Using a Comparative, Longitudinal Study with Upper Elementary School Students to Test Some Assumptions of a Learning Progression for Matter

Carol L. Smith, University of Massachusetts/Boston
Marianne Wiser, Clark University
David W. Carraher, TERC

Paper Presented at NARST, March 24, 2010

Related paper set: Evaluating proposed learning progressions: What can we learn from cross-sectional data and longitudinal studies?

Email: Carol.Smith@umb.edu

DRAFT: Please do no quote without permission of authors

The study is part of the Inquiry Project, supported by MSP Grant#0628245, NSF-HER-DRL.

For information about the Inquiry Project, including its curricular materials go to: http://inquiryproject@terc.edu.

Using a Comparative, Longitudinal Study with Upper Elementary School Students to Test Some Assumptions of a Learning Progression for Matter

Carol L. Smith, University of Massachusetts/Boston
Marianne Wiser, Clark University
David W. Carraher, TERC

Abstract

A learning progression is a research-based proposal for how ideas about a content domain could coherently evolve over long period of time, given appropriate instruction, to bridge a lower and an upper anchor. Our project is concerned with elaborating on a Learning Progression for Matter for children in grades 3-5 of elementary school. We have been guided by the view that many of the later difficulties students have with the atomic molecular theory stem from limitations in their *macroscopic* understanding of matter and that developing these macroscopic understandings involves deep and broad *reconceptualizations* of children's knowledge. We also believe that elementary school students can build these understandings with supportive curricular units, designed from an LP perspective. To test this assumption, we have been designing such curricular units as well as more in depth clinical assessments, and conducting a three-year longitudinal study in which we compare the progress of students within the same school who have our new units with those who do not. This paper discusses the results of first two years of this study regarding the initial conceptual difficulties of students, the steps they take in reconceptualization, and the effectiveness of our new units in fostering this process.

Rationale

Researchers in science education and cognitive development, science educators, curriculum and assessment designers, funding agencies and science education policy makers are paying a good deal of attention to the learning progression approach to science and math education (Corcoran, Moser, & Rogat, 2009, National Research Council, 2007).

Learning progressions focus on "big ideas," i.e., concepts or theories central to a discipline (e.g., evolution, the atomic molecular theory, energy); they characterize the acquisition and development of those big ideas over a large time span (e.g., K-12). Different researchers think of LPs in different ways. Our own view is that a LP is a research-based proposal for how ideas about a content domain could coherently evolve over long period of time, given appropriate instruction, to bridge a lower and an upper anchor (e.g., knowledge at preschool and desired knowledge at the end of high school). A central assumption is that bridging between the lower and upper anchor requires a series of broad reconceptualizations, because the concepts, beliefs, models, and practices of kindergartners are fundamentally different from those of scientists. We believe that the evolving knowledge is not simply becoming more and more complex or "scientific", but instead identify key stepping stones, i.e., states of knowledge that are not necessarily consistent with the scientific view, that allow students to later understand the scientific view, with proper instruction. We view those stepping stones as important intermediate "targets" of instruction that are currently overlooked.

In an initial paper, Smith, Wiser, Anderson, and Krajcik (2006) proposed a broad Learning Progression for Matter (K-12) by working "backwards" from the (simplified) atomic-molecular theory, which is part of the American high-school curriculum. They asked “What prior knowledge would allow students to understand the atomic-molecular theory?” They concluded from a review of the literature about students’ difficulties with the atomic-molecular theory, and innovative curricula aimed at remedying those difficulties that, while the sources of some difficulties lay in the atomic-molecular theory itself and require innovative approaches to teach it, others are to be found in students’ understanding of matter at the macroscopic level and in their epistemology, which are typically not sufficient to support a meaningful understanding of the molecular-atomic model. Thus their learning progression for matter (LPM) proposed building a set of “big ideas” about matter at a macroscopic level during the elementary school years, which included important ideas about measurement and modeling, to serve as an appropriate foundation or stepping stone for learning about the atomic-molecular theory of matter in the middle and high school years.

We believe achieving this stepping stone (of a sound macroscopic understanding of matter) is itself a complicated process dependent upon instruction that successfully fosters radical conceptual change. That is, students’ starting ideas about size, weight, and materials are incommensurable with those of scientists, as they are rooted in everyday sensory experiences and an epistemology based on trusting those senses to give reliable information, rather than supported by deeper theoretical and mathematical analysis. For example, young children’s ideas about weight are centered on how heavy things feel; they have developed few theoretical beliefs about why things weigh what they do (weight is often taken as something that just is, rather than something that needs to be explained), and they learn about the weights of objects by hefting (rather than measuring). Similarly, their initial ideas about object size are also based on observation, rather than careful measurement. These initial concepts of weight and size do not support the development of the belief that all matter has weight—a belief that is critically important in coming to use evidence of weight changes or constancy in tracing matter over time, since many light objects appear to “weigh nothing at all” nor do they support the differentiation and inter-relation of weight and density as distinct magnitudes.

The Inquiry Project. Currently, we are involved in a collaborative project that is elaborating on a LPM for children in grades 3-5 of elementary school. Building on prior LPM work, we are further specifying how children’s macroscopic concepts of material, matter, weight, volume, and density might evolve during the elementary school years with supportive curricular units, designing such curricular units as well as more in depth clinical interview assessments, and studying the progress elementary school students make with these concepts over a three year span of time. We are also studying the progress that other students in the same schools make with their traditional science curriculum.

We propose that the needed reconceptualization is extensive. Indeed, we posit that the development of these macroscopic understandings of matter involves two inter-related advances in students' thinking about matter and related concepts:

- A gradual shift from (a) *perception-centered thinking*, that is, understanding and explanation closely tied to perceptual judgment and appearances, to (b) *model-mediated thinking*, informed by views about matter and drawing upon a set of increasingly advanced, inter-related concepts and scientific habits of mind.
- The development of *quantitative reasoning and understanding of measurement* that students can use to make predictions, interpret, and explain relationships among physical

> quantities. Additive scalar quantities such as volume, mass, and length are gradually re-conceptualized as dense linear dimensions (Whitney, 1968a, 1968b) for which a metric supports additive comparisons (differences, e.g. "three cc more") and multiplicative comparisons (ratios, e.g. "three times the volume of") in terms of measures and their respective quanta. Students also learn to describe relationships between dimensions of different natures. Graphs and other representational systems will depict functional relations in a coordinate space.

Elsewhere we have written extensively about how the LP approach affected the design of our particular science curricular units (Wiser, Smith, and Doubler, in press) and how LP-inspired curriculum units differ from traditional ones. One difference is a focus on "advancing" a broad network of inter-connected concepts and beliefs, rather than on "coverage" of key topics. That is, our goal is to move a large knowledge network forward without destabilizing it rather than expose children to a series of independent experiences that do not allow a systematic and orchestrated development of inter-related concepts. This in turn requires appreciation of complexity of ideas, and the ways they are inter-connected, rather than working on topics in isolation.

For example, important aspects of reconceptualizating weight are understanding its measurement, differentiating scale weight from hefted weight, and making the relation of weight and material more complex-- from "iron objects are heavy" to "iron objects are heavy for their size" and then "iron is dense." Note that concepts take part of their meanings from each other—e.g., density can be attributed to material only after material is clearly differentitated from object and it meaning has changed from a collection of surface properties (aluminum is shiny and light grey and feels cold to the touch) to the stuff the object is *made of* (an object made of aluminum is aluminum all the way through; it can mentally be decomposed into little pieces of aluminum). This requires encountering weight, material, and size in the same curricular unit, through activities in which they are linked.

