This is the peer reviewed version of the following article: Doucerain, M. M., & Schwartz, M. S. (2010). Analyzing Learning About Conservation of Matter in Students While Adapting to the Needs of a School. *Mind, Brain, and Education*, *4*(3), 112–124, which has been published in final form at https://doi.org/10.1111/j.1751-228X.2010.01090.x. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Analyzing Learning About Conservation of Matter in Students While Adapting to the Needs of a School

Marina M. Doucerain[1] and Marc S. Schwartz[2]
[1] Concordia University
[2] University of Texas

We probed the impact of two teaching strategies, *guided inquiry* and *argumentation*, on students' conceptual understanding of the conservation of matter. Conservation of matter is a central concept in middle school science curriculum and a prerequisite upon which rests more complex constructs in chemistry. The results indicate that guided inquiry was particularly effective in improving students' conceptual understanding, as evidenced by pre/posttest results and by a skill analysis of in-depth interviews of student dyads. We also discuss how the challenges inherent to educational contexts can undermine the quality and limit the impact of empirical research carried out in many schools. We suggest how these challenges could be met in the emerging infrastructures for change called the Research Schools Network.

The purpose of this article is twofold.We examine the impact of two teaching approaches—*guided inquiry* and *argumentation*—on middle school students' conceptual understanding of the conservation of matter, in a study conducted in Eastern Canada. In parallel, we discuss how the challenges inherent to educational contexts can compromise the quality and limit the impact of school-based empirical research. In an era of evidence-based teaching, the hurdles to creating usable knowledge are as challenging for researchers as understanding complex ideas is for students. In this article we address both issues by presenting promising findings while highlighting the need for infrastructures that allow educators and researchers to feel confident in the strategies they advocate and use.

Conceptual Understanding

Science educators and education researchers look upon conceptual understanding—as opposed to algorithmic understanding—as a crucial learning outcome for students (Slavings, Cochran, & Bowen, 1997; Vosniadou, 2007). Conceptual understandings are mental constructs that support complex views of scientific ideas that students use to make sense of novel situations and solve scientific problems (Bowen & Bunce, 1997). However, numerous studies show that, even after years of formal science education, most students maintain a fragile understanding of science concepts, even though they make considerable progress in applying algorithms to solve problems(Abraham & Williamson, 1994; Pfundt & Duit, 2000; Sawrey, 1990). Fostering conceptual change remains a considerable challenge for science educators. With this focus in mind we probed the impact of a curriculum on conservation of matter that relies on two teaching strategies: guided inquiry and argumentation. These strategies, which can be seen as complementary, make very different assumptions about learning.

Inquiry Approaches

During the last few decades, research in science education has touted inquiry as a potent approach to promote conceptual understanding (Hameyer, van den Akker, Anderson,& Ekholm, 1995). In spite of its centrality in the realm of science education, though, the term inquiry still suffers from semantic confusion (Minstrell, 2000).Herewe explore several competing definitions.

In its popular form, inquiry is often equated with "handson" or "discovery" methods. Both share a strong emphasis on physical engagement as a means to investigate natural phenomena and on the centrality of students' choices and decisions (Holliday, 2004). However, although this approach supports initial growth in understanding, it quickly decays over time (Schwartz & Sadler, 2007).

In a more general fashion, inquiry, as framed by a constructivist paradigm (Tobin, Tippins,& Gallard, 1994), balances the complementary and interdependent roles of activities, discussions (Abell, Anderson, & Chezem, 2000; Tobin, 2006), and teacher support (Holliday, 2004). Inquiry is further characterized by attributes such as eliciting student prior knowledge and predictions when appropriate (Fischer,

Address correspondence to Marina Doucerain, Culture, Health, and Personality Lab, Department of Psychology, Concordia University, 7141 Sherbrooke W., Montreal, Quebec, Canada H4B 1R6; e-mail: marinadoucerain@gmail.com.

2000).

Guided inquiry (Magnusson, Palincsar, & Templin, 2004), the approach we used, features more structure and teacher support than the discovery form of inquiry (Martin, 2000). In our approach, the teacher (or curriculum) is responsible for identifying the goal of the activity, whereas students are responsible for choosing actions to achieve the goal. If viewed on a teaching methods continuum (Figure 1), guided inquiry would fall between the two extremes that are "expository" (i.e., traditional, textbook-based methods) and "discovery" teaching (Martin, 2000).

Argumentation Approaches

Although guided inquiry can support focused activity, the depth of understanding students achieve is not well understood. In the last decade, a growing body of science education researchers suggested shifting emphasis from "science as exploration and experiment" to "science as argumentation and explanation" (Abell et al., 2000). The concept of "argumentation" comprises a large range of related terms such as science talk (Lemke, 1990), conversation (Magnusson et al., 2004), or discussion (Vellom & Anderson, 1999). All are informed by the same teaching philosophy and are treated as synonyms.

Two main theoretical frameworks delineate argumentation in science education (Jiménez-Aleixandre & Erduran, 2008). One paradigm adopts a sociocultural perspective that views language and social interaction as playing a paramount role in shaping students' cognitive processes (Jiménez-Aleixandre & Erduran, 2008; Vygotsky, 1978). The second relies heavily on the importance of discourse in constructing knowledge (Magnusson et al., 2004).

Overall, studies examining argumentation in science teaching found that students' conceptual understanding increased when exposed to this approach(see von Aufschnaiter, Erduran, Osborne, & Simon, 2008 for a review). Several characteristics of argumentation, as employed in this study, can illuminate this positive outcome. First, discussions make cognitive processes visible through language by revealing students' "theories in action" (Karmiloff-Smith & Inhelder, 1975) and thus help uncover misconceptions (Bloom, 2001). Collaborative conversations also promote the coconstruction of knowledge by allowing students to confront their ideas with one another and negotiate new understandings (Barnes & Todd, 1995). Argumentation approaches create environments that rely on "consensus without coercion" (Vellom & Anderson, 1999, p. 181), which resemble scientific communities (Lawson, 2003).

Conservation of Matter

We focused on conservation of matter, which states that matter can neither be created nor destroyed. Therefore, in a closed system, the amount of matter present remains constant, regardless of the change processes at play in the system. Conservation of matter is not only central in middle school science curricula (National Research Council, 1996) but also a pivotal and prerequisite concept upon which rests more complex constructs in chemistry (Ozmen & Ayas, 2003). Consequently, it is particularly important for students to develop a sound comprehension of conservation for later study in chemistry.

However, as with other scientific concepts, most high school students show only a partial and fragile conceptual understanding of conservation of matter (Agung & Schwartz, 2007; Benjaoude & Barakat, 2000) and hold numerous misconceptions (Driver, Squires,Rushworth, & Wood-Robinson, 1994; Ozmen & Ayas, 2003). Formal instruction appears to help students solve algorithmic problems but not improve their understanding of the concept (Gomez, Pozo, & Sanz, 1995). Given its importance and challenges, conservation of matter is an ideal concept for studying the impact of guided inquiry and argumentation on science learning.

