Running head: Conceptual Bootstrapping

Bootstrapping Processes in the Development of Students' Commonsense Matter Theories: Using Analogical Mappings, Thought Experiments, and Learning to Measure to Promote Conceptual Restructuring

Carol L. Smith

University of Massachusetts at Boston

Cognition and Instruction (in press)

(Accepted for publication: April, 2007)

(Scheduled to appear in Fall 2007 double issue)

Abstract

This study explores whether the development of students' understanding of matter as something that occupies space and has weight involves conceptual change and restructuring rather than only simple belief revision. Based on an analysis of how the concepts in students' initial matter theory (henceforth MT1) may differ from the concepts in the matter theory that is a target of middle school instruction (MT2), I propose ways concepts in a given theory cohere with each other and identify the sources of the new ideas in MT2 and the learning processes by which those new ideas can be acquired. I test implications of these analyses by designing a curriculum unit that exploits these learning mechanisms, by using the curriculum with 4 classes of 8th grade Earth Science students, and by assessing 42 students' thinking about matter, object size, and weight (via individual interviews and written tests) before and after the teaching unit. Consistent with the hypothesis of conceptual restructuring, the data not only show coherencies in students' thinking about matter, size, and weight before and after teaching, but also coordinated patterns of change. The implications for the design of science education are discussed.

Bootstrapping Processes in the Development of Students' Commonsense Matter Theories: Using Analogical Mappings, Thought Experiments, and Learning to Measure to Promote Conceptual Restructuring

This study examines whether constructing an understanding of matter as something that takes up space and has weight involves a restructuring of a network of concepts and their meanings--a form of deep conceptual change that involves locally incommensurable theories--and if so how such restructuring might occur. Although science educators have long been aware of the immense conceptual challenges posed by the atomic-molecular theory of matter (e.g., challenges to basic assumptions that matter is continuous and inert), many have not assumed that there are major conceptual challenges to understanding matter at a macroscopic scale. Consequently, these core ideas about matter are often either assumed to be in place or presented to students explicitly as unproblematic first principles. In neither case do educators anticipate that students will have resistance to understanding these principles, or plan coherent sets of activities designed to promote understanding of them. This omission may be a mistake, as evidence is now accumulating that many students have alternative (macroscopic) concepts of matter, weight, volume, and material kind that persist into the middle school years (Smith, Maclin, Grosslight, & Davis, 1997; Smith, Snir, & Grosslight, 1992; Stavy, 1995) that significantly affect their learning of the atomic-molecular theory of matter itself (Lee, Eichinger, Anderson, Berkheimer, & Blakeslee, 1993; Snir, Smith, & Raz, 2003). Thus, if educators more clearly understood the extent to which understanding these ideas involves changes in the core meanings and relations among concepts rather than simple belief revision, they might better appreciate why the kinds of learning processes needed to facilitate these understandings are more complex than simple knowledge accretion processes. This knowledge in turn might encourage them to design more effective curricula that would help students build a sound understanding of matter during the elementary school years, one that would lay the foundation for later learning of the atomic-molecular theory during the middle school years (Smith, Wiser, Anderson, Krajcik, 2006).

Conceptual restructuring accounts of conceptual change rest on the assumption that students have some initial commonsense theories in which their everyday explanatory concepts are embedded and play a role (Carey, 1999; Gopnik & Meltzoff, 1997; Wellman & Gelman, 1992). These commonsense theories, although not self-consciously held, are assumed to be like scientific theories in consisting of inter-related concepts that resist change, that determine a concept's core (i.e., its central or essential properties), and that support inference making, problem solving, belief formation and explanation in a given domain. Conceptual restructuring occurs in cases of theory development where some of the concepts of the later theory (T2) are locally incommensurable with those of the earlier theory (T1) in the sense that some of the beliefs of one theory cannot be formulated in terms of the concepts of the other (Carey, 1991). This failure in translation occurs when there have been coordinated changes in how concepts are individuated in the theory (via coordinated conceptual additions, deletions, differentiations, and coalescences) as well as fundamental changes in the meaning of the concepts (e.g., changes in the core properties of a concept or a re-analysis of the concept's basic structure). In such cases, the concepts and beliefs of one theory are not intelligible in terms of the other. That is, students cannot literally represent the new ideas in terms of the concepts of their initial theory.

Of course, not all cases of knowledge acquisition or theory development call for conceptual restructuring. Carey (1999) contends, "Theories are often merely enriched as new knowledge accumulates about the phenomena in the domain of the theory. Theory enrichment consists of the acquisition of new beliefs formulated over a constant conceptual repertoire" (p. 295). In contrast, the learning mechanisms underlying conceptual change must be more complex than those underlying simple belief revision. As Carey (1999) has also argued, many standard learning mechanisms (such as, hypothesis testing, parameter setting, correlation detection, association) work over an existing conceptual base; hence they aren't sufficient to explain the learning that occurs in conceptual restructuring. How could one learn a theory that one cannot initially represent? Other learning mechanisms (such as analogical mapping, limiting case analysis, and inference to best explanation) may play an especially important role in these cases, because they help create the links between new networks of inter-related symbols and prior mental representations. Further, learning underlying conceptual restructuring is multi-faceted: It involves orchestrating a coordinated series of mutually reinforcing changes, with no one change by itself sufficient to create the new sustainable structure. For this reason, philosophers of science (e.g., Quine, 1960, Nersessian, 1989, 1992) have appealed to a "bootstrapping" metaphor to describe the sustained, dialectical learning processes that occur in conceptual change of this type.

Some of the clearest cases of major conceptual restructuring--changes in the meanings of large networks of concepts used to explain phenomena in a domain--have come in the history of science. (See Gentner et al., 1997, Kitcher, 1988, Nersessian, 1989, Thagard, 1992, Wiser & Carey, 1983 for worked examples.) An important question concerns whether children also (sometimes) go through such restructuring in the course of their knowledge acquisition and if so, how this restructuring might come about. In a seminal series of papers, Carey argued that they do and provided detailed sketches of two cases where such changes might occur. One case involved children's concepts of *animal, plant, alive, person, death, growth, baby, eat, breathe,* and *sleep* (Carey, 1988); the other their concepts of *matter, material kind, weight, volume,* and *density* (Carey, 1991).

Others have doubted that they do or that conceptual restructuring is in the best characterization of the changes that occur in students. On the one hand, some who believe children have intuitive theories doubt that children have the metaconceptual sophistication to undergo conceptual restructuring (e.g., Spelke, 1991). For them, conceptual restructuring is a form of conceptual change that only occurs (rarely) in mature scientists. Others have challenged the characterization of children and students' as having initial commonsense theories (diSessa & Sherin, 1998). They argue that everyday explanation is embedded in knowledge systems that are too loosely organized, fragmented and context dependent to be well characterized as intuitive theories. Thus, they feel that conceptual change is better characterized as a process of *organizing* these loosely organized elements than moving from one committed explanatory system to another. Still others acknowledge that such restructuring might sometimes occur, but are not yet convinced that they occur in these particular cases described by Carey, especially the case in biology, arguing that they may be cases of mild reorganization rather than full scale restructuring (Chi, 1992; Keil, 1999; Thagard, 1992).

At present, detailed examination of the issue has been limited, especially in the context of students' changing macroscopic understandings of matter, occupied space, weight, and density. Although there have been numerous studies documenting specific misconceptions, there have been fewer studies that test specific conceptual analyses of how these ideas might cohere and that consider the processes that might facilitate conceptual restructuring. In the next two sections, I review and extend prior conceptual analyses and show how these lead to specific predictions examined in the present study. These predictions concern how students' ways of articulating one concept should relate to their ways of articulating other (related) concepts and the kinds of resistance to change and coordinated patterns of change that should occur in response to teaching.

Two Contrasting Macroscopic Matter Theories[1]

In previous work, it has been proposed that children's concepts of matter, material kind, size, and weight are embedded in an initial commonsense matter theory (MT1) that is incommensurable with the scientists' matter theory taught in middle school (MT2) (Carey, 1991; Smith et al., 1997). In this section, I first present an extended analysis of the contrasting concepts in MT1 and MT2 (see Table 1) and the sense in which these concepts are incommensurable. These characterizations portray the contrasting concepts of an idealized (and simplified) starting and endpoint, not the many points of transition along the way. I assume (based on work both in history of science and science education) that theory development is a complex and gradual process that has both evolutionary and revolutionary aspects, and will later discuss some "transitional positions" that can occur as students move from one set of understandings to another. As Nersessian (1989) explains:

> In all major conceptual changes in science, whole complexes of concepts have changed. These changes are largely independent, and yet interconnected. They are independent in that emergence of a new concept or alternation of an existing one does not automatically lead scientists to see how to make the other changes that will eventuate in the new conceptual structure. At most the repercussions of change in one part of the conceptual network will spread throughout the network and will point to areas in need of revision. We see clearly what needed revision and why only in historical

> perspective….Local revisions, by themselves, do not force global changes. The instructional import of this is that in teaching a scientific conceptual structure, a number of concepts need to be targeted for revision at the same time and new concepts introduced in a coordinated fashion. Unlike the scientists who first constructed the conceptual framework, we can take advantage of hindsight and emphasize the relevant conceptual interconnections in instruction. (p. 177)

My conceptual analysis includes the new proposal that the physical quantities of MT1 are represented as analog magnitudes. In contrast, the physical quantities in MT2 are represented more precisely and symbolically using natural language or mathematical symbols (such as "twenty grams" or "2.5 cm^3"). In this way, the development of children's matter theories is connected with the extensive literature on the contrast between analog and symbolic representations of quantity (see Carey, 2001, 2004; Church & Broadbent, 1990; Dehaene, 1997; Gallistel & Gelman, 1992) and with recent work on the role of mathematical ideas and inscription in helping students to redescribe their everyday perceptual experiences (Lehrer, Strom, & Confrey, 2002). I then briefly summarize the prior evidence that supports these analyses and present the new predictions being tested in this study.

Analysis of the Concepts of Matter Theories 1 and 2

Table 1 presents my analysis of the main concepts in each matter theory. The concepts of the two theories are characterized as incommensurable for several reasons. First, there are changes in the core properties of the concepts of the two theories. Second, there are fundamental changes in how concepts are individuated in the two theories, with both conceptual differentiations and coalescences occurring as part of the process of theory change. Third, there are changes in what properties are explicitly represented in each theory, with MT2 containing important properties that are not even represented in MT1. Finally, there are changes in the core explanatory principles of each theory and the explanatory role that individual concepts play within the theory. Although I assume these changes are inter-connected and mutually reinforcing, for purposes of exposition, I will consider each in turn.

Insert Table 1 about here

Changes in Core Properties of Concepts

One hallmark of conceptual change is that properties that are essential to the meaning of concepts in the initial theory come to be seen as more peripheral to their meaning in the later theory (and often derivative from new core properties of concepts in the later theory). Table 1 describes change in what properties are core for children's concepts of matter, material kind, size, and weight from MT1 to MT2.

In MT1, matter is conceptualized as some objective stuff that is perceptually accessible--that is, something that can be seen, touched, tasted, smelled and felt. MT1 supports an ontological distinction between material entities in the physical world (that are perceptually accessible to others in multiple ways and give rise to correlated sensory experiences) and mental entities (that are private and not directly accessible to others) and licenses certain inferences. Although children are aware that objects are made from different materials, they think of these materials in terms of their characteristic perceptual properties rather than in terms of more underlying properties (such as density). Further, although children are aware that many material objects have some size and weight, they do not yet think that *all* of them do--hence occupying space and having weight are not seen as essential properties of matter.

In contrast, in MT2, matter is conceptualized more abstractly as a fundamental constituent of objects or aggregates that takes up space, has weight and is conserved across repeated subdivision.[2] That is, matter continues to exist as it is divided into tiny pieces and each of these small pieces continues to take up space and have weight even though no longer detectable by human senses. Thus, the properties of taking up space, having weight, and being conserved across subdivision are now more central properties than being perceptually accessible.

Similarly, there are changes in the core of children's concepts of object size, weight, and material kind. More specifically, the core of object size shifts from *perceived size*--an unanalyzed perceptual property--to a more precise, three-dimensional analysis of the volume of space occupied by matter that can be quantified and measured. Similarly, the core of weight shifts from *felt weight*, an unanalyzed perceptual property of entities that can be held, to an analyzed, extensive, property of matter that can be measured using balance or spring scales. Of course, there is more to children's initial weight concept than felt weight. Felt weight is central, however, in that it grounds their understanding of what it *means* for things to have weight and provides children with a way of assessing the weight of things. At the same time it supports their learning crude generalizations, such as bigger things tend to be heavier, and adding more stuff tends to make things heavier, as often those weight differences are perceptually quite noticeable. What it does not support is their understanding that *all* matter has to have weight or adding *any* amount of material makes something heavier--especially since some things feel like they weigh nothing at all.

Finally, material kinds, such as wood, plastic, or metal, shift from being conceptualized as complexes of perceptual properties (i.e., color, texture, taste, smell, feel) to being conceptualized as fundamental constituents that have underlying properties that need to be inferred (e.g., density) or that may not be even fully known (e.g., some underlying essence, chemical structure or make-up). As students experience how the perceptual properties of materials can change across different transformations, they increasingly favor tracing the identity of materials by historical means rather than by perceptual properties alone. For example, if sawdust comes from sawing a piece of wood, they would conclude that it was still the same kind of material even if it was no longer the same color or texture as the chunk of wood. Further, coming to see weight as a fundamental property of matter goes hand-in-hand with differentiating weight and density and recognizing density as a distinguishing underlying characteristic of material kinds.

Changes in How Concepts Are Individuated in the Two Theories

Theories are incommensurable not only because there have been changes in the core properties of key concepts but also because there is not a simple one-to-one correspondence between some of the concepts in the two theories. One theory may use a single concept where the other theory finds it essential to recognize two fundamentally different kinds of entities (conceptual differentiation); or one theory may abandon a conceptual distinction that was seen as essential in the other theory (conceptual coalescence). Indeed making conceptual differentiations and coalescences goes hand-in-hand with making changes in a concept's core.