An LP-inspired curriculum also differs from traditional curricula by identifying productive entry points--concepts that are salient in students' everyday thinking (although they need restructuring) and therefore can be used to "leverage" change. Based on previous literature we chose to start the curriculum with weight and material and to build on these precursor concepts first. The concepts of density, volume and matter itself are more effectively developed when they are leveraged by reconceptualized concepts of weight and material and by a better understanding of measurement. Thus, we fore-grounded the concepts of material and weight in grade 3, as children explored solid objects (rather than liquids and powders), which are prototypical material entities in the sense that children know they have weight and size. Among them were density cubes, a set of cubes all the same size, but made of different materials, which embody the relation between weight and material in the simplest way. Children actively inter-related these concepts to each other and other concepts in their knowledge network in class discussion as they worked to understand phenomena and met challenges, and used new representation tools (e.g., a weight line) to extend these ideas in productive ways.

Considerations of object size and heaviness for size for the density cubes was present, but in the background. We then extended students investigations in grade 4 to explore a range of Earth materials (rocks, powders and liquids). In these investigations, considerations of volume, heaviness for size, came to the fore, along with considerations of materials and weight. In the grade 5 curriculum (which is currently being implemented) we are extending investigations to consider phase changes in the water cycle, the material nature of gases, and conservation of matter across phase change. Our assumption is that the extensive prior work children have done exploring more readily accessible materials in the 3rd and 4th grade, has prepared them for taking this next step. Typically, these topics are introduced without the

Elsewhere, we have also written about how the LP approach has influenced our design of assessments (Carraher, Smith, Wiser, Schliemann, & Cayton-Hodges, 2009). Because we conceptualize LPs as involving reconceptualizations of dense networks of inter-related concepts, we designed assessments that probed different facets of (potentially) related concepts. The tasks in our interview probe different aspects of a broad network—including material, weight, size, matter, density, number, fraction, and ratio. The same interview is given every year to our participants, so we could trace changes in their understanding over time. The goal of the assessments is to capture changes in concept articulation, rather than test the knowledge of specific facts that might have been learned during a particular unit. Our assessment deliberately probes for ideas in informal (non technical ways) so that we might uncover relevant precursor ideas before they are covered in class and implicit understandings that could play a role in later learning. By using the same assessment over time, we can also assess how change in one idea impacts change in others.

In this paper, we will consider what we are learning based on the first two years of our three year comparative longitudinal study. Comparative, longitudinal studies are particularly useful in evaluating three core assumptions of this LPM, by examining whether:

(1) Understanding matter and related concepts is NOT well developed in elementary school children with existing curriculum;

(2) Achieving inter-related concepts of matter, material, weight, volume, density and states of matter is within the grasp of most elementary students with targeted science curricula that work over multiple years (developed from LP perspective); and

(3) Understanding matter at the macroscopic level is an important stepping stone for learning about the atomic-molecular theory in the middle school years.

Our longitudinal study focuses on evaluating assumptions one and two, which lays the groundwork for future studies that test assumption three. Although there have been some successful attempts to develop innovative science curricular units on these topics for middle school students (e.g., Lee, et al., 1993, Smith, Maclin, Grosslight, Davis, 1997, Smith, 2007), much less has been done with assessing the impact of innovative science curricular units on these topics in the elementary school years. An exception is the pioneering work of Lehrer, Schauble, Strom, & Pligge (2001) which showed the progress that grade 5 children could make in modeling material kind (and constructing inter-related concepts of weight, volume, and density), with an innovative, sustained, and integrated math and science curriculum. Whether (and how) elementary school children can make progress on these issues through revisions only to their science curriculum remains an open question. Through our comparative and exploratory longitudinal study, we hope to learn more about what may be the "productive stepping stones" in helping elementary school students build a sound macroscopic understanding of matter.

Design/Procedure

Our study uses a quasi-experimental, comparative, longitudinal design where each student receives the same two-hour individual interview at multiple points in time, although sub-tasks are structured so that the questioning can be adjusted somewhat to the thinking of the student. Clinical interviews allow us to present children with a variety of real-world materials, and to pose task questions around those materials. This is particularly important when working with young children and when trying to assess early forms of understanding that might be present before they have learned specialized vocabulary.

Treatment students (those who received the LP inspired science curriculum for nine weeks in each of grades 3, 4, and 5) are interviewed at four moments over two and one-half years: (a) early grade 3, before the new science unit; (b) end of grade 3 after the first unit; (c) end of grade 4 after the second unit; and (d) end of grade five after the third unit. These students are from 5 classrooms in two different schools that involve students from urban populations.

Control students (students from the same school who are receiving the standard science classroom instruction in grades 3-5) are interviewed at the end of grade 3, 4, and 5.

The interview consists of 10-multi-part tasks administered in two one-hour sessions. Eight of the tasks probe various concepts related to matter (matter, amount of material, amount of matter, weight, volume, density) in multiple ways; portions of these tasks will be the focus of discussion here. These include (1) asking whether tiny things (a visible speck of clay, an invisible speck of clay) take up any space or have weight; (2) making judgments about and developing measures of "the amount of space filled up by" two blocks (volume) using a variety of tools (e.g., tiles, cubes, a broken tape measure, paper clips of various sizes); (3) explaining why smaller (covered cylinders) could be heavier or the same weight as larger cylinders; making judgments (for visible cylinders made of Delrin, brass, and aluminum of) about whether or not one object in a pair is made of a heavier kind of material; and conducting investigations (using covered and uncovered cylinders of various sizes) to determine whether the covered objects could be made of aluminum, brass, or Delrin; (4) making predictions about how high the water level will rise when two cylinders (the same size but different weight) are put in; (5) making judgments about conservation of properties of balls of clay after reshaping (e.g., weight, volume/how much space it takes up); (6) inferring the volume, weight, and amount of plastic in two block rearrangements; (7) sorting items by matter/not matter as well as by whether they contain atoms and molecules (if children have heard of them); and (8) making judgments of the effects of transforming (e.g., grinding, melting) materials on the identity of the materials and the preservation of some of its properties.

Data are presented for those subjects for which we have complete data: 45 Treatment students who were interviewed at 3 points of time (at the beginning of grade 3, and after completing the grade 3 and grade 4 units) and 27 Control students who were interviewed at two points in time (at the end of grade 3 and grade 4). (In all we have189 complete interviews.) The treatment and control students were equally balanced across the two schools, with 60% coming from the school with three classrooms, and 40% coming from the school with two classrooms. Students were selected for interviews based on their returning of permission slips granting permission to be interviewed.

Students were interviewed by a team of experienced interviewers (graduate students and research assistants) who were knowledgeable about the ways young students may think about these issues, and hence were able to probe in relevant ways. At the same time, they were not centrally aware of the specific research hypotheses. Interviewers recorded student answers to questions and descriptions of student activities as they went along; in addition interviews were video-taped, so that interviewers could review the tapes to check student responses before entering the data in a data base.