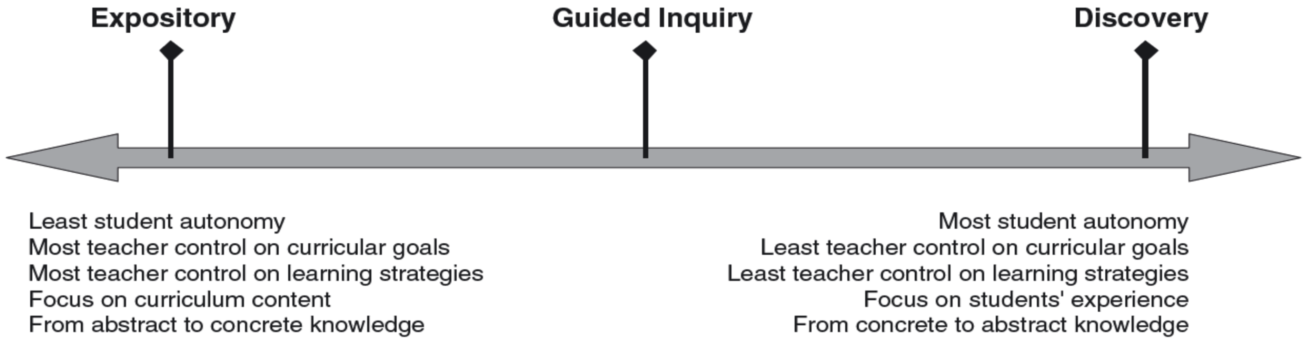
Conceptual Change, Dynamic Skill Theory, and Microdevelopment

In accordance with our focus on students' conceptual understanding, we adopted "conceptual change" as a general theoretical framework for learning (Posner, Strike, Hewson, & Gertzog, 1982). According to this view, learning involves a long, gradual, and complex process of change in cognitive structures that takes place in and is influenced by the larger sociocultural context of the learner (Vosniadou, 2008).

Within that framework, we used a developmental model, dynamic skill theory (DST; Fischer & Bidell, 2006), to characterize students' changes in conceptual understanding. DST is a neo-Piagetian model that presents learning as a nonlinear sequence of increasingly complex cognitive skill development, which can account for both contextual learning and generalization of knowledge and skills across task domains. DST describes behavioral changes in two very different time frames. The first extends from birth to approximately 30 years of age. During this interval behavioral changes can be divided into four tiers of development (reflexive, sensorimotor, representational, and abstract). The last three tiers constitute the ones of interest for educators. The second time frame is much shorter, where behavioral changes are measured in minutes, hours, or days along the same developmental continuum. The changes observed during this time frame, otherwise known as microdevelopment, refer to the application of DST to more immediate contexts,which influence how we use the skills that emerge as a result of maturation. Changes during this shorter time frame are congruent to learning observed in schools (Granott & Parziale, 2002; Schwartz, 2009;

Figure 1

Continuum of teaching approaches between expository and discovery



Note. Teaching approaches on this continuum vary along dimensions such as locus on control (of both curricular goals and strategies to achieve these goals) and direction of the progression between concrete and abstract knowledge. This figure, however, only serves illustrative purposes and is by no means an exhaustive representation of the wide variety of teaching approaches.

Schwartz & Fischer, 2004), and DST provides a metric to quantify changes in student understanding observed in this study.

Classroom-Based Research and Research Schools

Successfully bridging the methodologies of social science to the demands of educational contexts depends on finding ways to accommodate the political demands of educational systems while respecting the more uncompromising nature of science. The incongruous demands of both fields considerably hamper the rigorous research that science values in order to respect the needs of educators, administrators, and parents (Bryk & Gomez, 2008).

The methodology that emerges in this study is the result of numerous compromises, some that many education researchers have faced and some that reflect the unique dynamics of policy makers involved in deciding how education and research will unfold in their community. Ultimately compromises can leave in doubt whether the variables were well controlled or understood. This is a vexing situation especially if the study provides potentially important insights. However, the larger issue is whether educators and researchers can find the appropriate balance between needs and concerns that will allow for greater certainty in the methods and results as well as comfort for teachers and administrators in the process.

In a modest attempt to untangle the complicated nature of understanding in an equally complicated educational system, we accepted the conditions school administrators imposed upon us. However, we took note of the constraints, as an opportunity to better define what a community dedicated to evidence-based research would look like. Infrastructures for change have been explored in earlier articles in this journal (Coch, Michlovitz, Ansari, & Baird, 2009; Hinton & Fischer, 2009; Kuriloff, Reichert, Stoudt, & Ravitch, 2009); however, each infrastructure is often a solution that addresses local challenges. Creating an evidence-based culture is not a self-evident process, but nevertheless must be addressed so that all members can understand and appreciate not only its potential but also the demands it poses on all.

One strategy we explore later is a reconfiguration of John Dewey's Lab school. In this context the university and local school district are working with community leaders in defining and creating a Research Schools Network (RSN) committed to collaboration and evidence-based change. To that end, the RSN would become a practical infrastructure for making the goals, concerns, and constraints of all parties transparent and ultimately could build and support dynamic relationships between researchers, practitioners, administrators, and policy makers.

General Methodology—Context of the Study and Participants

The study took place in an Anglophone private school in Quebec during a 2-year period. Although our initial design included randomized samples and controls for the entire eighth grade, both authors needed to respond not only to the initial concerns of the administration, but also to the ongoing challenges faced by the teachers. To that end the methodology that emerged over the 2 years reflected the dynamic interaction between policy, the demands of evidence-based research, and teacher concerns.

The reader should also note that the data from this study are not only the results collected formally after the interventions but also the teachers' reactions to the data, which influenced methodological decisions. Their reactions, as well as the administration's policies, impacted how the study unfolded and the eventual conclusions drawn.

One of the initial conditions the administration set for work in their school was the elimination of controls. The administration felt that parents would not accept a situation where some students would enjoy the intervention while others would not. The second author offered to return and provide similar instruction for all students if the intervention was positive and significant; however, the administration felt this strategy was too complicated and would require too much time in a curriculum that already required teachers to cover a large number of concepts.

Additionally, there were unique political pressures in this Canadian province to teach some classes such as science in French. The political issues underlying which language of instruction could or could not be used in each class created an additional challenge for the teachers and researchers. Although all were bilingual, the language skills of students were mixed. Ultimately this variable did not seem to have an impact in the study's findings, but second-language instruction is an obvious concern in any study trying to document changes in student understanding and certainly a variable that needs to be controlled. However, given that the teachers involved in the study were interested in proceeding, we accepted the methodological compromises and concerns. At a minimum the project could serve as a pilot study.

Year 1 Methods

Five grade 8 classes (82 students total) taught by three teachers participated. As part of the middle school bilingual program, stronger students—as determined by a French placement test—learn science in French. They constitute four of the five grade 8 classes in this study. Teachers made the informal observation that most of the time students' level in French correlated to their general academic level.

During the first year, we first investigated the guided inquiry component (GIC) of the instructional unit. Because of practical considerations raised by teachers, the argumentation component (AC) was initially removed from the study, butwas reintroduced months later in response to teachers' reactions to the findings from this first investigation.