Table 1 identifies some of the important conceptual differentiations and coalescences that are part of the move from MT1 to MT2. These include differentiating weight and density as different *kinds* of physical magnitudes (as extensive vs. intensive quantities, respectively); differentiating length, area, and volume as different *kinds* of spatial extents (based on a mathematical reanalysis of space as structured arrays of units extending in one, two, and three dimensions); and differentiating air and empty space while coalescing solids, liquids, and gases as all phases of matter. Note that differentiating mass from weight is not part of this restructuring, as it would depend upon learning enough Newtonian mechanics to support such a differentiation. In Table 1, the term "mass/weight" is used to refer to children's weight concept in MT2 because it conflates elements of the physicist's concepts of mass and weight.

Changes in What Properties Are Explicitly Represented in the Two Theories

This repackaging of conceptual units also involves changes in what properties are included in the conceptual analysis, with some new properties added and others devalued or eliminated. In MT1 the perceptual properties "felt weight" and "perceived size" are central to children's concepts of weight and size. Further, children represent and reason about these physical quantities using perceptually-based, analog magnitude representations rather than more precise symbolic ones. Analog magnitude representations are approximate, fuzzy representations of quantity that are subject to Weber's law. In Weber's law, the size of a just noticeable difference increases with the magnitude of the stimulus. Thus, the ability to distinguish two stimulus magnitudes is a function of their ratio rather than their difference, and the same difference may be perceived as having different magnitudes in different contexts. In contrast, in MT2, children also have more precise, symbolic representations of these physical magnitudes. They conceptualize weight and size as fundamental properties of matter and have mapped these quantities to number via knowledge of relevant measurement procedures.

In MT1, children know that objects are "made of" different kinds of materials that can maintain their identity when cut into smaller pieces. However, analog magnitude representations of physical quantities take as inputs perceptually accessible entities; hence they cannot be used to represent and reason about perceptually undetectable physical amounts. Thus, children must draw upon other representational resources to construct the notion of matter as a "fundamental constituent" that characterizes MT2-- represented symbolically and recursively in terms of an imagined division of an arbitrarily small piece.

In addition, the perceptually-based, analog magnitude representations of weight and size of objects do not support: (a) students' making a decompositional analysis of these quantities, (b) their formulating the distinction between weight and density, or (c) their analyzing size in terms of different dimensions. Perceptual procedures produce global (unanalyzed) judgments of extent that conflate different senses of size and weight, are notoriously inexact and subject to illusions, and do not provide direct access to the characteristics of the "microscopic" world. For example, children assess how heavy an object is by picking it up and feeling the pressure and force of the object in their hands, with some objects having no felt weight at all. Judgments of felt weight are inherently influenced both by the weight of an object and the density of the material it is composed: a small, dense object feels heavier than it really is. Thus, the notions of "heavy" and "heavy for size" are conflated in one undifferentiated weight concept, and the child uses the same undifferentiated weight concept in generalizations about balance scales, materials, and sinking and floating.

Similarly, judgments of the size of objects are based on judgments of global bigness rather than an analysis of different spatial dimensions and can only be made for objects whose spatial extent is considered perceptually significant. However, given the greater sensitivity of the human visual sense than sense of touch (we can see things that exert no detectable pressure on our hands) and given the existence of common technological devices such as magnifying glasses and microscopes that directly increase the sensitivity of vision but not the sensitivity of the skin senses, children can more easily detect that something is still "there" or continues to exist through vision than touch. Thus, one would expect that students would first see occupying space as an essential property of matter (rather than having weight). Further, they would begin to make the construal of the continued existence of matter at a micro-level, based in part upon their knowledge of magnifying glasses and microscopes, before they construct the idea that all matter must have weight.

Changes in the Explanatory Agenda of the Two Theories

In MT1, children's notion of "matter" as "physical stuff" helps children draw the distinction between mental and physical realms. It helps them explain why objects produce correlated sense experiences (because there is some objective stuff in the world that simultaneously excites the different senses) as well as why different individuals can perceive the same objects. It also provides an initial framework for organizing their learning about specific kinds of materials. Because children rely on analog magnitude representations for physical quantities, however, they do not have clearly defined notions of "unit of matter" or "unit of weight"; thus, explaining the total weight of an object in terms of the weight of (imagined) component units is not part of their explanatory agenda. In contrast, the changes in concepts of matter, size, and weight in MT2 also change the explanatory power and agenda of their theory. For example, coming to see weight as a fundamental property of matter and matter as "underlying constituent" commits students to explaining the weight of a large-scale object in terms of the weight of its component parts. Such an analysis in turn requires that they make a principled distinction between weight and density: density is an intensive property of material kinds that is preserved in the decomposition, while weight is an extensive property of matter that reflects the sum of the weight of all the component parts.

Prior Evidence Supporting This Conceptual Analysis

Evidence for Changes in Core Properties of Concepts

One line of evidence that being perceptually accessible is central to children's initial concept of matter comes from an analysis of the set of entities children pick out as matter and the way they justify their choices (Carey, 1991; Smith et al., 1997; Stavy, 1995). The things that are most often judged as matter are things for which they have multiple inter-connected sensory cues--things that they can not only see, but also touch and feel--for example, a car, rock, sand, water. As the number of perceptual cues becomes fewer, there is a fall-off in children's judgments that the entity is indeed matter, with mental events such as wishes and dreams being clearly recognized as immaterial entities.

Individual pattern analyses conducted by Carey (1991) in her study of preschool to grade 6 children further revealed that the set of entities each child picked out as matter does not coincide with the scientist's extension. Children typically both underextended (i.e., failed to pick out some entities that scientists consider matter such as air, smoke, or dissolved sugar) and overextended (i.e., included some entities that scientists would exclude--such as heat, light or electricity). Further, children's justifications for their judgments revealed that what held this unusual extension together was that these entities are physically detectable or exert important physical effects--not that they take up space or have mass. Only a minority of older children uniquely picked out the set of entities that have mass as being matter (i.e., including solids, liquids, gases, and powders, while excluding heat, light) or justified their choices on this basis. The latter patterns give strong evidence that there has been a shift in the core properties of matter from being perceptually accessible to having weight or mass.

A quite different line of evidence for this same conclusion comes from an analysis of children's pattern of judgments to thought experiments about matter (Carey, 1991; Smith, Solomon, & Carey, 2005; Yair & Yair, 2004). For example, in one thought experiment children were asked what happens as a piece of Styrofoam is repeatedly divided. If being perceptually accessible is central to their concept of matter, then one would expect that they would assert that, as matter is repeatedly divided, it gets smaller and smaller and ultimately disappears. In fact, this is precisely how many young children answered this question. In contrast, older children often answered quite differently--confidently asserting that matter could not simply disappear; even if it were no longer perceptible, it must still really be there. These judgments and justifications suggest that being an underlying constituent that is conserved in decomposition has replaced being perceptually accessible as a core characteristic of matter.

Evidence for changes in the core properties of children's concepts of weight, size, and material kind also come from changes in their pattern of judgments and justifications as well as responses to thought experiments. For example, if felt weight is central to children's concept of weight, then children should judge that some light visible material objects, placed in their hands, weigh nothing at all. In contrast, if children believe that weight is a fundamental property of all matter, they should confidently judge that these entities must weigh something (even if not perceptually detectable) and maintain in thought experiments that even pieces of matter too small to see and feel continue to have some weight. In fact numerous studies have consistently found such shifts in patterns of judgments and justification about weight (Bovet, Domahidy-Dami, & Sinclair, 1982; Carey, 1991; Smith, Carey, & Wiser, 1985; Smith et al., 1997, 2005). Most young children confidently deny that some visible objects have any weight at all because they "feel like nothing" whereas many older children equally confidently maintain that they must weigh something because they are matter and matter takes up space and has weight. These same studies have reported similar shifts in judgments about whether something takes up any space, although the shift typically occurs a little earlier. Finally, a few studies have investigated children's judgments about whether material kinds maintain their identity when ground into tiny pieces. Young children often deny that material identity has been maintained when ground into fine powders that no longer have these typical properties (Dickinson, 1987) or when substances are dissolved (Au, 1994).[3] In contrast, older children are more likely to assert that a ground up powder must still be the same material because the original object was *made of* that material (Smith et al., 1985). These findings are consistent with a shift from reliance on perceptual properties to identify kinds, to greater reliance on historical reasoning and knowledge of transformational history.

Evidence for Conceptual Differentiations and Coalescences

In their pioneering studies of children's understanding of physical quantities and their conceptualizations of space, Piaget and Inhelder have provided extensive evidence for the claims that children initially conflate weight and density (1974) and have topological rather than Euclidian conceptions of space (1967). Multiple follow-up studies have also supported the assumption that they typically conflate weight and density across a wide range of contexts. These contexts include making direct verbal judgments about the weight of objects and the density of materials (Smith et al., 1985, 1992, 1997), reasoning about which objects will make a foam bridge collapse (Smith et al., 1985), which objects could be made of the same material (Smith et al., 1985, 1997), and which objects will sink or float (Smith et al., 1992). Although some children come to differentiate weight and density during the elementary school years (Piaget & Inhelder, 1974; Smith et al., 1985), many continue to conflate weight and density into the middle school, high school, and adult years (Duckworth, 1986; Gennaro, 1966; Hewson, 1986; Smith et al., 1992, 1997). Children also often inappropriately use units of length as measures of other spatial extents, such as area, volume, and angle (Lehrer, 2003; Smith et al., 1997), and confuse measures of surface area with measures of volume (Smith et al., 1997).

Evidence for Conceptual Coherence

Conceptual coherence refers to how well concepts fit together when applied to situations in ways that make sense and avoid contradiction. Any argument for conceptual coherence requires a detailed analysis of how the concepts are articulated in a given conceptual system that supports their working together across their domain of application, such as was given above. Clearly, the concepts of weight, volume (as occupied space), and matter, which are articulated in MT2 as measurable physical quantities, coherently fit together not only because they are explicitly inter-related and inter-defined and rest on decompositional models of matter and its properties, but also because solids, powders, liquids, gels, and gases *do* in fact share those two properties. Yet, the concept of matter, which is articulated in MT1 as something that can be seen, felt, and touched, provides an alternative basis for identifying and reasoning (reasonably) coherently about physical stuff, especially as most entities children encounter and recognize as material in their everyday lives *do* produce these correlated sets of perceptual experiences. Further, given an understanding of matter as perceptually accessible stuff, it makes sense for children to privilege commonsense perceptual properties (such as felt weight, perceived size, color or texture) in their concepts of weight, size, and material kind as well. In this sense, the concepts of matter, weight, and size as articulated in MT1 constrain each other and fit together, even though they are not explicitly interdefined.

Given such an analysis, one can then seek evidence that students who articulate a concept one way are more likely to articulate related concepts in ways that promotes their fit with each other. One can also seek evidence that as students change their articulation of one concept they also change their articulation of others and that they resist piecemeal change. However, it is important to note that evidence for empirical associations among understandings and resistance to change would not by itself (in the absence of a conceptual analysis) support the claim of coherence. For example, if students who did not know the capital of Massachusetts also did not know the capital of Montana and resisted learning the names of both capital cities, one would not want to conclude there were intrinsic relations between knowledge of the names of capitals. One would seek explanations of the observed associations and resistance to change elsewhere.

Few studies have explicitly addressed the issue of coherence in children's concepts of matter, weight, density, and occupied space either theoretically (through conceptual analyses) or empirically (through testing for the kinds of interdependencies in ways of articulating these concepts expected based on these analyses). However, a few prior studies have reported some similarities in children's ways of responding to matter, space, and weight thought experiments (Smith et al., 1997; Smith, Solomon, & Carey, 2005); and that students who differentiated weight and density also conceptualized weight as a fundamental property of matter (Smith et al., 1985, 1997). Further, students who distinguished between weight and density for objects made of common solid materials were able to generalize this understanding to liquids and to understand sinking and floating in terms of the relation between the densities of materials (Smith et al., 1992; Snir et al., 1995).

Of course, conceptual coherence does not require the simultaneous emergence of all new ideas of MT2 or complete consistency across contexts, but it does imply that coordinated adjustments are needed to move from one relatively stable structure to another and that there should be cascading patterns of change. For example, Carey (1991) argued children begin to reconceptualize matter and occupied space as extensive quantities before they reconceptualize weight as an extensive property of matter, with this earlier change preparing the ground for the later change. Further, students may understand a new set of ideas in clearer contexts first and then use this new understanding to understand more challenging contexts (as when they understand the logic of length and area measurement prior to working through the logic of weight and volume measurement--see Lehrer, Jaslow, & Curtis, 2003). They may also work out some implications of new ideas before others, with new knowledge, experiences, tools, debate and dialogue playing a critical role in the process. An interesting feature of theory change accounts of conceptual change, though, is that they suggest that the need to preserve conceptual coherence will be an important part of the process of change in both children and adults (Chinn and Brewer, 1993; Thagard, 1992).

Current Study: Further Tests of Coherence Among Conceptions in Matter Theory 1

This study further investigates whether students' concepts of matter, object size, and weight cohere in alternative commonsense matter theories. I attempt to replicate previous findings by examining whether students who conceive of weight as felt-weight have alternative concepts of object size and matter that support their conception of weight as an emergent, rather than as a strictly additive property of matter. I also go beyond previous studies in two ways. First, I present children with multiple tasks probing for a concept of weight centered on felt weight, and hence examine whether students are consistent in their way of reasoning about weight across different tasks. Second, I examine whether students who have alternative conceptions of weight, object size and matter have failed to internalize the formal measurement procedures that they have been exposed to in school. If students' initial ways of conceptualizing weight and size are incommensurable with the assumptions of measurement, then they should have considerable difficulty in understanding and retaining what they learn about measurement at school. Finally, I examine coherency by looking at these students' responses to instruction.

The predictions about patterns of change in response to instruction will be described below.