Coding of interviews was based in some cases on patterns of judgments (e.g., did they judge that a tiny speck of Playdoh weighed a tiny bit or nothing at all), in some cases on student explanations of their judgments (why did they think that), and in other cases on how students worked with objects or materials (e.g., how they used a set of one inch cubes to find how much space two different shaped blocks took up, etc.) Coding categories were informed both by theoretical considerations based on prior research as well as by what we found noteworthy and interesting or even surprising in the data. In this way, our study is still very much an exploratory one. In developing the coding systems as well as in coding the interviews, we were blind as to the grade and treatment status of the child (interviews were assigned random numbers; in coding we worked with ascending random number which included interviews of all grades and treatment status). In cases where interpretation or judgment was involved (as in coding justifications or patterns of activity), two coders first reviewed a subset of the data to develop the initial coding system and a larger subset to establish reliability; in cases where the reliability was very high (typically over 90%), one coder then proceeded to code the rest of the data. In some cases that relied exclusively on coding student explanations, we had two coders code the entire corpus, even though agreement was generally high (85-95%). Disagreements were resolved via discussion.

For some items, we distinguished different patterns of responding that we thought reflected qualitatively different ways of thinking about or approaching the problem, and where appropriate ordered them for degree of sophistication. To validate that these different judgments typically reflect qualitatively different ways of thinking about an issue, we also coded student justifications, so that we could correlate changes in judgment pattern with justification pattern. (This work has been completed for some tasks and is ongoing for others.) For other items, our analysis focuses more exclusively on student justifications.

We used the Wilcoxon matched pairs signed-ranks test to assess within student progress in understanding for an item. This allowed our assessment to be sensitive to progress that students were making in becoming more sophisticated, even when they did not achieve "canonical" understanding. We used chi square tests to compare the response distributions of patterns on a given item (or set of items) between the Treatment and Control samples.

Results

Although many finer grained data analyses are still ongoing, the results of the first two years of the study provide compelling evidence that:

(a) Children in our sample *start with* macroscopic concepts of weight, size, material and matter *at odds with scientists' conceptions*; this supports our assumption that new curricula may be needed to target these conceptual difficulties that are typically overlooked as important in traditional science curricula;

(b) Children receiving the Inquiry Project curriculum made *clear progress* in both grade 3 and grade 4 on *multiple aspects* of conceptual understanding of matter, weight, volume, and density; these changes provide evidence that they are engaged in a productive, but long-drawn process of reconceptualizing their matter network;

(c) Control students (receiving the standard science curriculum used in their school) made very little progress on items calling for major reconceptualization (as opposed to elaboration). This is further evidence that instructional support is needed in developing these ideas; these ideas do not simply come "for free" with development, as some have assumed, nor are they well developed with standard math and science curriculum.

In the sections that follow, we amplify on the above three points. We also consider some initial evidence that developing an understanding that even tiny things take up space and have weight functions as one of the important stepping stones within the curriculum.

Conceptions of Treatment Students on the Grade 3 Pre-Interview (Table 1)

Our interviews revealed that at the start of third grade, our treatment children's initial understandings of material, weight, and size were grounded in perceptual experiences with objects and materials rather than tightly inter-related in an explicit model of matter. Further, they did not yet conceptualize physical quantities as dense linear dimensions for which a metric supports additive and multiplicative relationships, had almost no conceptual understanding of fractions as numbers, and little conceptual understanding of measurement. A distinct concept of volume (i.e., a concept of volume differentiated from length, area, shape, and weight) is almost totally absent, as is a concept of density differentiated from weight. Thus the conceptual relations in these students' knowledge network were very different from the relations observed among experts. Overall, they also lacked a coherent concept of matter as something that includes solids, liquids, and gases, and as something that takes up space and has weight and did not think of materials as varying in density.

Let's now consider some of the specific evidence for these claims (see Table 1).

In keeping with our assumption that concepts of weight, size and material would be more accessible than volume, density, and matter, most young grade 3 students understood that one could use a balance scale as another way (from hefting) to compare the weight of objects. They also could correctly compare the sizes and weight, when the size and weight differences were perceptually obvious and salient (e.g., a small brass cylinder and a larger plastic one). Most also thought that changing the shape of a clay ball (from a ball to a pancake) did not change the object's weight or the amount of clay in the object, and most understood that grinding wood and iron produced small pieces of the same material.

At the same time, there is evidence that the students' concepts of weight and material were perceptually-based and context-dependent, rather than closely inter-connected in a model of matter, and therefore in need of further elaboration and restructuring. For example, it was common for children to give different predictions about whether the clay ball would be heavier or lighter than the clay pancake when hefted or when placed on the balance scale, even though they talked about weight in both situations. Thus, although approximately 80 percent of children judged that the ball and pancake had the same amount of clay and another 80 percent judged that they weighed the same, only about half systematically judged that the ball and pancake had the same amount of clay, AND that they weighed the same, AND that they would balance on the scale. Similarly, although most students judged that ground up wood was still wood, that ground up iron was still iron or that melted butter was still butter, fewer preserved material identity in all three cases, and less than half made relevant inferences about that some non-perceptual properties would be preserved (for example, that sawdust would burn or that iron powder would be attracted to a magnet).

One of the clearest indications of young children's more perceptually centered way of thinking about weight came in the interview's first task where we handed children a large pink ball of clay, and asked them if it weighed anything. After the child said that it weighed something, we broke off a tiny piece and handed it to the child, asking whether this tiny piece would weigh a tiny bit or nothing at all. At this point, rather than reason more theoretically that the tiny piece must weigh something because it is some physical stuff, and that if the large ball weighed something each individual piece must have some weight, virtually all the children (over 90%) judged that the tiny (although visible) piece weighed nothing at all. In justifying their answer, they typically commented on its perceptual insignificance—"it is too small (or light) to weigh anything," or, "it feels like nothing in my hand." A few even used their new knowledge of the scale to support this conclusion, noting that it would also have no effect on the scale.

More judged that the tiny piece took up some space, but again their justifications centered on perceptual observations (e.g., I can see it). Even here, over half of the students confidently told us that the small (visible) piece was too small to take up any space even though they could see it and hold it in their hand.

Of course, one could wonder whether this was just a semantic disagreement—what does "nothing" or "a tiny bit" mean to these third graders—and indeed having different ideas about what nothing means is part of the story. (For example, we have evidence in another part of the interview, that our treatment students had little understanding of fractions as tiny numbers, or that there were any numbers at all between two consecutive integers.) But we would argue it is also symptomatic of deeper differences in their concept of weight itself—they do not yet conceptualize weight as a continuous physical magnitude correlated to amount of stuff. For example, in another task children were shown two balls of quite different size and weight (the large ball was 4 times the size and weight of the smaller ball) and they were asked "How many weights do you think there could be between the weights of these two balls?" Rather than answer "as many as you want," most children shocked us by saying there could be only one or two!

Other evidence that children have limitations in their "explanatory" knowledge about weight comes from a task in which children were handed different size covered cylinders that varied in weight. We started with a large cylinder that was heavier than a smaller one, and asked why they thought it was so much heavier. In response to this question, most mentioned that it was heavier because of its size (it's bigger, taller, larger, etc.). We then brought out new cylinders where differences in size could not explain differences in weight and asked children to explain these "puzzling" cases (e.g., two same size cylinders with very different weights, a small cylinder much heavier than a much larger cylinder, and finally two cylinders of very different sizes that had the same weight). Some children were at a loss to explain these cases at all, although a common response was to consider that one cylinder might be empty while the other was filled (or that the two cylinders might be filled with different objects). When we told children that in fact the cylinders were solid all the way through and asked them if they could think of other explanations, quite strikingly most could not. (The cylinders were made of materials of quite different densities—brass, aluminum, and a hard plastic. We used covered cylinders because we wanted to probe children's explanatory ideas unprompted by observable perceptual cues.)