Instructional Unit

We used the DESIGNS[1] curriculum on conservation of matter, developed at the Harvard-Smithsonian Center for Astrophysics and adapted by the second author. The curriculum included two distinct components: guided inquiry and argumentation.

GIC

The guiding philosophy behind this part of the curriculum's design was "to make sure the goal of each activity was clear from the student's perspective so that they would recognize ways to take action" (Schwartz & Sadler, 2007, p. 996). This part of the curriculum focuses on establishing a balance between uncontrolled open-ended exploration and the need for ongoing teacher intervention. Additionally, the activities follow a microdevelopmental progression that allows students to "build and coordinate skills in a hierarchical manner from a lower tier to a higher tier" (Schwartz & Sadler, 2007, p. 996). An innovative element of the unit is a concrete physical model that students use and manipulate to develop a particulate view of matter. Table 1 presents the sequence and content of hands-on activities of the curriculum during 10 days for a total of six 60-min periods. Readings and class discussions that precede and debrief these activities are not shown.

AC

This aspect of the curriculum focused on structured discussions and debate around a phenomenon related to the conservation of matter that students could observe. Figure 2 shows an example of such a problem.

At the beginning of the class, the physical recreation of the problem was presented to students and then hidden. Students had to predict the outcome of the situation and reach a consensual explanation of the problem with peers who had made the same prediction. In a later step, students had to defend their position and try to convince others in a large class discussion. The teacher served as a mediator, requesting clarifications, probing questionable claims, and managing the overall affective and interpersonal dimensions of the conversation. The teacher refrained from providing the right answer, thus allowing students to build, defend, and debate possible answers. At the end of the class, the hidden physical setup was displayed and students had a chance to discuss the outcome with peers in order to make sense of the phenomenon.

The curriculum included three similar interventions each lasting 60 min. The first was a written assignment eliciting students' prior knowledge and introducing the scientific discourse. Table 2 explicates in more detail how the other two lessons unfolded. They were structurally identical but addressed two different examples of conservation.

The AC was initially removed from the curriculum. Teachers worried about the extra time required for both components. Despite their interest in the AC intervention, they pointed out that they had so much material to cover that they would appreciate any opportunity to condense the study. As

[1]DESIGNS is an acronym for Doable Engineering Science Investigations Geared for Non-Science Students'

Table 1
Description of Guided Inquiry Activities

Activity		Description
Discovering the vinegar/trona reaction	Goal:	Investigating the reaction between vinegar and trona
	Tier:	Sensorimotor
Properties of the vinegar/trona-baking soda reaction	Goal:	Investigating the properties of the vinegar/baking soda (trona) reaction
	Tier:	Sensorimotor (main focus) and representational
Beginning to quantify the vinegar/baking soda reaction	Goal:	Testing if a 1:1 vinegar to baking soda mass ratio can propel a ping pong ball out of a 90-cm tube
	Tier:	Representational (main focus) + sensorimotor
Finding the ideal ratio for the vinegar/baking soda reaction	Goal:	Finding the ideal vinegar:baking soda mass ratio to propel a ping pong ball out of a 90-cm tube
	Tier:	Representational (main focus) + abstract + sensorimotor
A physical model of the reaction	Goal:	Predicting outcomes of the vinegar/baking soda reaction by using the "intermediate model"
	Tier:	Abstract (main focus) + representational

the second author was fairly convinced that the GIC alone would lead to significant gains, and having already accepted the imposed limitations in the methodology set by the administration, he agreed to conduct the study without the AC.

Instruments to Measure Conceptual Understanding

Questionnaire on Conservation of Matter

We used a multiple-choice instrument designed at the Harvard-Smithsonian Center for Astrophysics (see Figure 2 for an example of the questions developed). The construction of this 25-item instrument was an outcome of numerous student interviews and answers to open-ended essay questions (Sadler & Schwartz, 2004) and was further validated with later work (Agung & Schwartz, 2007). All items include distractors, which identified misconceptions encountered during interviews or reported in the literature. For the purpose of this study, we only used the 15 grade-appropriate items that tested conceptual understanding. We translated the questionnaire from English to French and asked a few grade 9 students to validate the translation.

This instrument was administered to all students before and after the GIC to quantify changes in conceptual understanding. Students understood that these scores would not impact their grades and that they were participating in a study to better understand the effectiveness of the GIC.We report these data before continuing with how the methodology evolved as a result of the teachers' reaction to the findings.

Table 2
Description of the Argumentation Sessions

Step	Description
Introduction of problem	Physical setup presented to the class and then hidden. Students make prediction about outcome.
Small-group discussion	Groups formed based on students' predictions. Groups discuss to reach consensual explanation and elect a delegate.
Delegates' debate	Delegates persuasively discuss respective groups' position. Rest of the class is audience.
Whole-class discussion	All students can participate. Goal is to reach consensual explanation of problem.
Reevaluation of answer	Students have time to reconsider their position based on debate.
"Nature's answer"	Display of physical setup. Students write how they make sense of actual outcome.
Opening up	Teacher introduces related but different problem. Class discussion of how negotiated knowledge applies to new situation.

Year 1 Results

After completing the GIC, all three teachers and the second author met to compare and discuss the pre- and posttest scores. On average, scores had increased across the five classes by 8 percentage points, from 30% (SD = 13) to 38% (SD = 19). Though modest, this gain is statistically significant (p = . 0001). In terms of effect size, post inquiry test scores were on average half a standard deviation above pretest scores (Hedge's g = . 50). The 95% confidence interval for g based on the observed standardized mean change of .50 is .24–.75. That is, the observed effect size is just as consistent with a population effect size as low as .24 as it is with a population effect size as high as .75.

Given the teachers' interest, as well as the students', in the GIC, teachers could not understand the meager gain. Despite the time and energy invested in the GIC, the posttest average was still a failing mark. Although disappointed with the outcome, the meeting was a defining moment for the teachers and the methodology.

During this discussion the teachers began wondering

Figure 2
Example of a problem used in the argumentation component.

Look carefully at the picture of the balance. George balances a nail and beaker of water on each side of a balance. He then places the nail in the beaker on the right side. What will happen to the pan with the nail in the beaker of water?

a. The pan will move up.
b. The pan will not move; the pans are still balanced.
c. The pan will move down.
d. The pan will first move up and then down.
e. The pan will first move down and then up.

about the importance of the AC. They were curious about its role in the learning process. They also began talking about the process of building evidence-based research. The students' results and teachers' reactions created a compelling outcome they wanted to explore. The second author and teachers discussed the ad hoc nature of the evolving methodology and the emerging challenges in untangling the competing variables and that any new outcome (as measured by the questionnaire) might be difficult to interpret. Two of the three original teachers wanted to invest the additional 3 days necessary for the AC.

Four months later these two teachers returned to the topic of conservation during a week-long period of review set aside by the department and used the AC. The questionnaire was given a third time at the end of the week. Figure 3 illustrates how the guided inquiry and ACs were coordinated with the questionnaire.

Although students were seeing the questionnaire for a third time, the change in scores was striking. The scores had on average increased by 33 percentage points from the second attempt (postinquiry), from 38% (SD = 19) to 71% (SD = 19). This gain is not only statistically significant ($p <$. 0001) but also striking. In terms of effect size, the increase is about 1.3 SD (Hedge's g = 1. 34). The 95% confidence interval for g based on the observed standardized mean change of 1.34 is .95–1.73. Figure 4 summarizes all test results for the first year.