Mechanisms of Conceptual Change

If children use the concepts of MT1 with imprecise, analog magnitude representations of physical quantities and undifferentiated senses of weight and spatial extent, how can they come to represent the very different concepts of MT2? What representational resources, both within and outside the domain of their initial matter theory, can they draw upon in developing this new understanding? What kind of learning processes do they use in exploiting those resources? If the preceding analysis is correct, three points follow about what the change process involves from a cognitive perspective.

Change Involves Conceptual Restructuring Not Simply the Revision of a Specific Belief

The first point is that change does not simply involve revising a mistaken belief or learning a new procedure, but making changes in the underlying concepts used to represent those beliefs or understand those procedures. Consequently, change will be a protracted process, requiring coordinated adjustments in how several concepts are articulated, in order to create a new network of concepts that is viable. In addition, there should be resistance to change, as students initially try to assimilate new information using their existing concepts.

To illustrate these difficulties, consider why the change cannot be brought about simply by telling students "All matter takes up space and has weight." Students will not understand this assertion as intended if they are using the concepts of matter, size, and weight that are articulated in MT1. On the one hand, this statement may seem false to them as they can think of counter-examples (e.g., they may think that a grain of sugar is matter but weightless). Further, given their reliance on assessing weight by hefting and the fact that the weight of some small objects is hard to detect, it is not a simple matter to challenge this counter-argument. Alternatively, if students accept and attempt to reason from this assertion, they may adjust their criteria for determining if something has an amount of matter and erroneously conclude that a little piece of Styrofoam actually has no amount of matter at all, given that it has no detectable felt weight. In our experience listening to classroom dialogues on this topic, we have observed students making these arguments that are incoherent from our perspective. To understand this statement as intended, they need to simultaneously adjust both their criteria for determining if something is matter and has weight.

Similarly, if the core of students' concept of weight is *felt weight* then the logic and procedures of weight measurement should seem counter-intuitive to them. (See also Lehrer, 2003, for a review of evidence that many fundamental assumptions about length, area, volume, and weight measurement are in fact not obvious to young children.) A fundamental assumption of weight measurement is that the weight of a given piece of stuff is the same regardless of the amount of stuff around it. Thus, one can analyze the weight of an object as the sum of the weights of its component (unchanging) parts. But this assumption runs counter to one's everyday perceptual experience of felt weight. Psychophysical studies have shown that adding a fixed size piece to an object produces a more noticeable weight increment to a very light object than it does to a heavy one. That is, the weight of the given piece is not perceived as constant or unchanging regardless of the amount of stuff around it. Rather, the perceptual system is responsive to the ratio of the weight of the added piece to the weight of the whole piece, with equal ratios (rather than equal units) being perceived as being of the same psychological size.

In addition, felt weight is a global, unanalyzed perceptual judgment made for whole objects. It is affected by both intensive aspects (the density of the object and its pressure on the hand) and extensive aspects (total absolute weight), further obscuring an understanding of the additive nature of weight. Analysis of the weight of the whole object in terms of the weight of its parts depends upon having an explicit theory of the factors that determine weight. Students need to differentiate weight and density in order to analyze and predict the weight of an object from knowledge of its volume and density.

Change Involves Using Conceptual Understandings and Representational Resources from Outside MT1 in the Construction of MT2

The second point is that new representational resources are needed in the construction of MT2, some of which come from outside MT1. In this sense MT2 is more powerful than MT1. What might some of these resources be?

Some resources are the general symbolic tools and quantificational resources of natural language (including the natural language quantifiers *all* and *some*) that can be used in stating new conceptual relations and universal principles. For example, "All matter takes of space and has weight; therefore, if something is matter, it must have some weight or take up some space." Although I have proposed that the concepts of MT1 are inter-related, I don't believe that MT1 is formulated using such precise universal principles. Rather children express MT1 principles using generic forms such as "Matter can be seen, touched or felt," "Big things tend to be heavy," or "Steel tends to be heavy," which imply "most" or "typically" rather than "all." Given that most 4-year-olds have learned the meaning of the quantifiers *all* and *some*, however, they can potentially use these resources in reformulating these generic principles as stronger universal principles. Once they do, they can potentially find counter-examples to these principles as well as search for alternative principles that are more valid.

Other resources are the conceptual understandings, symbolic tools and quantificational resources provided by culturally transmitted counting systems and by formal mathematics including its written notations for number and numeric operations and its algorithmic procedures that act on these representations. Both are used when children learn to measure physical quantities. Measurement allows students to go beyond inexact analog magnitude representations of physical quantities and to provide more precise symbolic representations of the magnitudes of these quantities. Given that children typically learn to count in the preschool years and are introduced to written notations for such numbers in the early elementary school years, the mathematics of counting numbers is available fairly early to help students begin the construction of MT2. Extending these resources to include an understanding of rational number, the notational conventions of fractions and decimals, and the operations of division, however, is a more protracted process that calls for change in children's concept of number during the elementary school years--a change that is often very difficult for many children (Gelman, 1991, Hartnett & Gelman, 1998; Smith et al., 2005).

Still another resource outside of MT1 is children's capacity for recursive thinking. When they combine their understanding of the physical process of breaking and subdivision (part of MT1) with recursive thinking (outside of MT1) they can create the new notions of fundamental constituent and continuous physical quantity, two central ideas of MT2.

Strong empirical support for the role of mathematical ideas and inscriptional practices in helping children develop differentiated and inter-related notions of volume, weight, and density has been provided by Lehrer, Schauble, and their colleagues in two recent instructional design studies with elementary school children (Lehrer, Schauble, Strom, & Pligge, 2001; Lehrer et al., 2002). They argue that learning to measure volume is central to understanding what volume is, and that differentiating volume and weight as separate measurable quantities is a pre-requisite to their co-ordination in a concept of density. Their elegant studies broaden the range of important mathematical ideas used as resources for developing understanding of these science concepts, as they teach elementary school children the mathematics of data, measure, similarity, proportionality, not just the mathematics of number. Their cogent and detailed analyses also show how these powerful mathematical ideas can be developed in elementary school children in the first place, foregrounding the interplay among varying forms of inscription with classroom talk and argument. In so doing, they consider how children can be led to redescribe their everyday experience in progressively more mathematical terms and the kinds of tools and talk that make this redescription possible.

Change Involves Using Thought Experiments and Analogical Reasoning to Develop New Ways of Looking at Everyday Observations

The third point is that change does not simply involve gathering new data, but developing new ways of looking at existing data. Many writers have argued that thought experiments and analogical reasoning are central to the explicit model building that supports the conceptual changes that have occurred in the history of science (Clement, 1989, 1991; Gentner et al, 1997; Kuhn, 1977; Nersessian, 1992), with mathematical modeling being a special case of analogical model construction. To the extent that conceptual change in childhood is like conceptual change in the history of science, then orchestrating thought experiments, analogical reasoning, and mathematical modeling of physical reality may be important in the childhood cases too.

Thought experiments are abstraction techniques (Nersessian, 1992): a way of exploring the logical consequences of a set of ideas in various idealized situations, including imagining what happens when the effects of a variable become extremely small or are entirely eliminated. They can be used to clarify and enrich one's understanding of initial ideas, to identify and sort out conflicting conceptual criteria (Kuhn, 1977), and to expand the range of situations considered in testing the explanatory adequacy of these ideas. Further, once imagined, the idealized situation can become the "simple" case, used in the formulation of basic principles not self-evident in everyday life, and the everyday situation can be re-analyzed as the more complex case (i.e., a product of multiple interacting basic principles).

Analogical reasoning is a central way that new conceptual resources can be brought into a particular domain (from another domain) or re-arranged and prioritized within a domain. Analogies, of course, can be deep (involving the mapping of an entire set of relations from one domain onto another) or more superficial (involving the mapping of more isolated properties). Deep analogies are particularly important in the process of conceptual change because these new sets of relations provide the scaffolding for the new model and concept definitions.

Gentner has written extensively about the way deep analogical mappings may work to bring about conceptual change in a target domain through the coordinated processes of highlighting, projection, re-representation and drawing alignable differences (e.g., Gentner et al., 1997). Initially, one establishes a partial analogical mapping between a source analogy and the target domain. The function of this initial mapping is to establish some correspondence between the domains and to highlight certain common relations. One can then use the analogy to extend one's theory building efforts in a variety of ways. For example, one can assume that the two domains share other relevant relations and project some other relations of the source domain onto the target domain, thus bringing new elements into the representation of the target domain. Or, one can use the analogy to note alignable differences between the two domains. In pursuing why these differences occur, one can engage in re-representation or re-description of one of the domains (often through the use of additional analogies) in an effort to bring the domains into closer alignment. Gentner's analysis thus shows the extended process through which analogies work, how multiple analogies may be used in making mappings between domains, and how ultimately aspects of both the negative analogy and neutral analogy may be projected onto the target domain. The latter is important because often disanalogies between source and target are prominent; only after the conceptual change is made is the full extent of the analogy understood.

What thought experiments and analogies may play a role in the move from Matter Theory 1 to Matter Theory 2? One important thought experiment involves imagining what happens as one divides an object into smaller and smaller pieces: Can you still see or touch the pieces? Do they still take up space and have weight? Does the matter itself disappear? If children just think about large scale macroscopic objects, then the inherent contradictions in defining matter as both "what objects are made from" and as "something that can be seen, felt, and touched" are not evident. By asking children to think about what happens in this limiting case (an idealization that goes beyond their everyday experience), students must consider a situation in which these features no longer co-occur. In addition, the thought experiment requires that students begin to represent the relations involved in repeated division--a set of relationships that can be used to clarify and define the new notion of "a fundamental constituent" and to imagine that matter is continuous at every point. This new notion can ultimately replace their more commonsense understandings of "made of" as "fashioned or molded from." Although the thought experiment can be resolved in more than one way, it contributes by both calling attention to a conceptual puzzle and highlighting certain relations that can be considered in its resolution.

Thought experiments and analogical reasoning are both involved when students consider the important analogy between the properties of objects at a macroscopic and microscopic level, as students cannot do direct experiments on this micro level. To get this analogy going, of course, students must have some beginning reason to believe that there might be pieces of matter too small to see. Such reasons can be based on their encounters with microscopes, experiences with tasting dissolved substances, or deductions made in prior thought experiments in which they imagine how tiny parts can be united to make a whole.

Not all macroscopic properties of objects are preserved at a microscopic level, so students need to use their background theories, available knowledge, and new data to decide what properties are preserved. Further, macroscopic properties that seem initially inapplicable to the microscopic level may be re-described in order to make the analogy work. For example, the property "can see with eyes" can be transformed to the property "can see with the aid of tools" such as magnifying glasses or microscopes or "can be seen by a creature with more sensitive eyes than mine.” Similarly, the property "feels heavy to me" can be transformed to the property "feels heavy to some very tiny creature" with sensitive sensory discriminations. In our talks with children, we have heard them make reference to such hypothetical and imaginative constructions.

Learning to measure weight and volume also involves analogical mappings (in this case between physical quantities and the domain of number in the introduction and construction of a unit of volume or weight). It supports reconceptualizing weight and volume as additive quantities, especially if teaching encourages metaconceptual discussion and reflection on some issues about the logic of measurement (Lehrer, 2003; Lehrer et al., 2003).[4] Discussing the logic of the procedures used in formal measurement--the need to identify an appropriate unit for a given quantity, to subdivide into identical units, and to count the number of units it takes to "cover" the object on that dimension--allows students to represent the quantity's additive structure explicitly. Further, once students assign an exact numeric value to a physical magnitude through measurement, they can use the numeric operations of addition and division to reason more precisely about the effects of these transformations. In this way, they can construct an explicit mathematical model and representation of the quantity in question that extends their understanding beyond what was possible with simple perceptually-based intuitive judgments.

Specific Hypotheses About Change That Are Tested in the Current Study

To examine how change occurs, my colleagues and I designed a curriculum that involved students with many of the reasoning processes listed above, gathered data on student thinking before and after the teaching, and observed their progress throughout the curriculum unit. The lessons simultaneously engaged students on multiple fronts: (a) explicit theory building about the nature of matter (i.e., searching for universal generalizations about the properties of matter, and testing the adequacy of these generalizations through a search for counter-examples); (b) analogical mappings between physical quantities and the domain of number as students clarify distinctions between weight and volume, identify distinct units of weight and volume, and discuss how to use these units in constructing distinct measures of weight and volume; and (c) thought experiments about the properties of matter that were preserved at a microlevel, building on their mathematical models of those quantities. Further details about the specific lessons are provided in the Methods section.

The study was designed to examine the pattern of change in individual students' thinking in response to this curriculum, not to compare the effectiveness of this curricular approach with alternative teaching approaches. To the extent that learning involves conceptual restructuring, it should not be easy. Students with the most entrenched alternative conceptions about matter should resist changing their concepts of matter, weight, and volume, especially if they have less developed understandings of the resources in other domains that are explicitly drawn on in the teaching. At the same time, engaging students in these mutually supportive and complementary reasoning processes should permit some students to reconceptualize their ideas about matter, weight, and size. Thus, when change occurs, students should make multiple, coordinated changes on several related understandings, rather than piecemeal changes.

Methods

Design

The study was a teaching study with identical pre/posttest assessments. All four Earth Science classes of one teacher were taught the same 10 lesson Matter Unit. The lessons were designed by our research team in collaboration with the regular classroom teacher who taught the unit. The Matter Unit was the second unit in a longer three-unit teaching sequence that involved: (a) an initial unit on discrete, intensive quantities in which students explored and solved problems with various per quantities, such as dots per box or people per square mile, and used them to create measures of crowdedness; (b) the matter unit in which students engaged with questions about the properties of matter and were challenged to reconceptualize weight, mass, and volume as distinct extensive and measurable quantities; and (c) a final unit on density of materials in which students clarified their understanding of density as an intensive quantity distinct from, but also inter-related with, their concepts of weight, mass, and volume. The pre/posttests were administered before and after the three-unit teaching sequence that took approximately 2 and a half months. Only the data relevant to the matter unit is reported here.

