3.63, p < .05, \eta^2 = .028$). Follow-up analyses showed that levels of reasoning were significantly higher if teachers provided claims, and that levels of reasoning increased from first to last lessons in both groups if teachers provided claims ($p < .05$). In data set 2, the effect of teacher provides claim only approaches significance ($F(1, 1003) = 3.29, p = .07, \eta^2 = .003$), with slightly lower overall means if teacher provided claims.

Discussion

Summarizing the results of the previous analyses, it was evident that the average reasoning level did not easily rise during the course of inquiry-science curricula in elementary school. As has been found in previous analyses of classroom discourse in secondary school, whole-class discussions were dominated by claims that are unsupported by (empirical) evidence. Also in our two data sets of elementary science curricula, an average percentage of 63 reasoning units contained only premise and claim (reasoning level 1) in first lessons;

similarly the average percentage of level 1 reasoning was 65 in last lessons. A significant increase in reasoning levels from first to last lesson was only achieved in one of the classrooms, in which students had participated in instruction on the nature of science preceding the science curriculum on floating and sinking. This class already started with a higher average reasoning level than did the control group of this data set, while the level of reasoning in the last lesson of the curriculum was not significantly different in the two classes. This result supports the interpretation that students' understanding of the difference between theory and evidence, as it was promoted by the preceding curriculum on the nature of science, also supported their more sophisticated construction of scientific arguments. Students' ability to use evidence and models in support of their assertions about science phenomena thus seems to be significantly promoted by their insight into the iterative nature of scientific investigations and the use of appropriate experimental methods.

In data set 2, we did not find a significant rise of average reasoning level across the four lessons analyzed; rather, reasoning levels seemed to vary according to the topic of the lesson (for example, whether class discussions revolved around the topic of density or buoyancy). However, there was a tendency of increase in the average reasoning levels from lesson 1 to lesson 8 in the experimental group with high instructional support, while no such development was evident in the group of low instructional support. Considering that one transcript of a last lesson in the experimental group was not yet included into the analyses, this tendency may still turn into statistically detectable superiority of the curriculum of high instructional support.

In contrast to the average reasoning levels, the level of conceptual understanding was promoted in all of the curricula considering differences between first and last lessons. While the experimental variations of the two data sets did not show a great impact on differences in students' conceptual understanding, the frequency of level 2 (prescientific) and level 3 (scientific) conceptions was greater during the later lessons of the curricula than in the

beginning where level 1 (misconceptions) were most dominant. If the large number of reasoning units that could not be rated at a conceptual level is taken into account in the analyses, we find a superiority of the experimental class of high support in data set 2. This superiority was expected considering that the students of this group showed significant higher levels of conceptual understanding in both posttests and follow-up tests. Nevertheless, there does not seem to be an obvious relationship between these achievement test data and the quality of conceptual understanding in instructional discourse even though the coding systems differentiating between the three basic levels of understanding used in both types of analyses were largely the same.

The medium to low overall correlations between the conceptual understanding and the level of reasoning in respective reasoning units show that the relation between the two constructs is not as straight forward as one may have hypothesized. Particularly, our coding schemes for conceptual understanding include level 1 concepts (misconceptions) that frequently only consider single variables or observations and more sophisticated level 2 and 3 concepts that are necessarily based on relations between constructs. Thus, one may have expected that students who refer to the relation between constructs on a conceptual level may also be inclined to use supporting evidence that is relation-based. While this relationship between concepts and evidence-based reasoning did appear toward the end of the instructional units, especially the first lessons showed that data- and evidence-based reasoning did occur with corresponding low-level, one-dimensional concepts. Apparently, as students gained insight into the concepts underlying floating and sinking during the course of the curriculum and perhaps also adapted to the particular nature of classroom discourse promoted by the teacher, the association between the two constructs became closer. However, it may only be speculated about this possible causal relation between both constructs.