All three teachers and the second author met to discuss the third round of student results. The group easily concluded that the AC had played an important role, but it was not clear whether the dramatic improvement emerged as a result of the unfolding methodology, the teachers involved, an unknown interaction between the GIC and the AC, or other reasons. The teachers wished to explore these possibilities, and thus we created a new study design for the following year to accommodate the administration and teachers' concerns while better managing the variables.

Figure 3
Unfolding of year 1 interventions and test administration.

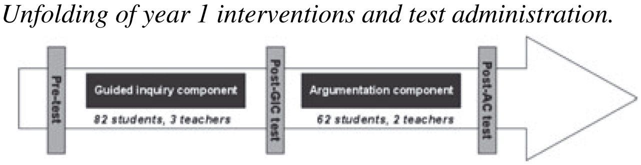
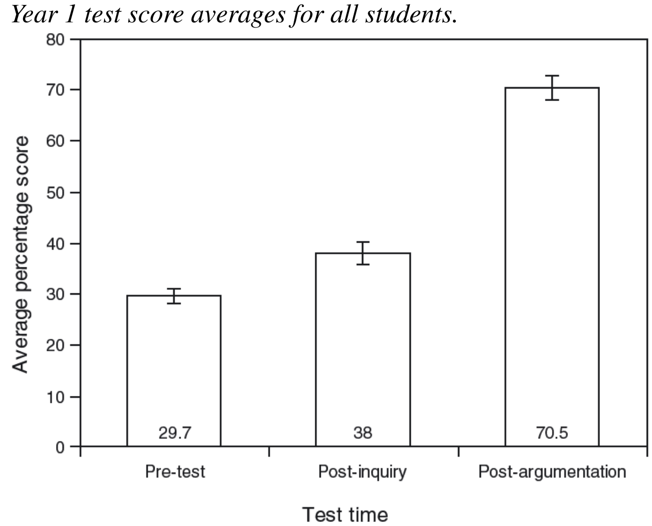


Figure 4
Year 1 test score averages for all students.

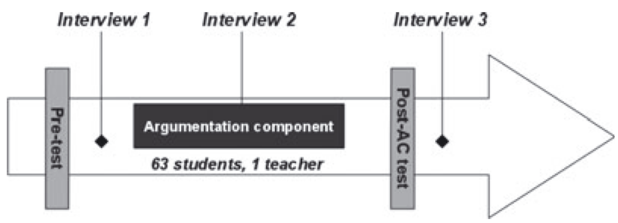


Year 2 Methods

By year 2, one teacher had changed schools and another was on leave. The school was facing a new reform movement in Quebec that introduced the possibility of a new battery of exams at the end of the year. Although not yet mandatory, the school felt compelled to participate. Given the new political context, and as new teachers decided not to participate in the latest version of the study, the remaining teacher (first author) discussed the possibility of investigating the strength of the AC with her three classes (N = 63). Consequently,

Figure 5

Unfolding of intervention and data collection during year 2.



during this stage of the study we looked at just the impact of the AC on student thinking. As the year before, the intervention covered three 60-min periods.We used the same questionnaire before and after the AC.

In addition, the first author decided to include an interview protocol to explore in further depth how student ideas were evolving during this component. This option appeared as the only remaining opportunity in our context to capitalize on what had been learned so far about changes in student understanding. Figure 5 summarizes the sequence of intervention and data collection steps in the second year of the study.

Interviews

The interviews were introduced in the second year of the study to unpack how—in addition to how much—student understanding evolved over time. In order to capture the students' thinking processes as they unfold through the social interactions embedded in the argumentation process, we conducted a series of interviews with student dyads. Half purposefully and half randomly, we selected three pairs of students who displayed different levels of general understanding in science: *low* (Larissa and Ludivine), *middle* (Matthieu and Mel), and *high* (Helen and Henry). We interviewed each pair three times. In each session students solved together a few problems illustrating the conservation of matter, while the interviewer's role was restricted to explaining the problems, asking for clarification, and mediating student interactions.

We performed a skill analysis of the interview transcripts by scoring each meaningful segment, here defined as a student's response, or a series of responses, presenting ideas on a single aspect of conservation of matter. We used the DST scale (Fischer & Bidell, 2006) to score the complexity of each unit of meaning. This scale encompasses eight hierarchical levels ranging from single sensorimotor actions (level 1) to abstract mappings (level 8). In other words, for each unit of meaning we assigned a score between 1 and 8 depending on the level of conceptual understanding of conservation of matter exhibited by the student's utterance. Table 3 shows a description of this scale.

For example, if a student said, "the flame is getting bigger," we coded the observation as a single representation and, therefore, assigned it a score of 4. If a student claimed that

Table 3

Microdevelopmental Scale Used for Skill Analysis

Score	*Complexity level*	*Description*
4	Single representations	Notes one observable aspect of the situation
5	Representational mappings	Coordinates two aspects of the situation
6	Representational systems	Coordinates two representational mappings
7	Single abstractions	Coordinates two representational systems into a single abstract concept
8	Abstract mappings	Coordinates two single abstractions

"Gas bubbles leave through the thing [hole in the can], andnow the whole can is less heavy," the explanation demonstrates a coordination between two sources of variation: the movement of gas bubbles and the weight of the can. The student's statement represents a mapping of representations and is assigned a score of 5. If a student says "Even if the Coke bubbles and becomes vapor, the bottle will still weigh the same. Because if the bottle is sealed, it's still the same amount of Coke," then he is reasoning at the system of representations level by demonstrating that he can coordinate four sources of variation: the connection between a change of state and the amount of matter, as well as coordinating the weight of the bottle with the fact that the bottle is sealed. This statement would be assigned a score of 6. Scoring of similar data had been performed in the past with the same content (Dawson-Tunik, 2004).

Year 2 Results

Test Scores

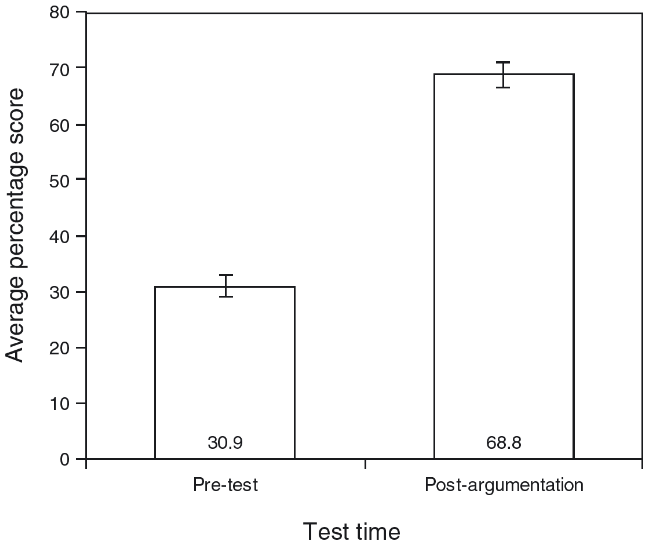
Students' scores increased on average by 38 percentage points, from 31% (SD = 15) on the pretest to 69% (SD = 18) on the postargumentation test. This gain, which is statistically significant ($p < .0001$), moved a majority of the class from failing (pretest score was lower than 60 for 93% of the students) to passing grades (posttest score was higher than 60 for 65% of the students) during the 3-day intervention. In terms of effect size, scores increased on average by about 1.5 SD (Hedge's $g = 1.51$). The 95%confidence interval for g based on the observed standardized mean change of 1.51 is 1.12–1.90. Figure 6 summarizes these results.