Of course in everyday life, it is common for children to have many experiences with objects where their weight differences are explained by whether they are empty or full (e.g., full suitcases weigh more than empty ones, as do soup cans). It may be less obvious to children that differences in materials are also relevant, for several reasons. First, they may have less experience with "chunks" of materials (rather than materials fashioned into different objects which may be hollow or not). Second, it may also be that one cannot see the insides of solid objects and therefore rule out they may be hollow. Given the salience of the empty/full explanation, children may naturally "project" this explanation in cases where it doesn't apply. Indeed, in one of our later in-class third grade activities with the density cubes (where the aluminum and copper cubes differed in weight), many students predicted that if we cut the aluminum cylinder in half it would be hollow inside!

In contrast, to questions about weight and materials, where we sensed children knew what we were talking about (after all they have been using words for weight and materials since preschool) but simply had very different ideas, in the interview questions designed to probe children's concept of volume, we found evidence that a concept of 3D volume (clearly differentiated from area or length) was simply missing. We designed several tasks to probe children's ideas about volume, without using technical language or vocabulary. In one task, we brought out two blocks of different shapes whose volume could not be easily inferred perceptually (the purple foam block was 8 x 2 x 1 inches, and the wood block was 3 x 2 x 2 inches), and challenged them to use a variety of tools (a ruler, tiles, set of one inch cubes, paper clips) to determine which block was bigger. We used the additional phrase "that is, which one fills (or takes up) the most space" and a sweep of the hand to indicate that we meant how much space it takes up all the way through (rather than how much area it covers on the table). After letting students first choose which "tool" to use in a spontaneous measure, we directed students (if they had not done so already) to think of a way of using the 1 inch cubes to determine how much space each block takes up—in this way, the suggested "tool" might further suggest a 3D sense of space. We found, however, that children were quite creative in using the "cubes" to make linear or area measures of the "size" of the blocks (e.g., measuring the length of one side, the perimeter of the block, or the surface area of one or more face) rather than using the cubes to determine its volume (e.g., by building a little replica of each object, or by measuring and coordinating different dimensions).

Similar difficulties were found in the classic conservation of clay task. In this task, children start with two balls of equal size, shape, and weight; one ball is then flattened into a pancake. In addition to asking children the questions about whether the ball and pancake have the same amount of clay, weigh the same, and will balance on a scale, they are also asked whether the ball and pancake take up the same amount of space. In this task, the most salient spatial dimension is the area of the pancake, and young children overwhelming judge that the pancake takes up more space. Again, attending to the volume of the ball and pancake (or even relating the amount of space the clay takes up with the amount of material) is much less salient to children than the area (which is a more directly accessible perceptual magnitude).

In a third task, we probed children's understanding that water displacement depends upon the volume (not the weight) of the submerged object by presenting children with two cylinders that were obviously the same size and shape but quite different weights, one made of brass, the other of aluminum. Children first predicted what would happen to the water when the aluminum cylinder was placed in the water; the cylinder was then placed in the beaker and they were asked to explain why they thought the water rose up. Next children were asked to predict what they thought would happen when the brass cylinder was placed in another beaker of water (filled to the same level), as well as explain the reason for their prediction. (In this case, children were not allowed to test their prediction, so the same task could be reused in a later interview.) At the time of the pre-interview, more than 90% of the children predicted that the brass cylinder would make the water go up higher because of its greater weight (focusing on dynamic explanations—it presses harder, pushes up the water more, etc.)

Taken together, these findings indicate that children do not just need to learn to measure volume, but also need to develop a concept of volume as 3D occupied space that is both differentiated from and inter-related with other physical quantities. We suspect that developing such a clear volume goes hand-in-hand with theory building about the contexts in which volume is important, and exploiting some of these contexts in learning to measure it.

We also developed three tasks to probe whether children were developing precursors to a concept of density, by probing their understanding of physical situations without use of any technical vocabulary. Again for most children, intuitions about density were generally lacking (although a few were developing relevant insights.) One task (already described) probes children's spontaneous explanations of why smaller objects can be heavier than larger ones. This is one context where children might focus on differences in the kind of material an object is made of, in particular the "heaviness of the kind of material" as a relevant part of their explanation. As mentioned earlier, at the beginning of third grade only a few children spontaneously offered such explanations.

Another task directly probes whether children can infer which of two objects was made of a "heavier kind of material" in situations where they need to distinguish the weight of the object from its density. Children are first given experience comparing the weights of three same size cylinders made of different materials—where the materials are clearly visible and distinguishable by appearance--brass, aluminum, and Delrin (a hard black plastic). They are then given different size pieces of those materials—(a) a small (same size) sliver of aluminum and Delrin (which are both light and indistinguishable in felt weight); (b) a small (light) sliver of brass and a large (much heavier) chunk of aluminum; and (c) a small and large piece of aluminum—and asked to judge whether (or not) one object is made of a heavier kind of material, and if so, which one. At the time of the pre-interview, few children systematically made model-based judgments across all three items, inferring which must be made of a heavier material from the prior relational information; instead the modal response was to focus simply on the weight (or felt weight) of the specific object pairs. In addition, some children showed evidence of undifferentiated patterns, where they shifted between a simple focus on weight to considering relational information across problems (without being aware they were making different kinds of judgments).

In another task, children are asked to figure out what material the different size covered cylinders might be made of (could they be made of brass, aluminum, Delrin, or were they made of something else), when the only uncovered cylinders available for comparison are the larger size (see Figure 1). In fact, one of the four cylinders was made of "something else", two were made of aluminum (a large and small cylinder), and one small cylinder was made of brass. In this task, children have "duplicates" of the small and large cylinders, which would allow them (if they had deemed it relevant) to stack the small cylinders in order to compare their weight to one of the larger cylinders. This (nonverbal) activity of stacking the objects is a wonderful way of showing whether children are beginning to reason about ratio relations in solving this problem. At the beginning of third grade, almost all students made (at best) direct comparisons of the weights of existing (different size) objects, instead of creating “composite objects” in order to create equal size “fair” comparisons with the uncovered cylinders. As a result, they were not able to correctly infer the identity of the materials.

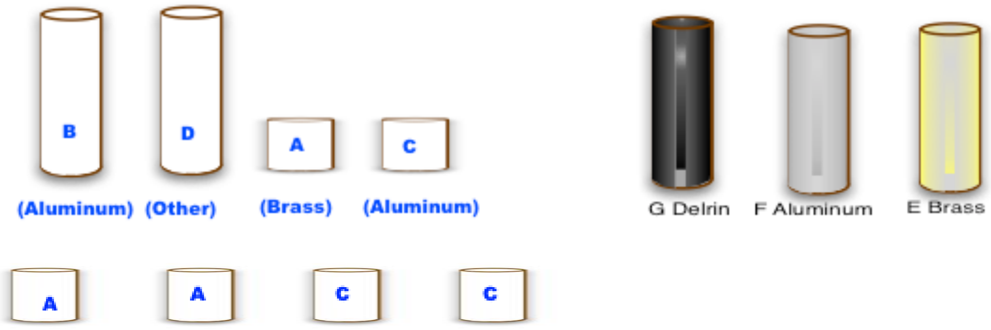

Figure 1. Schematic depiction of the covered and uncovered cylinders used in this task. B and D are the same size (height and cross-sectional area) as G, F, and E. A and C are one-third the size of the larger cylinders. Thus, if one stacked 3 A’s or 3 C’s one had a “composite cylinder” equal in size to G, F, or E so that one could then fairly compare the weights to infer if they were made of the same material.