Subjects

The data for this study is based on the 42 8th grade Earth Science students who participated in the curriculum and who also participated in individual interviews, given before and after the curriculum unit. The students were selected for interviews to include students with a range of abilities across all four classes. In addition, all 68 students in the four Earth Science classes were given pre and post written test assessments by their classroom teacher. An analysis of the written test data revealed that the interview subsample was indeed a representative sample of the class. We focus our analysis on this group of 42 students, because both extensive written and interview test data is available for them.

Assessment

Each student first took a written test (administered as a group test) and then participated in individual interviews, both before and after the entire three teaching units. The interviews, however, were conducted without knowledge of students' performance on the written test. These assessment instruments were devised to probe student understanding of matter, weight, and volume in a variety of ways as well as their differentiation of weight and density. Because only the curricular unit on matter is discussed in this paper, only the items that relate to their understandings of matter, weight, mass, and volume will be presented here.

Three tasks (two on the interview and one on the written test) probed for their qualitative understanding that even small and light things must weigh something. The three tasks used different materials (Styrofoam, clay, and sugar). Further questions about Styrofoam also probed their qualitative understanding that very tiny things must take up space and have some amount of matter, and that matter continues to exist as it is repeatedly divided into smaller and small pieces.

In addition, one task on the written test probed for their ability to select and use information about a given object to make quantitative calculations of its weight, volume, and mass. Another written test item probed for their ideas about what entities are matter and the properties that all matter has.

Table 2 summarizes each of the tasks and the kind of understandings assessed.

Insert Table 2 about here

Qualitative Understandings: Do Tiny Pieces of Matter Have Weight, Take up Space, and Have Some Amount of Matter?

Clay task: Weight & amount of clay judgments (Interview).[5] Students were shown a clay ball about the size of a squash ball. A tiny piece of clay (about the size of a BB pellet) was then brought out and added to the larger clay ball (and pressed in). Students were asked two key questions: (a) Did I change *the amount of clay* in the ball when I added that piece? and (b) Did I change the *weight of the clay ball* when I added that piece? For both questions, students were asked to explain their reasoning. If they denied that there had been a change in the amount of clay or the weight of the ball, they were asked whether one *could* change the amount/weight by adding and to indicate how much one would have to add in order to do so.

Sytrofoam task: Weight, space, and amount of matter judgments (Interview). This task had multiple parts and was our main assessment of students' qualitative understandings of weight, space, and amount of matter. It probed students' ideas about whether progressively smaller and smaller pieces of Styrofoam had any weight. It also probed whether they thought those pieces had some amount of matter and took up space. Finally, it gave them a thought experiment about whether matter continues to exist with repeated division.

At the start of the task, students were handed a piece of Styrofoam the size of a cracker (about 1.5 x 1.5 x .25 inches) and were asked whether that piece "weighed a lot", "a tiny, tiny bit, "or "nothing at all" and why they thought that. Then students were given another much smaller piece of Styrofoam about the size of a BB pellet and were asked the same questions.

Students were next asked whether there could be a piece of Styrofoam too small to see with the naked eye. (Most students readily agreed that there could be based in part on their knowledge of microscopes.) We then asked them two questions about such a very tiny piece:

Would that tiny piece of Styrofoam take up any space? How do you know? Would that tiny piece of Styrofoam weigh anything? How do you know?

Finally, we took them through a series of questions about the "amount of matter" in the Styrofoam pieces. We again gave them the two pieces of Styrofoam (one about the size of a cracker and the other about the size of a BB) and asked of each piece whether it "had a lot of matter," "a tiny, tiny bit of matter," or "no matter at all" and how they knew that. We concluded by presenting them with the following thought experiment. We asked them to imagine that they had special tools (such as laser beams) that allowed them to cut very small pieces of matter. The question was whether the Styrofoam could be cut in half forever or whether there would be a point at which there was no more matter left to divide.

The weight of grains of sugar task (Written test). Finally, the written test also included a brief assessment of students' beliefs about the weight of small pieces of matter. This item read as follows:

A large sack of sugar is very heavy. It contains millions of grains of sugar.

1. Does a pile of 10 grains of sugar weigh anything at all? How do you know?
2. Does one grain of sugar weigh anything at all? How do you know?

Quantitative Understandings: Measurement Tasks (Written Test)

On the written test, students were shown a drawing of a cube in balance with 9 individual gram pieces on a balance scale. The cube was also scored into 1cm^3 units to facilitate the calculation of volume (see Figure 1). Students were given the following instructions. "Here is an object that is in balance on the scale. Answer the following questions about the object, giving the appropriate numbers and units.

1. What is the volume of the object? How did you figure that out?

2. What is the weight of the object? How did you figure that out?
3. What is the mass of the object? How did you figure that out?"

Insert Figure 1 about here

Although this item does not assess the many facets of understanding of measurement that have been described by Lehrer (2003), it does probe (a) whether students' differentiate among relevant quantities (e.g., weight, volume, length, area, etc.); (b) identify a relevant unit for those quantities (e.g., grams vs. cubic centimeters vs. centimeters); and (c) use an appropriate procedure to figure out how many of those units "cover" the relevant measurement space.

Matter Categorization and Justifications Task (Written Test)

On this written test item, students were given a list of 12 entities and asked to decide for each whether or not they thought it was matter (that is, whether they thought it was made of some kind of physical material or not). These entities were: a rock, a dog, a tree, water, a grain of sugar, a particle of chalk dust, air, heat, shadow, echo, idea, wish. After making their judgments, they were asked two further questions: (a) What are the properties of matter? and (b) How can one tell if something is made of matter?

Coding and Scoring

Each task was coded separately, independent of knowledge of the subject's response on the other tasks by one coder. Coding of individual questions was based on objective criteria (e.g., judgment that an item was or was not matter, judgment that a piece of something weighed a tiny bit or nothing at all, correct determination that the object in Figure 1 weighed 9 g and had a volume of 27 cm^3). In cases where students changed their mind about a given judgment, codes were based on student's final considered judgment. Such cases almost always occurred on the pretest, where a few students spontaneously corrected an initially incorrect judgment. The coder also read students' justifications, which overwhelmingly were consistent with their judgments. In only a few cases (less than 5% overall) did the information in the justification lead the coder to assign the response to a different category than would have been suggested by the judgment alone. Further details about the coding categories used for each task will be provided in the results section.

The Matter Unit

Overview of the Ten Lessons

The teaching intervention was designed to help students move from the more perceptually-based concepts of size, weight, and matter that characterize commonsense Matter Theory 1 to the more abstract (and quantified) concepts of volume, mass/weight,[6] and matter that characterize Matter Theory 2. Briefly, students explored the properties of clear cases of matter (large-scale solids and liquids) in preparation for investigating less clear cases (small-scale solids and liquids, air). Students also learned about the measurement of mass and volume in the context of clarifying their understanding of these quantities and deepening their understanding of the fundamental properties of matter. A pivotal part of the lesson sequence involved using their emerging abilities to quantify mass and volume to derive the mass and volume of very tiny amounts of matter by division--a thought experiment that supported the conclusion that taking up space and having mass are fundamental properties of all matter. In the case of mass, this thought experiment was followed with the experience of finding the mass of very light objects using an extremely sensitive scale (capable of detecting masses of .0001g) to provide empirical confirmation that objects that have no "felt weight" nonetheless have measurable mass.

The teaching unit began by eliciting students' initial ideas about what entities are matter and what properties may be characteristics of all matter. The purpose of this initial class was to get students thinking more deeply about the topic and to create some disequilibrium. Students proposed different potential properties of matter (e.g., something that you can see, feel, touch, takes up space, has weight) and then were asked to make arguments for or against a particular proposal by looking for possible counter-examples. Thus, they considered whether they could think of something that was matter but (a) couldn't be seen or touched (b) didn't take up space or (c) didn't have weight. They also had a lively debate about whether air, smoke, heat, light, and other puzzling cases were or were not matter.

Students next embarked on a series of lessons designed to clarify their understanding of one potential property of matter--taking up space--by investigating how it could be quantified and measured and then using that quantification to reason about the amount of space taken up by very small things.   Students first engaged in a thought experiment to clarify that "taking up space" meant "occupying space" such that two things couldn't be in the same space at the same time. They were then challenged to go beyond crude perceptual judgments to create measures of the amount of space taken up by solids and liquids.   For example, they then were asked to determine whether three rectangular wooden objects of different dimensions (and shapes) took up the same or different amounts of space. The problem was a challenge because they could not tell from looking if the objects had the same volume. After eliciting their competing predictions based on everyday observation, the teacher asked students how they might be able to resolve their disagreements. In this context, she introduced the idea of trying to measure the volume of each object, using a standard unit of volume, via a cubing procedure. Students were divided into groups, given the three objects and a large number of cubes, asked to determine the volume of the objects using cubing, and discussed their methods and findings. This exercise not only introduced them to a unit of volume (1 cm^3), but also highlighted the additive structure of volume measurement (i.e., the volume of the whole object is the number of volume units it takes to compose or fill the object). Students also worked through a variety of other problems, including finding the volume of an irregular object (a staircase) and designing two objects with the same volume but different dimensions. The teacher then used student understanding of cubing to anchor their understandings of the mathematical formula for finding the volume of rectangular solids and their understanding of the measurement of the volume of liquids. Finally, the teacher asked students to use mathematical reasoning to estimate the volume of things (e.g., a drop of water) that were too tiny to measure directly. For example, to determine the volume of a single drop of water, students first found the number of drops that equaled 1 ml (11 drops) and then divided that volume by 11 to determine the volume of 1 drop (.09 ml). In this way, they used their mathematical understandings of fractions and decimals to bootstrap their understanding that very small physical quantities can have a magnitude that is less than 1 but greater than 0.

Students next pursued a series of lessons to introduce the term "mass," clarify their understanding of the property of mass (as related to weight and distinct from volume), to quantify this property, and to use this quantification to derive the masses of very light objects. Initially, students were given a variety of objects made of different materials (some the same volume and different mass, others the same mass, but different volume) and were challenged with the question: How could we compare the mass of these objects? Again, students discussed a variety of ways of comparing mass, such as by feeling, looking at the size, putting on a scale, and also discussed which way might be most reliable. Students used balance scales to make some initial (qualitative) comparative judgments, and considered the advantages of using balance scales (over felt weight judgments) in comparing mass. They also discussed the contrast between mass and volume, by noting objects could be the same mass, but different volume, and vice versa. In the next class, the students were further challenged to *quantify* the mass of a variety of objects, using the balance scale and a set of 1, 5, 10, and 100 gram pieces. This class introduced students to a gram as a standard unit of mass, and the activity of finding the masses of various objects emphasized the additive structure of mass measurement. Finally, students investigated the mass of small, light objects such as a single lentil. They were led to use their new understanding of mass measurement to estimate the mass of very tiny things through the use of division. To provide even more direct evidence that very tiny objects have measurable mass, they saw a video of some classmates making measurements of other very light objects (such as a piece of glitter) on a very sensitive scale (an analytical balance). The fact that even these items had measurable mass provided further evidence that matter continues to have weight as it is divided into tiny pieces. This class also provided a context for a discussion about the sensitivities of different scales.

Finally, the class returned to the issue of whether air was matter. In previous classes, students had investigated the proposal that having mass and volume are essential properties of matter, clarified their understanding of these properties as quantifiable and measurable magnitudes, and used mathematical reasoning to estimate the mass and volume of very tiny items. The class now considered if there were ways to determine the mass and volume of gases, such as air, as a way of investigating whether or not gases were in fact material. To demonstrate that air had measurable mass, the teacher showed that a balloon filled with air pulled down more on a balance scale than an unfilled balloon. To demonstrate that air takes up space, she did a lung capacity demonstration in which exhaled air displaces water in a tube. Students discussed the implications of these findings for the issue of whether air was matter.

Classroom Data and Its Analysis

There were two sources of data that provided information on student progress throughout the unit: (a) general classroom observation notes (made by research assistants who observed the classes); and (b) data from individual student portfolios (kept by the classroom teacher), which contained each student's classroom work, homework, and end of unit tests. The classroom observation notes were taken in real time, by an observer who took notes on the gist of whole classroom discussions (i.e., noting questions raised and the responses made by students and the teacher). When the class worked in small groups, the observer moved among groups and made general notes on how students were approaching the activities (e.g., did they understand the activity, or were they confused; what strategies were being used in solving the problems.) Data from student portfolios were also reviewed to provide more information about how individual students were engaging with the lessons. Here the relevant unit of analysis was the individual student response to the homework or in-class problems. Each problem was analyzed separately by one coder who noted which students (and how many) solved the problem correctly, the kind of reasoning that was used on the problem, and the kinds of confusions that occurred. Although the data sources for these analyses were not as rich as would have been the case if we had videotaped each class, or if we had followed in detail a target group of students and asked them probe questions while they were working on a problem to fully understand their reasoning, they did allow us to reconstruct the gist of student progress through the unit. That is, we could identify areas that provoked lively debate and student disagreement and problems that were solved relatively easily or with more difficulty.

Results

Pretest Results

Pretest results were analyzed to bear on three main issues: (a) whether students regarded *felt weight* as more central to their concept of weight than being a *property of matter*; (b) whether students' conceptions of weight cohere with their conceptions of size and matter; and (c) whether students who regard felt weight as central to their concept of weight lack an understanding of weight and volume measurement.

Qualitative Conceptions of Weight: Weight as Felt Weight vs. Property of Matter

Clay & Styrofoam tasks: Weight judgments (Interview). In both these tasks we asked students about the weight of pieces of matter that they could see and hold but which had no appreciable felt weight in order to see which characteristics of the piece would take priority in their judgments. If the core of weight is *felt weight*, then *felt weight* should take priority and students should assert that the small pieces weigh nothing at all. In contrast, if weight is conceptualized as an essential *property of matter*, then the fact that the pieces are matter should over-ride their negligible felt weight and lead to the assertion that they must weigh something.