Comparison Between Year 1 and Year 2

Year 1 and year 2 students' performances were statistically similar, both on the pretests ($p = .46$) and on the post

Figure 6

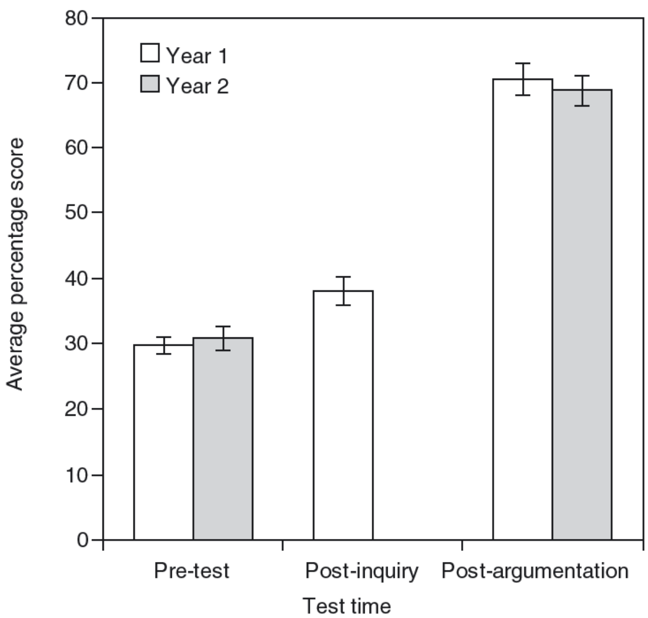
Year 2 test scores averages for all students.



Note. Error bars represent 1 SE above and below the mean.

Figure 7

Comparison of average test scores between year 1 and year 2.



Note. Error bars represent 1 SE above and below the mean.

argumentation tests (p = . 43), as illustrated in Figure 7. This rough measure is not a substitute for a control; however, the measures suggest that the students in year 1 and year 2 are comparable in terms of their performances on the questionnaire.

The differences in effect size between postinquiry and postargumentation gains suggest that guided inquiry was less effective than argumentation in helping students develop their conceptual understanding of conservation of matter. It also appears that the AC of the curriculum was equally effective, whether or not guided inquiry preceded argumentation sessions.

Microdevelopmental Skill Results

In this study, the complexity of students' utterances ranged from single representations (level 4) to abstract mappings. Assessing students' level of conceptual understanding in each one of their utterances allowed us to map out students' microdevelopmental pathways across their learning of conservation of matter. For each interviewee, we represented his/her microdevelopmental pathway by plotting the complexity level of utterances over time, as in Figure 8. In that way, each pathway shows how the complexity of a student's understanding of conservation of matter evolved as the three interviews unfolded.

Interviewees' score change between pre- and posttests is strongly correlated to their average change in skill level between interview 1 and interview 3 (r = 0. 55). Given the limited number of students included in this comparison, this correlation is just an indicator that both methods of assessing students' learning seem to capture different facets of the same phenomenon. Interviewees' microdevelopmental pathways provide a finer grained picture of the student's effort in achieving a deeper understanding of the conservation of matter. Although these pathways are very different from one another, they share several similar characteristics.

First, they all display a wave-like pattern.They fluctuate to a varying extent but none is linear, even in very short time frames. This variation in cognitive complexity is consistent with other studies that use a microdevelopment lens to characterize learning (Granott & Parziale, 2002; Yan & Fischer, 2002) and reinforces the idea that learning is not a linear process (Fischer & Bidell, 2006). Learning a new concept evolves through a sequence of cognitive regressions and progressions where students constantly reorganize their cognitive structures to meet the demands of the moment.

However, the pathways presented here demonstrate degrees of variation in cognitive complexity or frequency of oscillation between skill levels. Yan and Fischer (2002) suggest that this kind of variation is indicative of a process of large-scale skill reorganization that novices go through when they learn a complex new concept. The authors note that "during the early part of the construction process, novices show numerous, rapid fluctuations within a developmental range between upper and lower attractors, trying to build an appropriate skill but having difficulty sustaining it" (p. 152). In this study, all interviewees were novice with respect to conservation of matter and had to reorganize their cognitive skills. Based on Yan and Fischer's argument, a greater degree of variation along the microdevelopmental pathway would be indicative of more cognitive reorganization or learning. To examine this idea, we looked at the correlation between interviewees' average level of skill variation and their score increase between pre and posttests. For each interviewee, we assessed the average level of skill variation by counting the total number of wavelike fluctuations and dividing this value by the total number of utterances across all three interviews. This computation yields an admittedly crude estimate of average number of wave-like fluctuations per utterance. The correlation between this measure of variation and pre/posttest score increase is high ($r = .66$). This result suggests that indeed a higher level of fluctuation in skill complexity is related to more learning, as measured by the questionnaire.

Second, all the learning trajectories represented here fluctuate within a certain range. All pathways are mostly contained between the level of complexity required to fully answer most interview questions (level 7) and students' most basic descriptions of a phenomenon (level 4). Yan and Fischer (2002) highlight that "each task has a characteristic skill at a particular complexity level that is optimal for success" (p. 155). The upper boundary observed on individual microdevelopmental pathways typically matches the optimal level of complexity for the task at hand. They also note that for most tasks this ideal level of performance is below the highest level that individuals can reach. In this case, however, it seems that the ideal level of performance was not in easy reach for the students (otherwise, they would have constantly performed at that level). Instead, it represented a certain stretch that may have pulled students toward higher complexity. Ideally, educators would match the level of a learning task's optimal level of complexity not only to the specific developmental range accessible to the learners but also more specifically to the higher end of that developmental range.

Within this sextet of pathways, specific trajectories also exhibit patterns of similarity and difference. For Helen and Larissa, skill level averages of each interview continuously increased. Also, only their pathways contain clusters of consecutive high-level performances (labeled "high level clusters" on the charts). That is, they were both able to make more that one consecutive series of utterances at a newly reached level of complexity (in this case the abstract level).The presence of these clusters suggests thatHelen and Larissa were consolidating their complex understanding of conservation of matter, in accordance with Yan and Fischer's (2002) analysis of clusters of high performance. This phenomenon is also reflected in their questionnaire scores. On the posttest, Helen and Larissa scored 93 and 87%, respectively, the highest among the interviewees. Thus, both their test scores and microdevelopmental pathways suggest that they reached and to some degree consolidated a complex understanding of conservation of matter.