It also needs to be pointed out that the correlations between conceptual understanding and level of reasoning not only varied across the analyzed lessons but also between the two

data sets where overall, there is a medium-sized correlation within data set 1 and a much lower correlation within data set 2. It may be speculated that this finding is also due to differences in length between the two curricula, so that a reduced and focused learning progression as it was used in data set 1 will tend to produce a closer association between the constructs than the treatment of a wider range of concepts within a larger time-frame as it was attempted in data set 2. Overall, however, the low to moderate correlations support the validity of our coding scheme with regard to the rating of reasoning levels as a construct that is separate and separable from conceptual understanding.

How did the teachers support the quality of reasoning in whole-class discussions? As expected, teacher prompts for backing especially raised the level of reasoning in associated reasoning units. This effect was evidenced in both data sets revealing that students apparently took up these prompts by providing types of backing (data, evidence, or rule) as expected. The effect of teacher prompts was apparent despite interactions with type of experimental group or lesson, with consistently greater means associated with reasoning units that included prompts for backing. A similar effect was found in both data sets for units in which the teacher provided backing themselves. Considering that the prompting for backing seems to work in the intended way, however, the scaffolding of teachers to prompt students is preferable with regard to students' more active involvement in the construction of evidence-based arguments. While the effects of prompting and providing backing seem to be rather straightforward, the effects of teachers prompting for and providing claims are less consistent, as they varied in both data sets and involved multiple interactions with experimental groups and lesson. For example, in data set 1, reasoning units in which teacher provided claims were associated with significantly lower reasoning levels, while there was no such effect in data set 2. Likely, the association of multiple teacher codes (for example, teacher provides claim plus teacher prompts for backing) needs to be analyzed with more scrutiny in order to follow up these divergent effects.

Conclusion

As has been evidenced in previous studies, students will not spontaneously use backings such as data, evidence, or rules to support their claims in classroom science discourse. Rather, teachers need to prompt students to do so. Even in our highly developed curricula, intended to promote conceptual change and involving multiple opportunities for scaffolded reasoning, a large percentage of reasoning units involved unsupported claims. Nevertheless, teacher prompts for backing apparently were picked up on by the students so that the average reasoning level increased in these reasoning units. This result supports the conclusion that teacher scaffolding with regard to the appropriate employment of (empirical) evidence in class discussions is pivotal for an involvement of students in the active construction of scientific arguments.

References

- American Association for the Advancement of Science (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.
- Bransford, J., Brown, A., & Cocking, R. (1999). *How people learn*. Washington, DC: National Academy Press.
- Bybee, R. W. (2002). Scientific Literacy: Mythos oder Realität? In W. Gräber (Ed.), *Scientific Literacy. Der Beitrag der Naturwissenschaften zur allgemeinen Bildung*. (pp. 21-43). Opladen: Leske und Budrich.
- Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). An experiment is when you try it and see if it works. A study of junior high school students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, 11, 514-529.
- Carey, S. & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist*, 28, 235-251.
- diSessa, A. (2006). A history of conceptual change research. In K. Sawyer (Hrsg.), *The Cambridge Handbook of the Learning Sciences* (S. 265-281). Cambridge: Cambridge University Press.
- Davis, E., & Miyake, N. (2004). Explorations of scaffolding in complex classroom systems. *The Journal of the Learning Sciences*, 13(3), 265-272.
- Duschl, R. A. (2003). Assessment of Inquiry. In J. M. Atkin & J. Coffey (Eds.), *Everyday Assessment in the Science Classroom* (pp. 41-59). Arlington, VA: NSTA Press.
- Duschl, R. A., & Gitomer, D. H. (1997). Strategies and Challenges to Changing the Focus of Assessment and Instruction in Science Classrooms. *Educational Assessment*, 4(1), 37-73.
- Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Bristol: Open University Press.

- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the Norms of Scientific Argumentation in Classrooms. *Science Education*, 84, 287-312.
- Furtak, E. (2006). *The dilemma of guidance in scientific inquiry teaching*. Unpublished doctoral dissertation. Stanford University.
- Furtak, E., Hardy, I., Beinbrech, T., Shavelson, R., & Shemwell, J. (2008). A framework for analyzing reasoning in science classroom discourse. Paper to be presented at the Annual Meeting of the American Educational Research Association, New York, NY.
- Halford, G. S., Brown, C. A., & Thompson, R. M. (1986). Children's concepts of volume and flotation. *Developmental Psychology*, 22(2), 218-222.
- Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking." *Journal of Educational Psychology*. 98(2), 307-326.
- 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: Brookline Books.
- Kawasaki, K., Herrenkohl, L. R., & Yeary, S. A. (2004). Theory Building and Modeling in a Sinking and Floating Unit: A case study of third and fourth grade students' developing epistemologies of science. *International Journal of Science Education*, 26(11), 1299-1324.
- Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Jimenez-Alexandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). "Doing the Lesson" or "Doing Science": Argument in High School Genetics. *Science Education*, 84, 757-792.

- Kirschner, P., Sweller, J. & Clark, R. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.
- Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children. Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, 64, 141-152.
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.
- Lederman, N.G. (1992). Students' and teachers' conceptions of the nature of science: A Review of the Research. *Journal of Research in Science Teaching*, 29, 331-359.
- 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: Lawrence Erlbaum Associates.
- Linn, M. (2006). The knowledge integration perspective on learning and instruction. In K. Sawyer (Hrsg.), *The Cambridge Handbook of the Learning Sciences* (S. 243-264). Cambridge: Cambridge University Press.
- Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59(1), 14-19.
- Mc Comas, W.F. (1998). *The Nature of Science in Science Education*. Dordrecht: Kluwer.
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting Students' Construction of Scientific Explanations by Fading Scaffolds in Instructional Materials. *The Journal of the Learning Sciences*, 15(2), 153-191.
- 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: Klinkhardt.
- National Research Council. (1996). *National Science Education Standards*. Washington, D.C.: National Academy Press.
- Newton, D. P. & Newton, L. D. (2000). Do Teachers support Causal Understanding through their Discourse when Teaching Primary Science? *British Educational Research Journal*, 26(5), 599-613.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the Quality of Argumentation in School Science. *Journal of Research in Science Teaching*, 41(10), 994-1020.
- 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.
- Puntambekar, S. & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(1), 1-12.
- Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, 13(3), 273-304.
- Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, 64, 1617-1636.
- Simon, S., Erduran, S., & Osborne, J. (2006). Learning to Teach Argumentation: Research and development in the science classroom. *International Journal of Science Education*, 28(2-3), 235-260.
- Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, 62, 753-766.

- Sodian, B., Thoermer, C., Kircher, E., Grygier, P., & Günther, J. (2002). Vermittlung von Wissenschaftsverständnis in der Grundschule. *Zeitschrift für Pädagogik*, 45. Beiheft, 192-206.
- Sodian, B., Jonen, A., Thoermer, C., & Kircher, E. (2006). Die Natur der Naturwissenschaften verstehen – Implementierung wissenschaftstheoretischen Unterrichts in der Grundschule. In M. Prenzel, & L. Allolio-Näcke (Hrsg.). *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms*. Münster: Waxmann.
- Smith, C., Carey, S., & Wisner, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, 21, 177-237.
- Smith, C., Snir, J., & Grosslight, L. (1992). Using conceptual models to facilitate conceptual change: The case of weight-density differentiation. *Cognition and Instruction*, 9(3), 221-283.
- 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.
- Toulmin, S. E. (1958/2003). *The Uses of Argument*. Cambridge: Cambridge University Press.
- Tytler, R. (1998). The nature of students' informal science conceptions. *International Journal of Science Education*, 20(8), 901-927.
- Tytler, R., & Peterson, S. (2004). From „try it and see“ to strategic exploration: Characterizing young children's scientific reasoning. *Journal of Research in Science Teaching*, 41(1), 94-118.

- Tytler, R., & Peterson, S. (2005). A Longitudinal Study of Children's Developing Knowledge and Reasoning in Science. *Research in Science Education, 35*, 63-98.
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction, 15*, 317-419.
- 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 Publishing Company.
- Wilkening, F., & Sodian, B. (Eds.). (2005). Special Issue on “Scientific reasoning in young children”. *Swiss Journal of Psychology, 64*(3), 137-217.
- 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.
- Vygotsky, L. (1978). *Mind and society: The development of higher psychological processes*. Cambridge: Harvard University Press.