Overall, there were three main judgment patterns across the Clay and Styrofoam tasks:

(a) *Hard Core Felt Weight*. The first pattern was to consistently judge that very small light objects weigh nothing at all. These students said that adding a small (BB size) piece of clay did not change the weight of the ball; one would need to add a much larger piece to have any effect. In addition, they said that the visible BB size piece of Styrofoam weighed nothing at all. (Indeed, the majority even thought the cracker size piece of Styrofoam weighed nothing at all.) They also maintained these judgments in the thought experiment about whether an invisible piece of Styrofoam would weigh anything at all. Overall, 26% of the students

(11/42) had this pattern of response at the time of the pretest.

(b) *Transitional Felt Weight.* The second pattern was to judge that the small piece (BB size) piece of clay would make the clay ball heavier, but then to judge that the BB size piece of Styrofoam (as well as the invisible piece) weighed nothing at all. Although the small visible piece of Styrofoam and Clay were about the same size, Styrofoam is much less dense than clay and hence presents a much more extreme case. These students may be "transitional" in that they need more extreme cases to say that something weighs nothing at all. Overall, 29% of the students (12/42) had this pattern of response. (One student was included in this category who only judged the invisible piece weighed nothing at all.)

(c) *Weight as Property of Matter*. The third pattern was to consistently judge that even small pieces of matter must have some weight. These students not only said that a small piece of clay would definitely make the clay ball heavier, but also maintained that even very tiny pieces of Styrofoam--those too tiny to see with the naked eye must weigh something. Overall 45% of the students (19/42) showed this pattern of responding.

Analyses of student justifications supported the assumption that students with Hard Core and Transitional Felt Weight patterns were relying on felt weight judgments and did not have a general belief that all material objects must weigh something. More specifically, students who said that the small piece of Styrofoam weighed nothing at all appealed to its small size and lack of felt weight. For example, "it's too small, it has no weight," "it feels like nothing in hand," and "I don't feel any pressure when I pick it up," "I can feel how light it is, it doesn't weigh my hand down at all." One even explicitly articulated the non-additive view: "The tiny piece is only a part of big; it weighs nothing, but a bunch would weigh something."

In contrast, all of those who asserted that the Styrofoam weighed something articulated a belief in a *general principle* that forced them to conclude that the Styrofoam *must* weigh something despite its negligible felt weight. That is, their justifications used the words "has to" or "must", or appealed to universal generalizations about objects, matter, things that take up space or "everything." For example, they said: "It's matter, so it has to weigh something;" "Everything has to have some weight;" "Anything that takes up space has weight;" "An object has to weigh something;" "If larger weighs something, the smaller does too; it has to because it is a piece of the larger and the larger weighs."

The weight of grains of sugar task (Written test). If there are real differences in the way students conceptualize weight, then these differences should be reliably diagnosed with other tasks of similar sensitivity as well. To test this possibility, we developed a written test item that presented students with another extreme case: asking them whether ten grains and one grain of sugar weighed anything at all. Despite the fact that there were numerous differences in testing procedure between the Sugar and Styrofoam tasks (e.g., giving the task in written rather than interview form, asking students to write out rather than orally explain their reasoning, presenting questions about hypothetical objects rather than objects placed in their hands, and administering the task on another day), 90% of the students had concordant patterns on the two tasks (see Table 3, $x^2(1, N = 42) = 27.8, p < .001$). That is, all but three students who judged that a small piece of Styrofoam weighed nothing at all, also judged that one grain of sugar weighed nothing at all, and all but one student who judged that the small piece of Styrofoam must weigh something also judged that one grain of sugar must weigh something.

Insert Table 3 about here

Summary Measure 1: Weight judgments. All together, there were six items (4 in the interview; 2 in the written test) on which students were questioned about whether very small objects had weight: a small piece of clay, a medium, small, and invisible piece of Styrofoam, and 10 grains and 1 grain of sugar. A summary quantitative measure of the strength of students' belief that very small objects have weight (to be used in later quantitative analyses) gave students one point for each judgment that a small object had some weight. Scores ranged from 0 (no judgments that small objects had weight) to 6 (all judgments that small objects had weight).

A split-half reliability of this composite measure (comparing the quantitative score on the three Styrofoam questions with the quantitative score on the Clay and two Sugar questions) showed reasonable consistency: inter-correlations of .83**.** This finding supports the conclusion that students do vary in their reliance on felt weight in their judgments, and that this quantitative summary measure provides another way of describing variations in students' ways of conceptualizing weight that is complementary to our more categorical scoring systems.

The Relation Between Conceptions of Weight and Conceptions of Taking up Space, Amount of Matter, and Divisibility of Matter

If students' concepts are embedded in commonsense matter theories, then students who conceive of weight as felt weight should be more likely to have alternative, perceptually-based conceptions of taking up space and matter as well. Additional questions within the Styrofoam and Clay tasks were relevant to testing these hypotheses as they probed whether students thought tiny pieces of matter took up any space, had any amount of matter, and continued to exist with repeated division.

Categorical analyses. Table 4 shows there were dramatic differences in the way students answered these questions depending on their conception of weight. Those with Hard Core Felt Weight patterns typically said that the invisible piece of Styrofoam takes up no space at all, that a tiny visible piece of Styrofoam has no amount of matter, and that the matter itself will ultimately disappear with repeated divisions. In contrast, those with a Weight as Property of Matter pattern typically said that the invisible piece of Styrofoam takes up some space, that the small visible piece of Styrofoam has some amount of matter, and that the matter would continue to exist with repeated division. Those with Transitional Felt Weight patterns were in between on any given judgment.

Simple 2 x 2 chi-square tests of association relating MT1 vs. MT2 conceptions of weight to MT1 vs. MT2 conceptions of taking up space, amount of matter, and divisibility of matter revealed highly significant associations. In these analyses Hard Core and Transitional Felt Weight patterns were grouped together and contrasted with Weight as Property of Matter patterns. The specific values were: $x^2 (1, N = 42) = 17.3$ (Weight & Occupy Space); $x^2 (1, N = 42) = 12.3$ (Weight & Amount of Matter); and $x^2 (1, N = 42) = 8.2$ (Weight & Divisibility of Matter), all significant well beyond the .001 level. In addition, students with Weight as Property of Matter patterns were much more likely to be correct on all three space/matter items than those with Felt Weight patterns ($x^2 (1, N = 42) = 15.9$, $p < .001$).

Insert Table 4 about here

Summary Measure 2: Amount of matter and space judgments. Four items probed students' understanding that small objects have some amount of matter (a small piece of clay, the medium and small piece of Styrofoam, and the Styrofoam thought experiment); one item probed students' understanding that small objects take up space (questions about whether an invisible piece of Styrofoam takes up any space). A summary, quantitative Matter/Space Judgment score gave students one point for each judgment that a small object had an amount of matter or took up space. Overall, there was a substantial positive correlation between students' judgments that small objects have weight, on the one hand, and that they take up space and have some amount of matter, on the other (Pearson $r = .62$, $p < .001$). This provides further support for the hypothesized relations among students' conceptions of weight, taking up space, and matter, as suggested by our conceptual analyses of MT1 and MT2**.**

The Relation Between Conceptions of Weight and Matter

If students with Felt Weight patterns vary in their conceptions of matter, then these differences should show up in a task with a very different format from the Styrofoam task. The Matter Categorization and Justifications task asked students to decide which items are matter and not matter and to list explicitly the properties of matter. Again, students with Felt Weight (Hard core or Transitional) patterns had different patterns of matter judgments and justifications from those with Weight as a Property of Matter patterns.

Both groups of students made some distinction between material and immaterial entities: material entities were more likely to be judged as matter than non-material entities (see Figure 2). However, the distinction was clearer and sharper for students with Weight as Property of Matter patterns as compared to those with Felt Weight patterns. Students with Weight as Property of Matter patterns were virtually at ceiling for all items except air and heat. In contrast, students with Felt Weight patterns made many more under-extension errors on solids, liquids, and powders (*t-test*, $p < .01$, 1-tailed) and slightly more over-extension errors (heat, echo); hence, the contrast between their judgments on these items and non-material entities was less sharp.

The two groups of students also differed significantly in what they said explicitly were the properties of matter (see Table 5). About half (46%) of the students with Felt Weight patterns either left this question blank, gave examples rather than properties, or mentioned some totally irrelevant property (e.g., matter is something you can use, is natural, has parts or doesn't move). Another large group (41%) of students with felt weight patterns said that matter was something you could see, feel, or touch. In contrast, the students with Weight as Property of Matter patterns refrained from mentioning these commonsense perceptual properties. Instead, they focused on its being some kind of physical material (i.e., "the physical material things are made of"), "something that is solid, liquid, or gas", something that is "physically there" or that is "real and exists", or mentioned the objective properties of taking up space or having weight. Overall, 84% of the students with Weight as Property of Matter patterns gave these kinds of answers compared to 13% of those with Felt Weight patterns ($x^2 (1, N = 42) = 21.3, p < .001.$)

Insert Figure 2 and Table 5 about here

The Relation Between Qualitative Conceptions of Weight and Understanding Weight and Volume Measurement

If students with Felt Weight patterns do not explicitly conceptualize weight and taking up space as essential properties of matter, then they should not understand the logic of weight and volume measurement. Weight measurement involves analyzing the weight of the whole object as the sum of the weight of its component parts. Similarly, volume measurement involves analyzing the volume of the whole object as the sum of the volume of its component parts. Such decompositional analyses should not make sense to them if they think that small objects can weigh nothing at all or take up no amount of space.

Categorical analyses. Table 6 shows the relation between students' weight judgment patterns and their ability to provide correct weight and volume measurements for the object depicted in Figure 1. As expected, students with Hard Core Felt Weight patterns knew virtually nothing about weight measurement--only one counted up the gram pieces to determine the object's weight. Their most common responses for weight measurement involved some conflation of units or procedures for measuring space and weight (55% of the time) or to leave the question blank (27% of the time). Although they were no more successful at making a correct volume measurement, their pattern of errors was different. Most responses correctly focused on some spatial aspect (72%), but did not clearly differentiate volume from other spatial dimensions or attempt to calculate the spatial extent. Students with Weight as Property of Matter patterns knew more, although they were not yet at ceiling: almost half now had correct (and contrasting) procedures for measuring both the weight and volume of an object. Students with Transitional Felt Weight patterns were in between: 42% were able to add up the number of gram pieces on a scale to determine the object's weight, but only one (8%) also had a correct (and contrasting) procedure for measuring volume.

Chi-square analyses contrasting type of weight pattern (Felt Weight vs. Property of Matter) with incorrect or correct measurement patterns revealed significant associations for both Weight Measurement ($x^2(1, N = 42) = 4.4, p < .05$) and Volume Measurement ($x^2(1, N = 42) = 6.6, p < .01$).

Insert Table 6 about here

Summary Measure 3: Measurement of weight and volume. The third summary measure concerned students' understanding of weight and volume measurement. Three points were given for each fully correct measurement (correct numeric value, appropriate units, and explanation of procedure), 2 points for an almost fully correct measurement (correct numeric value and either correct units or procedure), and 1 point for an answer which articulated a relevant procedure (i.e. count all the cubes or multiply L x W x H), but did not provide the correct numeric value. The vast majority of scores were either 3 or 0.

There were substantial correlations among all three summary pretest measures. Both Weight/Volume Measurement and Amount of Matter/Space Judgments remained significantly correlated with Weight Judgments even when the effects of the other variable was statistically controlled (Table 7), arguing that each made a distinctive contribution to students' conceptions of weight. A step-wise regression analysis showed each variable made a distinctive and significant contribution to predicting Weight scores and that the two together accounted for almost half the variance in Weight Judgments at the time of the pretest ($R = .69$, $R^2 =.47$, $p < .001$).

Insert Table 7 about here

Analysis of Student Work & Class Discussion During the Teaching Unit

We observed all classes and reviewed all available student worksheets and homework in an effort to better follow the evolution of student thinking throughout the teaching unit. Because students worked collaboratively on classroom worksheets and the teacher often scaffolded their solution to problems, classroom worksheet data tends to overestimate the comprehension of individual students. Nonetheless, these data, especially when supplemented with our class observation notes, allow us to reconstruct the gist of student reaction to the various activities. It also provides further confirmation of the limitations in some students' initial understandings of matter, weight, and taking up space that we had documented in the pre-interviews.

Students enjoyed debating what was matter/not-matter in the opening class discussion, where it was clear that students had different views. After hearing the different viewpoints and arguments made in class, students were asked to reflect on the properties they were most confident that all matter had (at this point). The data from both the class discussion and student homework confirmed that "having weight" was not yet regarded as a central property of matter for many students: it was not argued for as often in the class discussions or on the homework sheets as were the properties of being perceptually accessible or taking up space. Further, even when it was listed on the homework problem, it was typically *added* to lists of commonsense properties rather than replacing them. Being perceptually accessible was especially important for those students who in the pre-interviews had judged that Styrofoam weighs nothing at all.

The volume lessons confirmed that most students initially had little idea about how to measure the volume of the objects and confused different spatial dimensions when attempting to quantify how much space something took up (length, area, perimeter, surface area, etc.). The opening task challenged them to determine the volume of three differently shaped rectangular objects by building the objects with cubic centimeters, but "cubing" objects was by no means a straightforward procedure. This activity revealed a number of divergent approaches (e.g., making a single row of cubes equal to the block's length or a single layer of cubes, ignoring the third dimension) that needed to be worked through and discussed. Successful completion of this task helped to anchor their understanding that they were really trying to measure the entire amount of space that the object occupied (not just a length or an area) and that a cubic centimeter was a standard unit for volume. In follow-up homework, they were asked to make drawings that represented the volume of objects and to use those representations to aid in making quantitative calculations of volume. For example, students were asked to score a picture of a 3-dimensional object (to show cubic centimeters) and to calculate its volume. They were also asked to use graph paper to design two different 3-dimensional objects that had a volume of 40 cubic centimeters. Again, these were challenging problems that brought to the surface a number of confusions. For example, some students drew objects that were equated in the area of the front face rather than total volume; others calculated the surface area rather than the volume. In retrospect, it would have been desirable to have even more problems of this type, and to have had more time to discuss these confusions in class. Nonetheless, there was evidence that students were making progress in sorting out some of their "buggy" procedures and misunderstandings. By the last homework in the volume section, the majority of students were now correctly attempting to compute the volume of rectangular solids by multiplying all three dimensions; only a minority made errors of computing an area or giving a length.