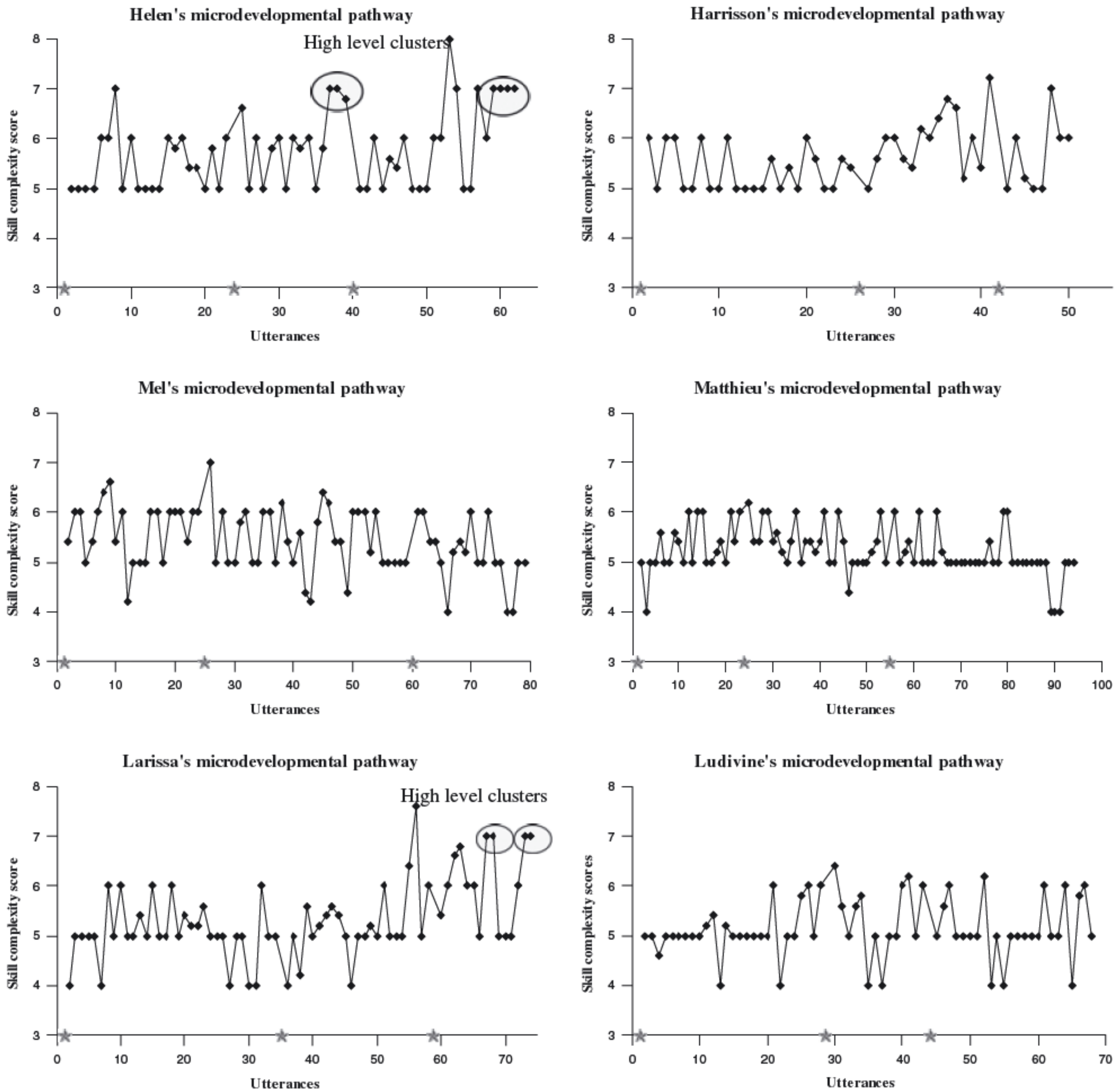
Conversely, in Henry's, Matthieu's, and Ludivine's cases, skill-level averages evince a different pattern. The averages increase from the first to second interview, but abate during interview 3.Apossible explanation for this motif is the absence of in-class support at the time of interview 3. The third interview takes place after the completion of argumentation sessions and the posttest, and thus occurs during a period of no contextual support. On the other hand, interview 2 took place during the intervention, and students' performances are likely potentiated by contextual support. Because argumentation sessions offer high contextual support, students perform at a higher cognitive level than they would in the absence of support. That students' performances evidence an overall drop in complexity when contextual support is removed is consistent with Rappolt-Schlichtmann, Tenenbaum, Koepke, and Fischer's (2007) study results.

In the case of Mel, interview skill level averages continuously decreased. In light of his 46-point gain on the posttest, this downward course is surprising. However, a closer look at interview transcripts helps shed some light on this puzzling contrast. During the second interview, Mel seemed to struggle between what he believed and what he had learned in class: He often used cautionary "according to the rules," or "in theory" to introduce what he perceived to be "the official version," and then transitioned to a misconception with a "but . . . . '' As the gap between personal and "official" (consensually constructed in class) understandings widened, Mel's responses gradually became shorter and less complex. He became increasingly reluctant (or unable?) to explain his ideas and sometimes closed his statements with "I can't explain it." Has Micro Developmental Pathway certainly reflects this struggle. However, although Mel was willing to be honest about his beliefs on conservation of matter during the interviews, he seemed to defer to (most probably chose to adopt) the "official version" during the posttest in order to obtain a good score. As such, it is very likely that Mel's score on the posttest reflects a "strategic" understanding of conservation of matter instead of his real ideas on the subject.

Discussion

This article emerged at the confluence of a compelling finding uncovered in a challenging research paradigm. It appears that the impact of a structured inquiry approach is limited without the opportunity for students to integrate their discoveries in a discourse that challenges their growing ideas. The study supports the notion that social interaction plays a powerful role in building knowledge. Well grounded in socio-constructivist paradigms (Vygotsky, 1978), this finding points to the importance of language and dialog in shaping cognitive processes. In this specific case, class discus-

Figure 8
Individual microdevelopmental pathways of interviewees.



Note. Stars on the x axis represent the beginning of an interview.

sions may have been particularly effective in forcing students to evaluate the validity of their ideas by confronting the arguments of their peers. Instead of relying on the traditional superimposition of the "scientific truth," delivered by the teacher and layered over existing preconceptions, students had to start with these preconceptions and work together to develop a coherent framework to think about the conservation of matter. This new challenge unfolded in the evolving dialog with their peers. This study, however, did not aim at investigating the mechanisms underlying how discussion facilitates learning, and so these suggestions about how language might have played a role are speculative and would require further empirical study.

The impact of guided inquiry and argumentation on students' conceptual understanding of conservation of matter points to a potentially powerful intervention for educators. Although both teaching approaches lead to statistically significant increases in test scores, the AC produced a clearly larger effect size. Surprisingly, the increase in test scores was similar in magnitude, whether or not guided inquiry preceded argumentation. The results of this study seem to support Abell et al.'s (2000) claim that shifting the emphasis from "science as exploration and experiment" to "science as argumentation and explanation" is an appropriate move to foster students' conceptual understanding of conservation of matter. However, as Harrison and Treagust (2001) point out high scores on a questionnaire do not necessarily reflect students' actual conceptual understanding. In the second year of the study, six interviewees provided a window into the learning process. The observable difference between Larissa and Henry's microdevelopmental pathways clearly illustrates the point raised by Harrison and Tregust: both students obtained 87% on the posttest, but only Larissa's pathway displayed clusters of high-level performances and a continuous increase in average skill level.