Finally, children are also given a “matter” “not matter” sorting task in which they sort 14 entities (i.e., wood, ice, dog, mosquito, sand, milk, water, air, steam, smoke, heat, shadow, light, dream) presented as words on cards, into three piles--”is matter,” “is not matter,” “don’t know,” and to explain the basis for their sorting. Because this task uses the word matter, and it is a word most young children do not know (or for which they have alternative meanings, such as whether something important), we begin the task by “glossing” the word as having to do with whether something is made of “stuff,” and providing some anchoring examples (a rock is matter, a plastic horse is matter and a real horse is too; but an idea or wish is not made of “stuff.”) In the 3rd grade pretest, only one child created a “canonical grouping” that included all solids, liquids, and gases while excluding non-material entities and no one used having weight, taking up space, or being made of atoms as the defining characteristics for matter. Some, however, had other systematic grouping patterns that suggested they were developing relevant precursor concepts.

One pattern was to include only a subset of items as matter (e.g., most typically only solids, or some solids and liquids, while excluding air, steam, heat, and light). These children most typically focused on matter as something that you can “see, feel, and touch.” Although this concept is radically different from its scientific counterpart, it does pick out a relevant subset of items and in many respects forms a basis for further investigation, testing and revision, as children first learn these things also take up space and have weight, and then (with further revision to their space and weight concepts and thought experiments about repeated division) come to see that taking up space and having weight trumps being perceptually accessible as a key criterion for the grouping.

Another pattern was a broad grouping that included air, steam, and smoke as well as all solids and liquids but as also some non material physical entities (heat or light.) These children also are well prepared for achieving a scientifically compatible concept of matter, as they learn that matter has weight and occupies space, and that heat and light do not.

However, we found that at the start of third grade, over half of the children created idiosyncratic and (to us) incoherent matter groupings (groupings that simultaneously excluded some material entities such as water, air, sand and also included some non-material entities such as heat, light, or shadow. We do not see these patterns as productive stepping stones for further learning, but rather as evidence of absence of a precursor concept and for the destabilization that emerges when children who start with the intuition that solids and liquids form an ontological category because they can be touched and held, rote learn that air is matter before they are conceptually prepared to understand what this means. (This is not to say that these children will not learn a scientific view of matter but rather than their classification cannot be used as a stepping stone for it.)

Progress Among the Treatment Students (Table 2)

Children receiving the Inquiry Project curriculum made *clear progress* in both grade 3 and grade 4 on *multiple aspects* of conceptual understanding of matter, weight, volume, and density, although they are not yet at ceiling on most measures. These changes provide evidence that they are engaged in a productive but long-drawn process of reconceptualizing their matter network.

Major Changes by the End of Grade 3 (Table 2)

Our interviews reveal that Treatment children made progress on multiple items probing their understanding of weight and materials by the end of third grade. Some (limited) progress was also made in understanding of volume, "heaviness of materials," and matter, although in these cases the majority of students did not yet achieve a canonical understanding. The fact that children's understanding of materials and weight leads volume and density is consistent with the emphasis of the curriculum as well as their being more accessible "lever concepts."

By the end of third grade, children were virtually at ceiling on understanding that grinding did not change materials (91%) as well as understanding that reshaping clay did not change heft, weight measured with a balance scale, or amount of material (84%). Children also made progress in weight measurement, with the majority able to use a set of gram weights to measure the weight of an object.

Lagging behind was that reshaping did not change the amount of space the clay took up (38%). We know from the pretest that these children had a very poor understanding of volume, and were much more likely to compare the sizes of objects in terms of the areas they occupy. The difference in area occupied by the clay ball and the clay pancake was very salient; moreover, fewer lessons in the 3rd grade curriculum were devoted to volume.

Children made dramatic progress in their understanding that tiny things both have weight and take up space (from 10% in the pretest to 60% in the posttest), typically justifying their choice by arguing "Everything must weigh something." Some further explained, "although I can't feel it, I could detect the weight of the piece with a 'sensitive scale'". As evidence that this is a watershed idea, two-thirds of the students who understood that the tiny (visible) piece had weight were also willing to infer that even (invisible) pieces of clay--pieces too tiny to see--would have weight and take up space as well.

Finally, there was also dramatic progress both in their knowledge of specific material kind terms and their use of material kind differences in explaining why larger objects are not always heavier. Indeed, the majority (over 70%) were considering that differences in the kinds of things the cylinders were made of could explain why some were heavier than others, with half started to talk about differences in the "heaviness of the kind of material" and considering specific materials the cylinders might be made of. At pretest, students (at best) used generic material terms (e.g., wood, metal, rock) in their explanations, and (at most) mentioned only one material by name. In contrast, at the posttest their vocabulary was much more differentiated (e.g., poplar, pine, oak, aluminum, brass, copper, PVC, acrylic, etc.), and children were more commonly considering multiple contrasting materials.

Lagging behind was children's ability to attend to the invariant of the volume of the blocks in the volume measurement task and to systematically distinguish heaviness of kind of material from weight (in the judgments about the cylinders). Although children made clear progress in their patterns on these tasks, only about a third of the students had achieved the target understandings by the end of grade 3. For example, in the volume measurement task, of the 40% of students who progressed, half moved from measuring length or area to measuring volume, and the other half moved from measuring length to measuring area. Similarly, in the task calling for children to differentiate heaviness of material from weight of objects, some of the children who made progress moved from the lowest pattern (pure weight judgments) to patterns where they based some of their answers on relational reasoning and others on simple weight judgments, while the other half moved from pure weight or undifferentiated patterns to consistent relational responding. Children made no significant progress on volume displacement (a topic not yet investigated in the curriculum) nor in using stacking of cubes to make inferences about the kind of material the mystery cylinders were made of.

Somewhat surprisingly, children made limited but highly statistically significant progress in their understanding of *matter*, even though the grade 3 curriculum never explicitly discussed the term "matter." This may be a form of "implicit" learning engendered by the curriculum; whether it is truly a sign of progress remains to be seen. More specifically, the number of "incoherent" groupings in children's matter/non matter sorting dramatically declined (from 52% at pretest to only 18% at posttest). Children were increasingly picking out more meaningful sets of items, although the number of fully correct sorting was still negligible. Most typically, they now coherently picked out a *subset* of material items (e.g., solids; solids and sand; or solids, sand and liquids), focusing on things that they thought could be seen, felt or touched. Children were not yet mentioning taking up space or having weight as central properties of matter sorting task.

Major Changes by the End of Grade 4 (Table 2)

By the end of grade 4, Treatment students had maintained many of their understandings of weight and materials (as well as implicit understanding of matter) and continued to improve on the challenging volume and density tasks. In many ways, their understanding of volume and density, which had begun to develop in third grade, seemed to "catch up" with their understanding of weight.

More specifically, students continued to improve on the volume measurement task. Forty percent of students focused on volume, rather than length or area in this task-- all the students who had already done so at the end of grade 3, and some students who had been focusing on length or area in the third grade. In addition, some of the students who had only been focusing on length at the end of grade three showed "improvement" by measuring area. (The literature documents that understanding the measurement of length, area, and volume is a stage-like process (e.g., Clements & Samara, 2009), so that measuring area rather than length, even though one should measure volume, is evidence of progress.)