The lessons on volume concluded by challenging students to determine the volume of a single drop of water. Students first measured how many drops of water it took to equal 1 ml on a graduated cylinder (about 11). The class then discussed how one could use that information to infer the volume of a single drop (by dividing 1 ml by 11 drops)--assuming each "drop" was the same volume. Students generally understood the logic of using division to obtain the volume of a single drop of water, although this lesson was a highly teacher-scaffolded activity and some had difficulty with the math algorithms for division and placing the decimal point correctly.

Although many students did not understand how to measure weight or mass at the outset, these procedures were generally more straightforward for them to learn than those for volume, as students simply needed to put an object on a balance scale and then see how many gram masses they had to add to the other side until the two sides were in balance. However, being able to carry out these procedures by no means meant that students' understood weight as an extensive property of matter. Both classroom observation data and worksheet responses indicated that this was still a live issue for some students. For example, after learning about the measurement of large-scale solids and liquids, students were asked if they thought all matter had mass and to list examples of things that might be matter but have no mass. In addition, they were asked if 50 lentils and a single lentil had any weight. Over half (65%) of the students with initial Felt Weight patterns either judged that the lentil weighed nothing at all or listed something that they thought might not have weight. (Among the items listed by these students were: flea, dust mite, grain of sand, chalk dust, a hair and a snowflake.) Thus, two full class periods were devoted to the question of whether a wide range of very light objects had any mass.

To help students investigate this issue, students were directed to find out how many lentils weighed 1, 2, and 4 grams. They were then challenged to look for patterns in their data (i.e., it took approximately twice as many lentils to weigh 4 grams as 2 grams) to see if they could figure out a way to estimate the weight of 1 lentil. Most students were able to estimate the mass of 1 lentil by division (determining the number of lentils it took to have a mass of 1g, and dividing 1g by that number of lentils to get the mass of 1 lentil). However, 40% of the students with initial Felt Weight patterns did not clearly use division either with this problem or a related problem asked on a unit test given by the teacher. They typically left the problem blank or gave entirely inappropriate answers that were greater than 1 g.

Students continued to investigate the weight of very tiny objects in the next class, by considering the weight of other very light items (e.g., a drop of water, a piece of glitter, a dust ball, a piece of hair, an ink signature). Class votes were taken and a few students were still unsure about these items, especially as they got more extreme (e.g., a dust ball, a piece of hair, an ink signature). The students then watched a video in which these items were massed on an analytical balance (which had a digital read-out) and recorded their masses. These ranged from .03 g for the drop of water to .0009 g for the hair or ink signature. Students were very interested and involved in making predictions. By the end of the class, sentiment was high that even these tiny items had measurable mass, which increased their confidence that all matter had mass and weight.

In the final class on the properties of matter, students returned to the question of whether or not air was matter--a question about which there had been considerable debate in the first class. Students now were prepared to try to settle this issue by finding out whether it had measurable volume and mass. The members of the class were interested in--and for the most part quite convinced by--two demonstrations showing that air had measurable mass (comparing an empty and filled balloon on an equal arm balance) and that it took up space (lung capacity demo showed that air pushed out water; balloon demo showed that air expanded the balloon). In response to teacher questions, students concluded that air was matter; most also considered having mass and taking up space as more central properties of matter than being directly accessible to our senses.

Posttest Results

Posttest analyses bear on three specific questions: (a) Did students change their ways of conceptualizing weight, space, and matter and learn how to measure weight and volume? (b) If so, how did change occur? Were changes coordinated or piecemeal? (c) What understandings, if any, predict students' readiness for change?

Did Students Change Their Ways of Conceptualizing Weight, Space, and Matter and Learn How to Measure Weight and Volume?

Students made clear progress from pretest to posttest in all tasks, although some students were still not at ceiling on the posttest tasks. In this section, a simple summary of the extent of progress is given. In the next, the pattern of change is considered.

Styrofoam/Clay tasks: Weight judgments. Students were asked whether small light pieces of Styrofoam weigh anything and continue to weigh something as they are divided into pieces too small to see; they were also asked whether adding a small piece of clay to a clay ball makes the ball heavier. There were three main patterns of response: (a) Hard Core Felt Weight patterns (consistently judging that small light objects weigh nothing at all) (4 out of 42 or 10%); (b) Transitional Felt Weight (making at least one judgment that some pieces of matter weigh nothing at all) (7 out of 42 or 17%); and (c) Weight as Property of Matter patterns (consistently judging that all pieces must have some weight because they exist and are matter) (31 out of 42 or 73%).

Table 8 summarizes the changes in students' conceptions of weight, as assessed by the Clay and Styrofoam tasks used in the interview (4 items). Two-thirds of those with Hard Core or Transitional Felt Weight patterns improved in their pattern from pretest to posttest, although almost half still did not have full Weight as Property of Matter patterns. Among those with Hard Core Felt weight patterns, 36% improved to Transitional Felt Weight patterns and 36% improved to Weight as Property of Matter patterns. Among those with Transitional Felt Weight patterns, 67% improved to a Weight as Property of Matter pattern. Thus, although there was evidence of improvement, there was also evidence of resistance to change. All students who had consistently judged that tiny pieces of matter have weight at the time of the pretest maintained this strong pattern on the posttest.

Insert Table 8 about here

Styrofoam/Clay tasks: Amount of matter/space judgments. At the time of the pretest, almost all students judged that the medium Styrofoam and BB size clay had some amount of matter. However, only 67% judged that a BB size piece of Styrofoam had an amount of matter; only 67% judged that a piece of Styrofoam too small to see took up space; and only 55% judged that Styrofoam matter continued to exist with repeated division. By the time of the posttest, students had improved on these judgments, but still were not at ceiling. Now 93% judged that the BB size Styrofoam had some amount of matter; 79% judged that the invisible piece of Styrofoam took up some space; and 71% judged that matter continued to exist with repeated division.

Matter categorization & justifications (Written test). The Matter Categorization and Justifications task provided further evidence of changes in students' ways of conceptualizing matter. Students were asked to categorize 12 entities (e.g., rock, tree, grain of sugar, water, air, heat, shadow, etc.) by whether or not they were matter and to provide a general characterization of the properties of matter. By the time of the posttest, students were more accurate in making judgments about what was matter: indeed they were virtually at ceiling on 10 of the 12 items, only making errors on "air" (83% correct judgments) and "heat" (67% correct judgments). Similarly, there was a big shift from pretest to posttest in students' knowledge that matter was something that took up space or had weight. At the time of the pretest, only 29% explicitly mentioned that matter was something that takes up space or has weight in this task. In contrast, by the time of the posttest 80% of the students mentioned either taking up space or having weight as an important property of matter.

Weight, mass, and volume measurements (Written test). The Measurement tasks gave students information about an object from which they were asked to infer its weight, mass, and volume (see Figure 1). At the time of the pretest, only a minority of students could use this information to determine the weight, mass, or volume of the object correctly (40%, 5%, and 21%, respectively). By the time of the posttest about three-quarters of the students understood how to use the balance scale information to determine weight and mass and almost two-thirds understood how to use information about the cubic structure to determine correctly the total volume of the object.

Summary. Students made improvements on every task in our battery. The proportion who improved was almost 100% for the easier items but lower (50-60%) for the more challenging items. The fact that some difficulties remained on the posttest suggests that achieving these insights is not simple, even with a curricula targeted to making these points.

How Did These Changes Occur? Were Changes Coordinated or Piecemeal?

Coordinated vs. piecemeal patterns of change. One way to assess the nature of change is to examine the performance of the subgroup of students who had Felt Weight patterns on the Styrofoam/Clay tasks at the time of the pretest. As Table 8 indicates, half of these students progressed to Weight as Property of Matter patterns whereas the other half did not. If change is piecemeal, change on one task should not be related to change on another task; both groups should make similar amounts of progress, albeit on different tasks. In contrast, if change is coordinated, then those who progressed to the Weight as Property of Matter pattern should have made simultaneous progress on several related tasks, while those who maintained Felt Weight patterns should have kept essentially the same way of responding on most tasks.

To test this hypothesis, students' pattern of change in weight judgments (Styrofoam task) were related to their performance on three other tasks: (a) amount of matter and space judgments (Styrofoam task); (b) weight measurement (Measurements task); and (c) identification of properties of matter (Matter Categorization task). For each task, a pattern of performance was identified that would reflect the concepts and beliefs of Matter Theory 2: (a) correct judgments that tiny items have some amount of matter, take up space, and continue to exist with repeated division (4 items correct); (b) correct weight measurement; and (c) explicit statement that having weight or mass is a general property of matter (in the Matter Categorization task). For each student, we tallied the number of tasks for which they had Matter Theory 2 patterns at the time of the pretest and posttest and used these numbers to determine each student's change score (i.e., the number of tasks for which they moved from Matter Theory 1 to Matter Theory 2 patterns between pretest and posttest.)

The two groups showed no difference in sophistication on the other tasks at the time of the pretest. Across both groups, 60 percent had *no* Matter Theory 2 pattern for any task; another 30% had only one. Thus 90% of students with Felt Weight patterns showed fairly consistent Matter Theory 1 patterns across the tasks.

In contrast, the two groups were dramatically different by the time of the posttest (see Table 9). Those who remained with Felt Weight patterns typically made no change on the other tasks or at most changed on only one. Those who changed from Felt Weight to Weight as Property of Matter patterns, however, made coordinated changes on the other tasks as well. A 2 x 2 Chi square analysis comparing no or piecemeal change (change on 0-1 tasks) with coordinated change (change on 2 or 3 tasks) revealed that a highly significant difference for the two Weight groups (x^2 (1, N = 42) =16.3, $p < .001$)

Insert Table 9 about here

Relations among posttest conceptions. Another way to examine the degree to which change preserves conceptual coherence (either through resistance to change or coordinated patterns of change) is to analyze the organization of posttest conceptions. In fact, conceptions were just as systematically related at the time of the posttest as they had been at the time of the pretest.

Consider, first, the relations between students' conceptions of weight and their conceptions of space, amount of matter, and divisibility of matter. Table 10 shows that most students (84%) who had Weight as Property of Matter patterns also understood that tiny visible pieces had some amount of matter, invisible pieces took up some space, and matter continued to exist with repeated division. In contrast, only one (9%) with Felt Weight patterns had all three understandings (Fisher exact test, $p < .001$)

Further support for this conclusion comes from an examination of these students' responses in the Matter Justification part of the Matter Categorization task (Table 11). The most common response of students with Felt Weight patterns was to mention commonsense perceptual properties. In contrast, the most common responses for those with Property of Matter patterns was to list weight or space while refraining from mentioning any commonsense properties (Fisher exact test, $p = .002$).

There was a strong relation between weight patterns in the Styrofoam and Clay task and students' understanding of weight and volume measurement (see Table 12). Students with Hard Core Felt weight patterns had virtually no understanding of either weight or volume measurement while those with Weight as Property of Matter patterns had a strong understanding of both. Those with Transitional Felt Weight patterns were in between. 2 x 2 comparisons (Felt Weight vs. Property of Matter; Correct Measurement vs. Incorrect Measurement) revealed significant relations for both Weight (Fisher exact test, $p = .001$) and Volume Measurement (Fisher exact test, $p = .005$).

Insert Tables 10, 11, 12 about here

Finally, the pattern of correlations among our three posttest quantitative summary measures was very strong as well, consistent with the hypothesis that these understandings are interdependent and mutually constraining. The pattern of relations at the posttest essentially replicated the pattern of relations we observed at the time of the pretest (see Table 7). Students' posttest scores for Amount of Matter/Space and Weight/Volume Measurement each remained significantly correlated with their Weight scores, even when the effects of the other variable was statistically controlled. A step-wise regression analysis showed each variable made a distinctive and significant contribution to predicting Weight judgments, and that the two variables together accounted for over half the variance in Weight Judgments at the time of the posttest ($R = .75$, $R^2 = .56$, $p < .001$).

What Factors Predict Readiness for Change?

As mentioned earlier, half of the students with pretest Felt Weight patterns moved to Weight as Property of Matter patterns at the posttest (after the teaching); the other half remained with Felt Weight patterns. Could we predict which students would be most "ready" to re-conceptualize weight based on their pretest understandings? Although an earlier analysis showed that the two groups did not differ when performance on tasks was scored in a strictly categorical fashion, it is possible that some differences would emerge with the more continuous measures. These measures consider the total number of correct judgments about weight, the total number of correct judgments about divisibility of matter, amount of matter and space, and the number of correct measurements (summing over both weight and volume).

The two groups did not differ significantly on two of the pretest summary measures: Weight Judgments and Weight/Volume Measurement. However, the two groups did differ significantly on the third summary measure: Amount of Matter/Space Judgments. The group that progressed was correct on these items 68% of the time, whereas the group that did not progress was correct only 50% of the time (*t-test*, p =.05, 1-tailed).

In addition, the only pretest variable for these two groups that made a significant contribution to predicting their posttest performance was pretest Amount of Matter/Space Judgments. Indeed, this variable was more strongly correlated with posttest Weight Judgments (.53) and Weight/Volume Measurement (.40) than it was with itself (.37) ($p < .05$, 1-tailed).

Analyses of available worksheet and homework data for the two groups showed that they generally had similar levels of performance during the teaching sequence. Both groups gave evidence that they did not already believe that all matter had weight and judged some material objects to be weightless. However, they differed in their reactions to the lentil thought experiment in which they had to infer the weight of a single lentil using division. All of those who ended up concluding that even an invisible piece of Styrofoam must weigh something had understood how to use division to solve this thought experiment or a related follow-up problem. In contrast, the majority (55%) of those who maintained that Styrofoam weighs nothing had not used division on either one of these problems. This finding suggests that understanding the logic of this thought experiment may have contributed to changes in students' conception of weight.