We agree with Harrison and Treagust (2001) that test scores are limited measures of students' conceptual understanding and for that reason we attempted to capture a qualitatively richer picture of a few students' understanding by performing a skill analysis of interviews. As Yan and Fischer (2002) point out, the resulting microdevelopmental pathways "are like fine documentary movies recording each individual's learning history, whereas the mean and standard deviation are like a few blurred snapshots sketching group members' silhouettes" (p. 148). For example, Mel's microdevelopmental pathway visually recorded the unfolding of his "strategic understanding" pattern of responses, a phenomenon that would have been impossible to capture with only test scores. The "fine documentary movies" gathered through the skill analysis of student dialogue reveal some of the cognitive structure underlying the dramatic change observed on the pre–post questionnaires. Unfortunately, the nature of schools in general does not facilitate the development and use of such time-consuming and labor-intensive methods of assessment. Studying the learning process in order to facilitate this process highlights a challenge for educators and researchers.

Infrastructures of Change

Although schools are facing serious time constraints, only research can help educators effectively confront the challenge of determining how best to use what time is available. To successfully test hypotheses or search for meaningful patterns in student behavior, however, researchers and practitioners need a supportive context. Given that educational systems encounter and endure numerous forces and changes, the focus during this study often shifted to envisioning an infrastructure that could respond to political demands, parental concerns, and administrative needs. Indeed, although challenging, the context of this study also provided a window into the process that leads to potentially promising educational interventions. The process itself became the subject of scrutiny and possible research.

What this study reveals is the complex interaction of forces that impact how educational research unfolds. Although we did not intend to study the engine (which contained a few "gears" that we wanted to understand), we found ourselves considering how the engine works.We observed that learning was a community process that involved not only the students but also the teachers and researchers. The only voice missing in the ongoing dialog was the administration's. Although they tried to accommodate the teachers and researchers, they had no way to evaluate the effectiveness of self-imposed constraints in the unfolding methodology. The administration's decisions made sense given the political climate they faced, but had they been able to experience the doubt accompanying the study outcomes, they might have been more empowered to face the board and parents with a more compelling argument for why research needs to unfold in a more meticulous manner.

As highlighted several times in this article, the validity of our results is weakened by the absence of control groups, the methodological changes in response to evolving political pressures, and other unfortunate (but typical) setbacks that undermine the methodological soundness of the study.During the first year, one teacher withdrew from the study before the AC because of time pressure: she felt she could not spare the hours needed for the AC and still cover the regular curriculum. During the second year, the second grade 8 teacher was new to the school and preferred not to engage in a teaching experiment.These problems are inevitable in a dynamic system like school districts, but still it is important to ask: how do the members of such a system (i.e., schools, universities, and legislative bodies) achieve individual success while contributing to the more general goal of improving education?

The dynamic environment needs to be the subject of research, and a research question that follows might be: What association of people and rules best address and respond to the ongoing and emerging forces that impact researchers, educators, students, and those who administer the system? Alternatively, what is the most effective infrastructure of people, purposes, and resources that can support and test proposed changes in education?

One Infrastructure for Change: The RSN

The creation of a RSN has become one ongoing experiment that has emerged in response to this study and the research questions proposed. The RSN, as an "infrastructure for change," has become a working hypothesis. The RSN seeks to solve problems at multiple levels of analysis—the classroom, the school, the district, and the state. It seeks to support the ongoing learning process of teachers like those in this study who learn from participating in research. Once these teachers recognized the complexities that emerge in research and the choices that are required to create rigorous findings, they were willing to invest their own time and their students' in the research process.

Although there are no data to report at this time, this RSN infrastructure is growing to manage the problems encountered in this study. In the same community the dean of the college of education and the local school superintendent have agreed to create an executive council consisting of members of both institutions. In turn, this council identifies problems and resources in the educational community. Its members address how to ensure individual success while focusing on more strategic goals for the community. The council seeks to work with its local state representative to coordinate new research and its potential findings with state laws. If such coordination is unfruitful, then the RSN looks to create parallel pathways for students and teachers so that the work of creating rigorous findings does not penalize anyone taking the risk of changing course, challenging assumptions, and testing hypotheses. The council seeks ways of increasing trust in this larger educational community by initially sharing resources through talks or workshops open to all teachers, as a basis of starting conversations that ultimately uncover the problems that need to be solved. The same conversations help teachers recognize where they fit along a continuum of interest ranging from those who prefer to be passive consumers of research to those who want to participate in research to those who wish to return to the university and become more proficient in producing new knowledge.

In parallel the university is creating a master's degree in Mind, Brain, and Education (MBE) to encourage teachers to take part. To support this process, the College of Education offered scholarships for the first course in the MBE program so that teachers could decide if this degree and program is the right path to follow. At the same time, the executive council is establishing a working model of communication for granting institutions, so that those attempting to understand this infrastructure of change can visualize how classroom problems can translate into research ideas. Ultimately the RSN can be a context where everyone is vested in the process of educational reform, as well as in its outcomes, because both the process and outcomes are products of the community.

Conclusion: Unpacking the Learning Process

The study reported here is in part a serendipitous event that allowed us to consider the importance of the learning process for everyone involved in the study—teachers, researchers, and students. Additionally we were able to partially tease apart the relative impact on student learning of two components of a curriculum exploring the conservation of matter—guided inquiry and argumentation. Although both teaching strategies resulted in positive results, the argumentation approach was much more promising. These results, however, lead to many open-ended questions. Indeed, researchers need to explore howmuch guided inquiry isnecessary to render argumentation effective. Similarly, how does the topic being taught influence the balance of approaches when seeking to achieve the most expeditious results? Should the balance of attention given to both approaches shift as students and teachers move to different topics in science? These are compelling questions in our current culture of testing, which has become more complicated as districts consider how tightly they can compress the curriculum.

Research such as this suggests possible strategies and highlights the need to find ways to measure how, how well, and how much student understanding changes as a result of different teaching methods. Such research as reported here would be best served by concurrent research exploring the nature and depth of students' understanding.

The purpose of this article was as much to report promising findings as it was to propose a new level of analysis in exploring how to face the challenges and requirements of researchers and practitioners in education. These challenges could be met in emerging "infrastructures for change" we call the RSN. Indeed, the network is the product of collaboration between researchers, educators, and policymakers attempting to strike the right balance between individual needs and the conditions necessary to work together to find solutions to educational issues that the community is facing.

References

Abell, S. K., Anderson, G., & Chezem, J. (2000). Science as argumentation and explanation: Exploring concepts of sound in third grade. In J.Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 65–79).Washington, DC: American Association for the Advancement of Science.

Abraham, M., & Williamson, V. (1994). A cross-age study of the understanding of five chemistry concepts. *Journal of Research in Science Teaching*, *3*, 147–165.