Students also improved dramatically in realizing that the size of an object, not its weight, predicts water displacement (58% of students, as opposed to 7% in the pretest and 18% at the end of grade 3). The task could be solved successfully by simply comparing the heights of the cylinders, which had the same cross section, but it did call for reconceptualizing that spatial extent, not weight, is relevant to water displacement   Of course, students did explicitly address this topic in the curriculum, but it is also a notoriously difficult topic for students to learn, so we take this progress to be encouraging and a sign that the curriculum was effective in fostering the development of a concept of volume as "occupied space" stable enough to integrated into a new account of water displacement. This is particularly striking because students' initial account based on weight, was held by over 80% of third graders, and offered with great assurance. Overall, by the end of grade 4, 30% of the students had a fairly "consolidated" understanding of volume (being correct on both the volume measurement and water displacement task), while approximately two-thirds (69%) had moved to having strong insights on at least one of these tasks.

Students also continued to improve in developing precursors to a formal density concept. When asked which of two objects was made of a heavier kind of material, half the students were able to systematically relate weight and size, rather than base (at least some of) their answers on weight only. In addition, and quite impressively, half the students were able to use a sophisticated stacking strategy in order to correctly infer what material two small covered cylinders were made of. (That is, they stacked three of the same small covered cylinders together to make them the same size as the uncovered cylinders of known materials, and then compared the weight of the stack to each of the uncovered cylinders their strategy for inferring material.) Recall that, in third grade, most had directly compared the weight of one small covered cylinder to the weight of the uncovered cylinders, ignoring the difference in size.

We take progress on both these tasks as strong signs that students are developing meaningful precursors of density. Further, we note that although the curriculum discussed "heavy for size", students were never asked to use weight information to infer what a mystery material was; so this kind of task represents a novel extension of their class work. Overall, 44% of the students did distinguish heaviness of kind of material from weight in the judgment task *and* used the stacking strategy in the other; two-thirds of the students showed insight in at least one of these two demanding tasks.

Although students made no further progress in understanding that tiny visible objects take up space and have weight, they did maintain the insights they had developed in third grade. Further, they continued to make progress in their judgments about whether invisible pieces of stuff would take up space and have weight. In answering these questions, students can: (a) deny that an invisible piece would take up space and have weight, or simply deny the existence of invisible pieces (so no further questions were asked); (b) acknowledge the existence of invisible pieces, and assert they would take up space *or* have weight, but not both; or (c) acknowledge the existence of invisible pieces and maintain that they would take up space *and* have weight. Almost half of the students judged that the invisible pieces can exist, and that they take up space and have weight; in addition, another 27% of the students acknowledge that the invisible pieces would either take up space or have weight (overwhelmingly, when they acknowledge only one, it is taking up space). This is in contrast to what we found at the end of third grade where almost half still denied that invisible would take up space and weight.

Their insights on these questions about invisible pieces of stuff are particularly interesting, because again, the 3rd and 4th grade curricula never explicitly broach questions about the existence of invisible pieces of stuff or the properties of those pieces (this is part of the 5th grade curriculum). Thus, we see the progress students have made in grades 3 and 4 as "preparing the ground" for the work ahead in grade 5 where they consider the weight of invisible pieces of stuff in multiple contexts (e.g., dissolved salt, gases) and the transformations of matter in the water cycle. Traditional curricula typically introduce students to the topic of the water cycle and the three states of matter without the extensive conceptual preparation about weight, material, volume, and their relations that the Inquiry curriculum provides.

Significant restructuring has taken place in the knowledge network of the majority of the Treatment students by the end of 4th grade. Weight, was initially measured by hefting conceived as a property of both objects and materials (i.e., "heavy" had the same meaning in "iron is heavy" and "this big cylinder is heavy"), and not tied to matter (some materials, e.g., Styrofoam, weighs nothing; small pieces of stuff weigh nothing). It is now a property of objects, or rather of the amount of material an object is made of (as demonstrated by students' judgments that weight does not change when a clay ball is turned into a pancake), measured with a scale, and an inherent property of any piece of solid or liquid material, however small (as when they argue "everything has weight and takes up space"). Weight and size are both taken into account when thinking about the weight of objects made of different materials—objects made of brass are heavier for their size than objects make of aluminum. A concept of volume as "occupied 3D space" is now part of the network.

In keeping with the learning progression approach, we believe attention to restructuring students' knowledge network prior to introducing phase changes, the idea that matter includes gases, and a formal treatment of density is critical if these more advanced ideas are to be meaningfully integrated into students' knowledge network and hence to move it forward toward the atomic molecular theory without destabilizing it. The data we gather with these same students in fifth grade will be a critical further test of this assumption. We can see why longitudinal studies, though hard to do, are so important. If the Inquiry Curriculum is indeed "preparing the ground" in robust ways, students will be able to integrate the idea that gases as matter into their matter knowledge network—a notoriously counter-intuitive and challenging idea for students.

Of course, some may question whether this development is entirely a good thing, as they may see our curriculum as promoting the conflation of weight and mass—something that must be "undone" at a later point in a physics curriculum. We acknowledge that distinguishing weight and mass will be important, but propose that believing that weight is an inherent property of matter is an important "steppingstone" in elementary school for further learning about matter and density; that the scientific concept of mass is not within the reach of elementary school children, and will be much easier to grasp after learning the atomic molecular theory; and that elementary school children do distinguish between amount of stuff and weight, a distinction that can be elaborated on later, with proper instruction into the relation between weight and mass. From a LP perspective, what is important is moving the network forward in ways that prepare one to take the next step, even if that means delaying the time that one introduces beliefs that match the experts.

In fact, we already have some evidence that developing an understanding that tiny visible things take up space and have weight was strongly linked to children's beginning to differentiate weight and density. Those children who judged that small tiny things have weight and take up space were the ones who were most likely to consistently use prior relational information (about weight differences among equal size cylinders made of different materials) to make correct inferences about the heaviness of kind of material of new object pairs made of those same materials (see Table 3). The new item pairs were critical tests that students differentiated weight and heaviness of material because direct perceptual comparison of the weight of the new pairs would lead to incorrect judgments. This finding strongly suggests that their idea that tiny pieces have weight because everything has weight was not just some rote learned slogan, but a meaningful and productive belief. It is also consistent with the prior work of Smith and her colleagues with middle schoolers (1997, 2007) in which they have found similar relationships.

Comparisons with the Control Students

The fact that children in the Treatment group were improving (and in ways that matched targeted goals of the Inquiry curriculum) strongly suggests that they were benefiting from the Inquiry Project curriculum. However, it could also be that the existing science curriculum (not designed from a learning progression perspective, but with separate units on balance scales, materials, solids, liquids, and gases, and water cycle) is also effective in promoting the development of understanding matter at the macroscopic level or that some aspects of student improvement primarily stem from increasing maturation or from learning outside their science curriculum (e.g., increased mathematical facility; learning about volume in the math curriculum), in which case current work on LP has no "value added."

To test the LP hypothesis we need to compare the progress students make with LP curriculum to progress they make with a more standard approach. More specifically, how much progress did students in the same school make with their standard science curriculum? As part of our comparative longitudinal study, we interviewed students who were a year ahead of our treatment students, and who had had their schools' standard science curriculum.