Discussion

The purpose of this study was to investigate whether helping middle school students understand that matter takes up space and has weight requires restructuring a network of concepts. In discussing our findings, three main questions will be addressed: Why assume that students initially have different underlying concepts rather than simply mistaken beliefs? What specific processes may contribute to conceptual change in this case? What implications do these findings have for the design of effective science instruction?

The Case for Conceptual Change via Restructuring a Network of Concepts

The first issue is why assume students have fundamentally different concepts of weight and matter that are incommensurable with the concepts of scientists rather than simply more limited (or mistaken) beliefs. Is it not possible that the students with "felt weight" patterns have the same core concepts as the others, but simply lack some relevant knowledge or have some mistaken beliefs about what exact question was being asked? For example, some students may have thought we were asking them about the felt weight (rather than objective weight) in the Styrofoam task because we were placing the objects in their hands--a procedure that made felt weight salient. Or, some students might have the false belief that felt weight is a reliable indicator of weight because they lack knowledge of formal measurement procedures for weight. Still others may be unable to articulate the properties of matter, because of lack of formal schooling in these issues, but have an implicit understanding of matter as a fundamental constituent of objects. According to this view, formal schooling provides additional knowledge that helps students to elaborate on basic core understandings--not transform them.

There are four main ways to decide whether a given case involves deeper conceptual change (via restructuring) rather than knowledge enrichment over a constant conceptual core: (a) by analyzing the concepts involved and identifying the alternative theory in which they could be embedded; (b) by probing for consistency in a given aspect of conceptual understanding in multiple ways; (c) by examining if there are systematic relations among concepts in a way that fits with the conceptual analysis of what the alternative theory is like; and (d) by investigating student responses to direct instruction. Faulty beliefs, by definition, can be quite local and independent of each other. For example, one could lack knowledge of appropriate procedures for measuring weight without lacking knowledge of procedures for measuring volume or without lacking knowledge that all material objects have weight. Further, if one already has appropriate underlying concepts of weight, space, and matter, acquiring formal definitions of concepts or formal measurement procedures should be a relatively straightforward process. Concepts, in contrast, are by their nature inter-related and mutually constraining; hence, changing one concept involves making changes in others. Further, if one does not have an appropriate set of concepts for understanding a set of measurement procedures or taught definitions, there will be resistance to change and misunderstandings as students first try to assimilate the new knowledge to their initial set of ideas. Further, students who manage to improve in one task should make coordinated improvements in several.

The introduction has already provided the first type of evidence: a specific conceptual analysis of an alternative matter theory (MT1) in which students' initial concepts of size, weight, matter, and material kind could be embedded that is incommensurable to MT2. This section identifies the other three kinds of evidence by discussing the extent to which students had the kinds of within concept consistency, between concept coherence, and resistance to change that would be expected on a conceptual change via restructuring account.

Consistency in Reasoning About Weight Across Task Formats and Times of Testing

Students were consistent in their reasoning about weight when probed in an interview and in a different manner on a written test, suggesting there were reliable individual differences in how students' conceptualized weight. In both tasks, they were questioned about the weight of light things. In the interview, a small piece of Styrofoam was placed in their hands, an aspect of the procedure that would make felt weight more salient. In contrast, in the written test, students were simply asked to think about the weight of grains of sugar. Each task probed whether there would be a point at which students thought the piece no longer weighed anything at all (Felt Weight patterns), or whether they consistently maintained even the small pieces had some weight (Weight as Property of Matter patterns). At the time of both the pre and posttest, 90% of students responded in the same way to the question about whether the small piece of Styrofoam and 1 grain of sugar weighed anything.

In addition, both groups of students (those with Felt Weight and Weight as Property of Matter patterns) made their judgments confidently and provided clear justifications, consistent with their having committed beliefs. Those with Felt Weight patterns thought it was obvious some pieces weighed nothing, because they felt like they weighed nothing, and denied that a microscopic piece would have any weight as well. In contrast, those with Weight as Property of Matter patterns thought it was equally obvious they *must* weigh something because they were matter. They also continued to maintain that judgment with an invisible piece, again providing a principled reason for their judgment. Overall, one is struck by the fact that these two groups of students find opposite conclusions equally intuitive.

In earlier work, Carey (1991) found similar evidence for cross-task consistency concerning student conceptions of the homogeneity and continuity of matter among tasks that were even more different. In one task, students were asked whether a piece of Styrofoam took up any space at all for progressively smaller and smaller pieces. In the other, students were asked whether one could see "all the iron" in a piece as it was cut into progressively smaller pieces (to see if they realized there was always more iron inside). Students who thought you could see "all the iron" in the piece were generally same students who thought that tiny pieces would take up no space at all.

Of course, students with Felt Weight patterns did not *always* judge that small light objects weighed nothing at all, and in our categorization system we distinguished between those with "Hard Core Felt Weight" patterns and those with "Transitional Felt Weight" patterns based on how they responded to the Clay task (used in the interview). In this task, students were shown a small BB size piece added to a clay ball and asked if *adding* the piece changed the *amount of clay* in the ball and then asked if it also changed the *weight* of the clay ball. Given that most students thought that adding a piece increased the amount of clay in the ball, this task might be thought to prime for an answer that the weight also changed, especially if students (implicitly) connect weight with matter.

In fact some students with felt weight patterns on the Styrofoam and Sugar tasks confidently resisted this suggestion and, consistent with their other judgments about Styrofoam, asserted that we did *not* change the weight because the small piece was too small and weighed nothing as all. When asked if we could change the weight of the clay ball by adding a piece, they all agreed that we could, but indicated a much larger piece was needed to do so. We dubbed these patterns "Hard Core Felt Weight" patterns because these children were not only quite consistent across the Sytrofoam and Clay task in judging that a small BB size piece weighed nothing at all, but also thought even some larger pieces weighed nothing at all.

In contrast, those with Transitional Felt Weight patterns indicated that the small BB size clay piece *had* made the ball heavier because it weighed something, even though they also later judged one could get to a point where the Styrofoam or sugar didn't weigh anything at all. How should we interpret these Transitional Felt Weight patterns?

There are at least two different interpretations of these patterns, which cannot be definitely distinguished given our data, because we did not further probe students on whether they thought an even tinier piece of clay would also have some weight. One interpretation is that these children have *conflicting beliefs* about different materials and hence are being *inconsistent* in their reasoning. That is, although they believe pieces of some materials (such as clay) would always have some weight, they do not think this is the case for other materials (such as Styrofoam or sugar). An alternative interpretation is that these students have similar beliefs about clay and Styrofoam, and would have predicted that we would get to a point where the clay piece also weighed nothing at all. The main difference between those with Hard Core and Transitional Felt Weight patterns concerns how extreme a case they need to judge something weighs nothing at all. Thus, they are still operating from Commonsense Matter Theory 1 (in the sense that that they do not yet think of weight as an essential property of matter), but they are enriching it with further knowledge and crude generalizations about weight. This enrichment leads to some gradual weakening of the centrality of felt weight, without yet the formulation of a principled generalization that all matter must have some weight.

Several features of our results support the latter interpretation. First, consistent with this interpretation, those with Hard Core Felt Weight patterns were more likely to say that the cracker size piece of Styrofoam also weighed nothing at all than those with Transitional Felt Weight patterns, although both groups agreed that the small piece of Styrofoam weighed nothing at all. Second, there were no differences between the Hard Core Felt Weight and Transitional Felt Weight students on their grain of sugar judgments, another very extreme case. Third, analysis of these students' justifications on the Clay task indicated that they did not support their "yes" judgments with principled assertions (e.g., it *must* change the weight because all matter weighs something, or even because all clay must weigh something). Instead they made more local comments about the particular piece (e.g., the small piece weighs something; you added more clay to the ball).

Nonetheless, future studies could question students about the weight of tiny pieces of a broader array of materials (e.g., solids of a range of densities; powders, liquids, and even gases) to determine if students think that very tiny pieces of each of these materials weigh nothing at all. Such a study would not only provide a stronger test of the consistency of those with "Hard Core Felt Weight" patterns, but would also provide further insight about the meaning of these "Transitional Felt Weight" patterns. Children might think that small pieces of (dense) materials always have some weight before they extend this understanding to less dense solids (such as Styrofoam), powders (such as sugar), or even liquids. This pattern would be consistent with the recent proposal of Nakhleh & Samarapungavan (1999) that as children make a transition to a new way of thinking they first have more local frameworks that get extended, elaborated, and applied to a broader range of materials. Alternatively, children might make the switch more generally for a broader range of materials as long as they consider them to be prototypically matter (e.g., most solids, powders, liquids).

I suspect that the pattern will be closer to the latter, rather than one that is highly specific to particular materials, because in my experience teaching students about density, those who get the idea seem to generalize it readily to both materials in solid and liquid form and of a broad range of densities. At the same time, I would not expect that students would as readily generalize that all gases have weight for several reasons. First, gases are much less well understood by students, certainly not considered prototypically material, and are often not even classified as material at all. Second, gases pose an extreme case because students have no experience (or perceptual evidence) that even very large amounts of gases have weight, as is the case for solids and liquids. Thus, deciding how to think about gases will not come "for free" but will require further experiences and systematic investigations.

Coherence Among Concepts

The strong inter-task correlations and specific relations among qualitative patterns of responding on weight, occupied space, and matter tasks fit with the hypothesis that student concepts of weight, size, and matter are coherently organized in both MT1 and MT2. Students with Felt Weight patterns were the same students who denied that the Styrofoam took up space or had any amount of matter. In contrast, students who asserted that the Styrofoam must weigh something were also those students who maintained that the Styrofoam must take up space and have some amount of matter.

Of course, one might wonder whether these relations can be better explained because of very similar task structure and questions in these cases. Two arguments, however, caution against this "task specific" explanation. First, individual students with Hard Core or Transitional Felt Weight patterns rarely provided *identical* answers to the amount of matter, space, and weight questions even within the Styrofoam and clay tasks. In virtually all cases, they gave at least one discrepant judgment (for a given size item), and in all those cases the judgment was to assert that the item had some amount of matter (or took up some amount of space), but did not have any weight. This pattern of judgment indicates that students were making some differentiation among the questions and that students were always (slightly) more sophisticated in thinking about amount of matter/space than weight, in keeping with previous findings of Carey (1991) and Smith et al. (2005). This pattern of better judgments about amount of matter than weight held both for the clay and Styrofoam questions, even though the two tasks had different orders of questioning, again arguing against simple task or order effect explanations of the superiority. (In the clay task, the amount of clay questions were asked first; in Styrofoam, the weight questions were asked first.) At the same time, students with Felt Weight patterns were rarely perfectly correct across all amount of matter and space items. Thus, one needs to distinguish statistical superiority on a given question (which might reflect more incremental and elaborative processes of learning) from systematically correct patterns of response across a range of questions (which might reflect a stronger commitment to new principles).

Second, there were strong relations among students' performance on weight, space, and matter tasks even with tasks that probed student understandings of weight, volume, and matter in quite different ways (e.g., the Measurements task and the Matter Categorization and Judgment task used on the Written test).[7] If these relations were to be explained simply in terms of similarity of task format rather than underlying constraints among conceptual understandings, then one might not expect systematic relations between students' judgments about the weight of tiny pieces of matter and their patterns on these tasks. In fact, students with Felt Weight patterns were much more likely to focus on commonsense perceptual properties in characterizing matter in the Matter Categorization task and not to know how to make volume and weight measurements. In contrast, those with Weight as a Property of Matter patterns more often focused on objective properties of matter and were more knowledgeable about how to make formal weight and volume measurements. These systematic relationships were observed both at the time of pretests and posttest, with the contrasts being especially large between students with Hard Core Felt Weight patterns (virtually no understanding of either Weight or Volume measurement at either time of testing) and Weight as Property of Matter patterns. Individual pattern analyses also confirmed that students with both Hard Core and Transitional Felt Weight patterns at best showed insight on only one other task, while those with Weight as Property of Matter patterns showed insight on multiple tasks. This pattern is consistent with our assumption that children with Transitional Felt Weight patterns have begun to enrich and elaborate on MT1 understandings without yet restructuring them.

Analyses of justifications also revealed these ideas were richly inter-connected for both groups of students. However, because they had different conceptions of space and weight, they often arrived at very different conclusions. For example, one group of students argued that Styrofoam must weigh something because it's there, it takes up space, it exists, it's matter, or because it is part of the big piece and each part must weigh something. The other group argued that Styrofoam weighs nothing not only with arguments based on felt weight (e.g., it feels like nothing in my hand), but also with arguments based on perceived size (e.g., it is too small to weigh anything). Indeed, one student explicitly said "the tiny piece is only part of the big; it weighs nothing but a big bunch would weigh something." Note this student acknowledges the tiny piece is part of the whole, just as students do who think weight is a property of matter. However, this fact has entirely different significance, used to support an "emergent weight" rather than "additive weight" argument.

Of course, it is always possible that another (unmeasured) third variable may better account for these systematic relations. For example, some might wonder whether variations in students' mathematical understanding or general intellectual ability might explain these results, without resorting to claims about intuitive theories. Because we did not have data on these students' general intellectual abilities or mathematical understandings, we cannot rule out these competing explanations. Nonetheless, the fact remains that there do seem to be some important and systematic relations to be accounted for, and that the specific pattern of results is highly consistent with the kind of coherence predicted on our conceptual analyses. That is, our data consists not just of general task correlations, but also of the ways specific understandings of weight, size, matter, and measurement of these quantities are observed to co-occur.