Agung, S., & Schwartz,M. S. (2007). Students' understanding of conservation of matter, stoichiometry and balancing equations in Indonesia. *International Journal of Science Education*, *29*, 1679–1702.

Barnes, D., & Todd, F. (1995). *Communication and learning revisited: Making meaning through talk*. Portsmouth, NH: Heinemann.

Bloom, J.W. (2001). Discourse, cognition, and chaotic systems: An examination of students' argument about density. *Journal of the Learning Sciences*, *10*, 447–492.

Benjaoude, S., & Barakat, H. (2000). Secondary school students' difficulties with stoichiometry. *School Science Review*, *296*, 91–98.

Bowen, G.,& Bunce, D. (1997). Testing for conceptual understanding in general chemistry. *The Chemical Educator*, *2*(2), 1–17.

Bryk, A. S., & Gomez L. (2008). Ruminations on reinventing an R & D capacity for educational improvement. Prepared for the American Enterprise Institute Conference, *The Supply Side ofSchool Reform and the Future of Educational Entrepreneurship*, October 25, 2007.

Coch, D., Michlovitz, S. A., Ansari, D., & Baird, A. (2009). Building mind, brain, and education connections: The view from the upper valley. *Mind, Brain, and Education*, *3*, 27–33.

Dawson-Tunik, T. (2004). "Good education is. . ." The development of evaluative thought across the life span. *Genetic, Social, and General Psychology Monographs*, *130*(1), 4–112.

DESIGNS II. (in press). *Physical science course*. Kendall Hunt: Harvard Smithsonian Center for Astrophysics.

Driver, R., Squires, A., Rushworth, P.,& Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. New York: Routledge.

Fischer, K.M. (2000). Inquiry teaching in biology. In J.Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 258–280). Washington, DC: American Association for the Advancement of Science.

Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of psychological structures in action and thought. In R. M. Lerner (Ed.), *Handbook of child psychology, Vol. 1: Theoretical models of human development* (5th ed., pp. 313–399). New York:Wiley.

Gomez, M.A., Pozo,J.I., & Sanz,A. (1995). Students' ideas on conservation of matter: Effects of expertise and context variables. Science Education, 79, 77–93.

Granott, N., & Parziale, J. (2002). Microdevelopment: A processoriented perspective for studying development and learning. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 1–28). Cambridge, UK: Cambridge University Press.

Hameyer, U., van den Akker, J., Anderson, R. D., & Ekholm, M. (1995). *Portraits of productive schools: An international study of institutionalizing activity-based practices in elementary science*. New York: State University of New York Press.

Harrison, A. G., & Treagust, D. F. (2001). Conceptual change using multiple interpretive perspectives: Two case studies in secondary school chemistry. *Instructional Science*, *29*, 45–85.

Hinton, C., & Fischer, K.W. (2009). Research schools: Grounding research in educational practice. *Mind, Brain, and Education*, *2*, 157–160

Holliday, W. G. (2004). A balanced approach to science inquiry teaching. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and the nature of science: Implications for teaching, learning, and teacher education* (pp. 201–218). Boston: Kluwer.

Jiménez-Aleixandre, M. P., & Erduran, S. (2008). Argumentation in science education: An overview. In S. Erduran & M. P. Jim´enez- Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 3–28). Dordrecht, The Netherlands: Springer.

Karmiloff-Smith, A., & Inhelder, B. (1975). "If you want to get ahead, get a story." Cognition, *3*, 195–212.

Kuriloff, P., Reichert, M., Stoudt, B., & Ravitch, S. (2009). Building research collaborations among schools and universities: Lessons from the field. *Mind, Brain, and Education*, *3*, 34–44.

Lawson,A.E. (2003). The nature and development of hypotheticopredictive argumentation with implications for science teaching. *International Journal of Science Education*, *25*, 1387–1408.

Lemke, J. L. (1990). Talking science: Language, learning, and values. Norwood, NJ: Ablex.

Magnusson, S. J.,Palincsar, A. S.,& Templin, M. (2004).Community, culture, and conversation in inquiry-based science instruction. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and the nature of science: Implications for teaching, learning, and teacher education* (pp. 131–156). Boston: Kluwer.

Martin, D. J. (2000). Elementary science methods: A constructivist approach. Belmont, CA: Wadsworth.

Minstrell, J. (2000). Implications for teaching and learning inquiry: A summary. In J.Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 471–493). Washington, DC: American Association for the Advancement of Science.

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

Ozmen, H.,& Ayas, A. (2003). Students' difficulties in understanding conservation of matter in open and closed-system chemical reactions. *Chemistry Education: Research and Practice*, *4*, 279–290.

Pfundt, H.,& Duit, R. (Eds.). (2000). *Students'alternativeframeworksand science education*. Kiel, Germany: Institute for Science Education (IPN).

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

Rappolt-Schlichtmann, G., Tenenbaum, H. R., Koepke, M. F., & Fischer, K. W. (2007). Transient and robust knowledge: Contextual support and the dynamics of children's reasoning about density. *Mind, Brain, and Education*, *1*, 98–108.

Sawrey, B. A. (1990). Concept learning vs. problem solving:Revisited. *Journal of Chemical Education*, *67*, 253–264.

Sadler, P. S.,& Schwartz, M. S. (2004). *Research tools to evaluate student understanding*. Cambridge, MA: Harvard-Smithsonian Center for Astrophysics, Science Education Department.

Schwartz, M. S. (2009). Cognitive development and learning: Analyzing the building of skills in classrooms. *Mind, Brain, and Education*, *3*, 198–208.

Schwartz,M. S., & Fischer, K.W. (2004). Building general knowledge and skill: Cognition and microdevelopment in science learning. In A. Demetriou & A. Raftopoulos (Eds.), *Emergence and transformation in the mind: Modeling and measuring cognitive change* (pp. 157–185). Cambridge, UK: Cambridge University Press

Schwartz,M. S., & Sadler, P. M. (2007). Empowerment in science curriculum development: A microdevelopmental approach. *International Journal of Education*, *29*, 987–1017

Slavings, R., Cochran, N., & Bowen, C. W. (1997). Results of a national survey on college chemistry faculty beliefs and attitudes of assessment-of-student-learning practices. *The Chemical Educator*, *2*(1), 1–55.

Tobin, K. (2006). Verbal and non verbal interaction in science classrooms. In K. Tobin (Ed.), *Teaching and learning science* (pp. 79–90).Westport, CT: Prager

Tobin,K., Tippins,D. J., & Gallard, A. J. (1994). Research on instructional strategies for teaching science. In D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 43–93). New York: MacMillan.

Vellom, R. P., & Anderson, C.W. (1999). Reasoning about data in middle school science. *Journal of Research in Science Teaching*, *36*(2), 179–199.

von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, *45*(1), 101–131.

Vosniadou, S. (2007). Conceptual change and education. *Human Development*, *50*, 47–54.

Vosniadou, S. (2008). Bridging culture with cognition: a commentary on "culturing conceptions: from first principles". *Cultural Studies of Science Education*, *3*(2), 277–282.

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

Yan, Z., & Fischer, K. W. (2002). Always under construction: Dynamic variations in adult cognitive development. *Human Development*, *45*, 141–160.