Progress within the Control Students from Grade 3 to Grade 4 (Table 4)

When we interviewed the control students at the end of grade 3, we found many of their conceptions remarkably similar to those of our treatment students at the time of the pre-interview (see Table 1). This provides further evidence that many key ideas in elementary children's matter network are in great need of reconceptualization if children are going to have an appropriate foundation for learning about the atomic molecular theory in middle school.

Further, on the key tasks calling for reconceptualization (e.g., understanding weight as a fundamental property of material, constructing a concept of volume as occupied 3D space and differentiating contexts where it is relevant rather than area or weight, differentiating weight and density), their developmental trajectories so far have been relatively flat, with no significant changes from end of grade 3 to grade 4 on any of these key measures (see Table 4). This finding is consistent with the assumption that standard science curricula are not highly effective in fostering reconceptualizations. Whether these Control students will make more progress in Grade 5 or "catch up" later remains to be seen. We suspect there is value in developing foundational ideas about weight, volume, and material, as early as possible as a stepping stone for further learning; we also think the kinds of understandings we are targeting cannot be transmitted quickly through simply telling, nor acquired from hands on experiences that are not designed from an LP perspective. They involve the coordination and reorganization of both implicit and explicit understandings that require a carefully orchestrated sequence of new experiences and careful discussion of those experiences. The reorganization of the knowledge network about matter is long drawn, as evidenced by the gradual progress among students even in the Treatment group.

Significantly, we found some of the same patterns of answers across tasks in the two groups. These similarities are also of importance for LP work in suggesting important "inherent constraints" among concepts that any curriculum needs to be responsive to. For example, for both students in the Treatment and Control group, it was easier to infer that tiny pieces too up space than that they had weight, although student beliefs about space and weight were highly correlated. Similarly, students who judged that the tiny piece had weight felt compelled to give theoretical justifications for this judgment which were similar in both groups (if it exists, it must weigh something; everything takes up space and has weight). Further, in both groups, the majority of those who believed that tiny pieces had weight or took up space, also continued to infer that invisible pieces would too; they also were more likely to have developed some relevant insights differentiating weight and density (see Table 3).

All these findings point to the importance of engaging children with the question of whether tiny *visible* things have weight or take up space. Existing curricula might consider this too obvious and trivial to take time to discuss and move directly to more exotic topics about gases. We would argue, instead, that this is a key stepping stone that should not be overlooked. It presents many of the same epistemological challenges of the more exotic cases, but in more accessible and tractable form.

Direct Comparisons with Treatment Students (Tables 1, 5)

So far, we have compared the progress treatment students made at two intervals (from early to end of grade 3, from end of grade 3 to end of grade 4) and the control students made at one (from end of grade 3 to end of grade 4). For both intervals, the treatment students made significant progress on multiple items; for the one interval assessed so far for the control students, the control students did not.

So how did the understandings of the Control students at the end of Grade 4 compare with those of the Treatment students?

Table 5 shows that across multiple measures (beliefs about whether tiny things have weight or take up space, measuring volume, knowing volume, not weight, is the relevant variable in water displacement, using ideas about heaviness of materials in spontaneous explanations, developing some intuitions about the similarities among diverse forms of matter, more tightly inter-relating weight and taking up space with amount of matter) there were significant differences among the two groups. Although the differences between the two groups did not yet reach significance on two of the four density measures, they were in the right direction for both tasks. Further, if one considered the number of students showing a systematically correct pattern on at least one of these two tasks, the difference was clearly significant (68% in the Treatment group vs. 44% in the Control group.)

Of course, making such direct comparisons is tricky, as it depends upon the comparability of the students in the samples to begin with, and our design did not allow us to interview the control students at the beginning of grade 3. For this reason, we think within child measures of progress are to be preferred, as each child is then his or her own control.

Is there any reason to believe that students in the Control group made less progress because they were "less good" students or that they were starting further behind the Treatment group of students? In fact, the evidence suggests the reverse.

To evaluate this assumption that they might be "less good students", we gathered grade 3 MCAS Reading and Math raw scores for the two groups of students, and compared their "total" scores. We found no evidence for this assumption; at one school the Control group's "total" score was statistically superior to the Treatment group. At the other school, the two groups were quite comparable and included a diverse range of student abilities.

Discussion

These initial results are encouraging and support two basic assumptions of LPM: (a) that children have matter networks in need of major reconceptualization; and (b) that LPM designed instruction allows children to make progress on these concepts in ways that traditional science instruction does not. Of course, these developments take time and revisiting across years. If this pattern continues in the next year of our study, it will have implications not only for appropriate targets of elementary school instruction (i.e., weight, volume, material, matter and density), but for the value of an instructional approach that focuses on building concepts over time.

Our next steps are to continue to do fine-grained analyses exploring various hypotheses about the *relationships* among different understandings. For example, we already mentioned the hypothesis that understanding even tiny things have weight and take up space is a stepping stone for differentiating weight and density and coming to believe that all matter has weight and takes up space. That is, the children who clearly differentiate weight and density should be the ones who understand that even little tiny things take up space and have weight. We are also very interested in exploring the *relations* among children's reasoning about number and physical quantity in the emergence of children's understanding physical quantities as dense linear dimensions (an aspect of our study not focused on here). These kinds of analysis are central to testing the crucial idea of stepping stones embodied within LP work. Further, these analyses could reveal whether the *same patterns of relations* hold among Treatment and Control students even if they differ in level of attainment. If so, it would support the idea that some elements of learning progressions reflect constraints among the concepts themselves and hence are fundamentally similar across different curricular environments.

We believe such analyses have and will also shed light on ways the curriculum may have fallen short or could continue to be improved. For example, one area for improvement would be strengthening student's understanding of and the measurement of volume.

One of the major advantages of comparative longitudinal studies is its ability to test hypotheses about the causal role of both instructional experiences and stepping stones in LP in ways much more powerfully than cross-sectional data can. One limitation is, of course, the difficulty of carrying out such studies and their smaller sample sizes. Ultimately, a judicious combination of both approaches will be essential.

References

Carraher, D., Smith, C., Wiser, M., Schliemann, A., & Cayton-Hodges, G. (2009) Assessing students' evolving understandings about matter. Paper presented at LeaPS Project Conference, Iowa City, June, 2009.

Clements, D. H., & Sarama, J. (2009). Learning and teaching early math: The learning trajectories approach. New York: Routledge.

Corcoran, T., Mosher, F., and Rogat , A. (2009*). Learning Progressions in Science: An Evidence-based Approach to Reform*. The Consortium for Policy Research in Education. Philadelphia, PA.

Lee, O., Eichinger, D. C., Anderson, C. W., Berkheimer, G. D., & Blakeslee, T. D. (1993). Changing middle school students' conceptions of matter and molecules. *Journal of Research in Science Teaching*, *30*(3), 249-270.

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

National Research Council (NRC). (2007). *Taking science to school: Learning and teaching science in grade K-8.* Committee on Science Learning, Kindergarten through eighth grade. Duschl, R.A., Schweingruber, H.A., & Shouse, A.W., Eds. Washington DC: The National Academy Press.

Smith, C. (2007) Bootstrapping processes in the development of students' commonsense matter theories: Using analogical mappings, thought experiments, and learning to measure to promote conceptual restructuring. *Cognition and Instruction, 25(4),* 337-398.

Smith, C., Maclin, D., Grosslight, L. and Davis, H. (1997) Teaching for understanding: A comparison of two approaches to teaching students about matter and density. *Cognition and Instruction, 15 (3),* 317-393.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and atomic-molecular theory. *Measurement: Interdisciplinary Research and Perspectives, 14*(1-2), 1-98.

Whitney, H. (1968a). The Mathematics of Physical Quantities: Part I: Mathematical models for measurement. *The American Mathematical Monthly, 75*(2), 115-138.

Whitney, H. (1968b). The Mathematics of Physical Quantities: Part II: Quantity Structures and Dimensional Analysis. *The American Mathematical Monthly, 75*(3), 227-256.

Wiser, M., Smith, C., & Doubler, S. (in press). Learning Progressions as Tool for Curriculum Development: Lessons from the Inquiry Project. To appear in A. Alonzo & A. Gotwals (Eds.), *Learning Progressions in Sciences*. Sense Publishing

Table 1: Overview of Student Initial Ideas for Grade 3 Pre-Treatment Group and Grade 3 Control Group

	G3 Pre-Treatment (Beginning of G3) N=45	**G3 Control (End of G3) (N=27)**		**Comparison of Distributions (Chi-Square)**
Properties of Tiny Pieces (Task 1)				
Visible speck has weight	9%	11%		N.S.
Visible speck takes up space	44%	52%		N.S.
Visible speck takes up space and has weight	7%	11%		N.S.
Volume Tasks				
Volume measure (Task 2)	7%	7%		N.S.
Water displacement (Task 4)	7%	0		N.S.
Materials and Density (Task 3)				
Use difference in heaviness of kinds of materials to explain why larger is not always heavier (Problems 2 or 3, Task 3, Part 1)	11%	15%		N.S.
Coordinate material and size in explaining why objects can be same weight but different size	0%	15%		N.S.
Distinguish heavier object from heavier material (across 3 items, Task 3, Part 2)	18%	27%		N.S.
Correctly infer identity of both covered materials via stacking (Problems 3&4, Task 3, Part 3)	8%	22%		N.S.
Matter Sorting (Task 7)				
Include all solid, liquid, and gas while excluding non-material items	2%	7%		N.S.
Other coherent patterns: Only under-extend; or only over-extend)	49%	49%		
Transformations: Amount, Weight, Volume, Material (Tasks 5, 9)				
Shape change does not change amount of clay, weight of clay, and whether items will balance	53%	74%		N.S.
Shape change goes not change amount of clay, weight of clay, balance and volume of clay	18%	8%		
Grinding/melting does not change material identity (all 3 items)	64%	81%		N.S.

Table 2: Overview of Grade 3 to Grade 4 Findings (Treatment Group)

	Treatment Group				Significant Progress in Patterns (Wilcoxon, p-value)		
	G3 Pre	G3 Post	G4 Post		G3 Pre to G3 Post	G3 Post to G4 Post	G3 Pre to G4 Post
Properties of Tiny and Invisible Pieces*							
Visible speck takes up space and has weight	9%	60%	62%		P < .001	N.S.	P < .001
Invisible pieces take up space and have weight	7%	42%	47%		P < .01	P = .017	P < .001
Volume Tasks							
Volume measure*	7%	27%	40%		P < .001	P = .014	P < .001
Water displacement	7%	18%	58%		N.S.	P < .001	P < .001
Materials, Weight and Density							
Use difference in heaviness of kinds of materials to explain why larger is not always heavier	11%	49%	58%		P < .001	N.S.	P < .001
Coordinate material and size in explaining how two different size objects are same weight	0	20%	38%		P < .01	N.S.	P < .001
Distinguish heavier object from heavier material across three items*	18%	33%	49%		P < .05	P < .01	P < .001
Correctly Infer material via stacking for two items	8%	9%	51%		N.S.	P < .001	P < .001
Matter Sorting*							
Include only solid, liquids, and gases	2%	4%	2%		P < .001	N.S.	P < .001
Other systemic patterns (only under-extend or over-extension errors for a given child)	49%	78%	80%				
Transformations*							
Shape change does not change amount, weight, and balance	53%	84%	91%		P < .001	N.S.	P < .001
Shape change does not change amount, weight, balance and volume	18%	38%	53%				

*Change on these tasks was assessed through an ordered set of categories (see text for description); this table shows just percent achieving the highest category (or highest two categories).

Table 3: Relation Between Judging that Tiny Things Take up Space and Have Weight and Consistently Using Relational Reasoning in Inferring Heaviness of Kind of Material for Treatment and Control Students

Pattern on (Visible) Tiny Piece of Clay	**Grade 4 Treatment:** **% Consistently Using Relational Reasoning in Inferring Heaviness of Kind of Material**	**Grade 4 Control:** **% Consistently Using Relational Reasoning in Inferring Heaviness of Kind of Material**
Does not take up space or have weight	[1/8] 13%	[1/12] 8%
Either takes up space OR has weight, but not both	[3/9] 33%	[3/7] 43%
Both takes up space and has weight	[13/28] 64%	[5/8] 63%

Table 4: Overview of Grade 3 to Grade 4 Findings (Control Group)

Aspect of Understanding Probed	**Control Group**			**Significant Progress in Patterns (Wilcoxon, p-value)**
Properties of Tiny and Invisible Pieces*	**End of Grade 3**	**End of Grade 4**		**End of Grade 3 to End of Grade 4**
Visible piece both weight and space	11%	30%		N.S.
Invisible piece has wt and space	7%	19%		N.S.
Volume				
Volume measure*	7%	11%		N.S.
Water displacement	0	15%		N.S.
Materials and Density				
Use difference in heaviness of kinds of materials to explain why larger is not always heavier	15%	30%		N.S.
Coordinate material and size in explaining how two different size objects are same wt	15%	15%		N.S.
Distinguish heavier object from heavier material* (all 3 items, Part 2)	27%	33%		N.S.
Infer material via stacking	22%	33%		N.S.
Matter Sorting*				
Include only solid, liquids, and gases	7%	7%		N.S.
Other systematic patterns (only over-extension or under-extension errors)	49%	41%		
Transformations*				
Shape change does not change amount, weight, and balance	74%	78%		N.S.
Shape change does not change amount, weight, balance, and volume	8%	6%		

*Change on these tasks was assessed through an ordered set of categories (see text for description).

Table 5: Comparisons of Treatment and Control (Grade 4 Findings)

Aspect of Understanding Probed	**Treatment**	**Control**		**Treatment vs. Control**
Properties of Tiny and Invisible Pieces	**End of Grade 4**	**End of Grade 4**		**Significant Difference in Distributions (Chi Square, p-value)**
Visible piece takes up space & has weight	62%	30%		P = .017
IP exists and has wt and space	47%	19%		P = .014
Volume				
Volume measure	40%	11%		P = .018
Water displacement	58%	15%		P < .001
Materials, Weight and Density				
Use difference in heaviness of kinds of materials to explain why larger is not always heavier	58%	30%		P = .028
Coordinate material and size in explaining how two different size objects are same wt	38%	15%		P = .044
Distinguish heavier object from heavier material in judgments (all 3 items, Part 2)	49%	33%		N.S.
Infer material via stacking for two problems	51%	33%		N.S.
Matter Sorting				
Include only solid, liquids, and gases	2%	7%		P < .01
Other systematic patterns (only over-extension or under-extension errors)	80%	41%		
Transformations				
Shape change does not change amount of clay in ball, weight of ball, or whether item will balance	91%	78%		P < .01
Shape change does not change amount of clay in ball, its weight and volume, and whether it will balance	53%	6%		