Coordinated Patterns of Change

Consistent with the conceptual restructuring hypothesis, there was also evidence for coordinated patterns of change. Some students showed resistance to change in response to direct instruction. Not only did they maintain Felt Weight patterns, but they also failed to fully internalize that weight was a property of matter or the procedures for measuring weight. In contrast, students who changed their ways of conceptualizing weight (as evidenced in the Styrofoam/Clay task) made coordinated changes in other conceptions as well. The entire pattern of results, therefore, seems more consistent with the assumption that students have alternative ways of conceptualizing weight, space, and matter that need to be restructured than with the assumption that they simply lack specific pieces of knowledge, hold faulty beliefs, or have largely independent perceptual concepts. The latter account provides no explanation for the coherence among student answers both when they are "right" and "wrong", the difficulties students' experience in acquiring new understandings, and their coordinated patterns of change.

Learning Processes That Support Conceptual Change

What kind of learning processes may have supported conceptual change in this case? In my view, conceptual change is complex as it calls for multiple coordinated changes in student thinking and the introduction of some new, symbolic representational resources (from outside the domain) that contribute to the building of a new network of ideas as students grapple with understanding real world phenomena. The new representational resources give students a foothold for making inferences that would not have been licensed by their initial ideas, allowing them to apply their ideas to the world in novel ways. At the same time, they must initially be at least partially mapped to students' existing ideas to anchor them and give them meaning. Thus conceptual change only occurs as students explore possible mappings between their existing ideas and these new representational resources. The process is heuristic, not algorithmic as it is never clear ahead of time what aspects of the analogy (if any) will prove most fruitful or what adjustments need to be made (on both sides of the analogy) to making the mapping work and to provide a good explanation for observed phenomena. It is also aided by explicit metaconceptual discussion that identifies key assumptions that might need to be re-examined. No one change is sufficient by itself to create a new stable structure; multiple changes are needed to reinforce others. Yet each change provides part of the scaffolding for making new inferences that help elaborate the new structure, a dialectical process well-described as "bootstrapping." Consider how such conceptual bootstrapping may have worked in our teaching unit.

The first component of our teaching involved engaging students in an explicit metaconceptual discussion about the most essential properties of matter. This discussion helped students formulate explicitly, symbolically, and precisely (using the quantificational syntax of natural language) properties that might be essential characteristics of matter, each of which were listed on chart paper by the teacher who was recording the discussion. As multiple ideas were expressed in this discussion, students had an opportunity to consider proposals that might be different from their own. For example, although many students suggested that matter was something that you could see, feel, or touch, others suggested that matter was something that took up space or that both took up space and had weight. The teacher then encouraged students to test these proposals against each other by asking them to search for counter-examples: Could they think of something that was matter that they could not see, feel, or touch? Could they think of something that was matter that did not take up space or have weight? In other words, she encouraged them to consider whether any of these were *essential* properties of matter. Although the purpose of the opening discussion was to raise questions for students rather than present a new framework, it served to enlarge the number of potential properties of matter that students were considering. Indeed in their homework, many students began to add takes up space or has weight to their long list of commonsense perceptual properties, without yet considering them as more essential core properties. As Carey has argued (1999, 2004) these symbolically represented new relations may at first be only partially interpreted or understood, but will ultimately provide part of the scaffolding for extending and revising students' initial theory. As students gain a deeper understanding of how to measure volume and weight and imagine what happens to materials when ground into tiny invisible pieces, they are ultimately able to see that these properties are more essential than the perceptual properties that were central to their initial framework.

What may have made this aspect of instruction effective was that it engaged students in active debate with each other--with making arguments and counter-arguments about the essential properties of matter--rather than presenting a simple definition or conclusion. In fact, the debate that ensued was quite lively and contentious, with some students arguing that weight wasn't a property of all matter because of the existence of such counter-examples as dust, air, or smoke! This finding highlights the non-algorithmic nature of bootstrapping processes and the need for multiple coordinated change in order to fashion a new (more) stable structure. In particular, students needed additional reasons (and evidence) for believing that very small entities had some extent and new ways of conceptualizing occupied space and weight.

Thus, a second component of our curriculum involved extending students' conceptual understanding of occupied space and weight as extensive properties of matter by teaching students about the measurement of weight and volume. Learning to measure involves the construction of more powerful, symbolic ways of representing weight and volume, in terms of the specific number of units of each quantity, which transcends the more approximate analog magnitude representations of these quantities that are used in perceptually based representations. Indeed, a unit of measure is a new type of quantity--simultaneously a discrete and continuous quantity--which combines features from the domain of number with features from the domain of the specific dimension being measured. Because units of measure are discrete entities, they can be set in one-to-one correspondence with integer number tags and counted. At the same because each unit is a continuous quantity with a given size tied to the qualitative dimension that is measured, the units counted must all be of equal size and fill the measurement space in order for the measurement to be valid. Further because each unit itself has an extent, it can be further subdivided into fractional units.

Learning to measure not only provides students with additional ways of determining the weight and volume of objects that is more precise than their initial perceptually-based analog representations, but also gives them a way of analyzing the weight or volume of an object as the sum of its unit parts. In this way, an explicit network of additive relations is added to their concepts of weight and volume, which in combination with the conception of weight and volume as essential properties of matter, underscores the extensive nature of both quantities. Although students may initially favor the more familiar perceptually-based ways of conceiving of weight and volume, with time and experience they come to see the system defined by formal measurement as more reliable, valid, and central to their understanding of each physical quantity.

Learning about measurement also opens the door for discovering new conceptual relations (e.g., the weight of an object is directly proportional to its volume for any given kind of material; the weight of a piece is independent of the weight of surrounding pieces, etc.) and for extending their reasoning to very large and small pieces. Thus, the third component of our teaching involves students with both thought experiments and direct experiments that helped them confirm that entities with no felt weight nonetheless had some determinable objective weight. For example, after determining that 25 lentils had a mass of 1 gram, students were asked what they thought the mass of 1 lentil might be (something they could not determine directly, given the sensitivity of the scales they were using.) Many were able to reason that 1 lentil would have a mass of approximately 1/25 of a gram--thus, using their ideas about fractions and fractional units to reason about this new situation. This "thought experiment" gave them a situation where they had to reason deductively from measurement principles using a mathematical argument. It was followed with direct investigations using much more precise measuring instruments in order to confirm that very tiny things such as a piece of glitter or a dust ball did in fact have measurable mass and to give them further confidence in the inferences they had made based on their understanding of measurement principles.

Overall, our curriculum was effective in helping about half of the 8th graders who started with Felt Weight conceptions to make coordinated changes in their conceptions of matter, weight, and object size. The fact that these students made coordinated changes in multiple understandings is consistent with a restructuring account of how this conceptual change works. Similarly, the fact that students’ judgments about matter and space and understanding of measurement each made distinctive contributions to predicting students’ conceptions of weight is consistent with the fact that multiple representational resources are drawn on in establishing their new conception of weight (e.g., using recursive thinking to develop the notion of an underlying constituent, and using mathematical ideas to develop an explicit representation of weight as an additive, measurable quantity.) But why was the curriculum highly effective for only half of these students?

One characteristic of the bootstrapping processes outlined above is that they draw on ideas that are more clearly represented in one context to aid in the restructuring of ideas in another. For example, if students understand that things they can see, feel, and touch have some amount of matter, take up space, and have weight, then they can use this understanding to inform their thinking about pieces of matter too small to see, or to argue that things that have some amount of matter must also take up some amount of space or have some weight. If they have some idea of an iterative unit of measure from previous experiences with the measurement of length, they can elaborate on these ideas in learning about the measurement of area, volume, and weight (Lehrer et al., 2003). And if they have developed a concept of fractions as rational numbers, then they can use this refined notion of number to represent increasingly small quantities (Smith et al., 2005). Each of these building block ideas is, of course, a product of a prior learning and instructional history. When these anchoring ideas are not yet developed and clearly represented, however, our curriculum cannot work as intended to bootstrap change. One hypothesis is that the group that made coordinated changes had clearer anchoring intuitions to build upon than the group that made more piecemeal progress.

Several lines of evidence support this working hypothesis. First, among the group that was not already at ceiling on their pretest conceptions of weight, the only pretest understanding that significantly predicted posttest understandings of weight was their pretest understanding of amount of matter/space. More specifically, those who did not make progress made more initial errors on amount of matter/space judgments than those that did progress. Thus, although it was rare for students in either group to be fully correct in their initial amount of matter/space judgments, students who progressed more frequently judged that small objects had some amount of matter or took up some amount of space.

Second, there was evidence that those who progressed had a better understanding of the weight measurement thought experiment in which students had to infer the weight of one lentil from knowledge that the weight of 25 lentils weighed 1 gm. All the students who progressed were able to make the inference that one lentil would weigh only a fraction of a gram or solve a similar follow-up problem. In contrast the majority (55%) of students who did not progress had difficulty with this problem as well as a similar follow-up problem on the teacher's end of unit test. These difficulties may have stemmed from difficulties working with fractional numbers or with understanding the logic of weight measurement. Clearly, if they did not understand this thought experiment, they could not use it to reinforce their understanding of the additive structure underlying weight measurements.

Given the limitations of the curriculum intervention for this one group of students, one could ask whether *alternative* approaches to instruction might have proved more effective in helping them bootstrap their understanding. Recent work by Lehrer, et al. (2001) with a 5th grade class of students provides a wonderful example of a more elaborate (and much lengthier) route into teaching these topics that seems quite promising. Like us, they assumed that the underlying logic of measurement is not transparent to young children and that learning to measure physical quantities can play an important role in clarifying students' conceptual understanding of those quantities. Their approach to bootstrapping, however, focused first on building relevant *mathematical* understandings of measure and similarity prior to building explicit models of material kind in which the relevant *physical quantities* of weight, volume, and density are quantified and coordinated. Thus, the initial organizing framework they used to inform instruction was a more general *theory of measure* rather than a *theory of matter*. Further, children played a much more active role in *constructing* measures for different geometric and physical quantities, debating procedures, and discussing the logic of measurement than they did in our curriculum. In their view, premature introduction of standard units or measuring tools often hides these thorny conceptual issues; hence students come to use tools without deeply understanding their logic. They also seriously engaged students with discussing contrasting epistemological issues in mathematics and science, including an understanding of measurement error, issues which were not really considered in our curriculum.

Although they did not present pre and posttest data on individual children's understanding of the mathematics of measure and similarity or of children's understanding of the measurement of weight and volume, the data from the classroom discussions of these issues was quite impressive. Further, they did collect some posttest data probing individual children's ability to distinguish weight & volume and co-ordinate these quantities in a concept of density. Their data indicated that the majority (82%) could do this on at least in some tasks. The latter was an especially impressive result, considering they were working with much younger group of students (5th graders versus 8th graders).

Many aspects of their approach are consistent with the general analysis of conceptual change presented here. These aspects include: (a) the need for students to make coordinated changes in their thinking as part of building a new stable structure; (b) the role of mathematical ideas in helping students to bootstrap more sophisticated understanding of physical quantities; and (c) the need for much lengthier (and thus more thorough and explicit) metaconceptual discussion about both the nature of measurement and matter or materials than is provided in traditional curricula.

Implications for Instruction

One of the problems with many current science and math curricula is that they typically do not recognize the serious *conceptual difficulties* that learning about measurement and matter pose for children nor the complex kinds of learning processes involved in bootstrapping conceptual change. Instead, these curricula typically introduce students to new knowledge, procedures, and beliefs, without helping them develop the new concepts and conceptual frameworks needed to make sense of these procedures and beliefs. For example, it is assumed that the logic of measurement should be "obvious" to students; hence, that logic is neither problematized nor discussed. Further, children often learn about measurement as a disembodied set of procedures, without attention to developing the underlying concepts about matter, material kind, weight, and volume that would allow those procedures to make sense. Indeed learning about measurement is often considered part of the elementary school math rather than science curriculum, divorced from any discussion of the nature of matter or material kinds. The limitations in these approaches may explain why, despite repeated encounters with measurement in elementary grades, so few 8th graders in our study initially had any idea about how to measure the weight, mass, and volume of a simple cube.

Acknowledging that students have serious conceptual difficulties understanding matter, weight, and the logic of measurement does not imply that the problems are insurmountable or that we should delay when these topics are introduced. Indeed, the present study shows the great strides these same 8th graders made in understanding matter and measurement in a focused curriculum that addressed these issues. Similarly, others (e.g., Lehrer et al., 2001) have shown the remarkable progress that elementary school students can make in constructing a deep understanding of the measure of length, area, volume, and weight in even more sustained curricula that engages students in constructing an explicit theory of measure and models of material kinds.

One of the advantages of doing a cognitive analysis of the sources of conceptual change is that it helps to demystify the process and explain where the new ideas may come from. This analysis suggests that many of the component understandings that must be drawn on in developing MT2--understandings of natural language quantifiers, symbolic representations of number, some initial concepts of matter, weight, and taking up space, and a capacity for abstraction and recursive thinking--are already in place in the early elementary school years. In addition, other key understandings--such as the development of a concept of fractions as part of conceptual change in the domain of number--are central targets of instruction during the elementary school years. (See Smith et al., 2005, for an analysis of this important conceptual change as well as evidence that it is linked to conceptual change in the domain of matter.) These analyses suggest that conceptual changes within both domains may be linked--iterative mappings between number and physical quantity may not only transform children's understanding of physical quantity but of number itself--and that elementary school students are more able to engage in serious theory building about these issues than has commonly been realized. Thus a reasonable goal for elementary school math and science curricula might be for students to develop an explicit theory of matter (as something that occupies space and has mass) as well as of measure. These important understandings not only reinforce each other, but also provide a powerful foundation for later science learning (Smith et al., 2006).

At the same time the simple availability of these component ideas does not guarantee that they will be woven together to support new conceptual structures. That kind of learning calls for the sustained orchestration of classroom activities, such as active classroom dialogues and debates, search for examples and counter-examples, analogical reasoning and argumentation (including analogical mappings between the domains of number and matter), and thought experiments that help students abstract from everyday experience and represent relations not obvious in everyday experience. These kinds of activities were central both to our curriculum and those developed by Lehrer, Schauble, and their colleagues, but unfortunately are currently still not typical in most school science classrooms.

References

Au, K-F. (1994). Developing an intuitive understanding of substance kinds. *Cognitive Psychology*, *27*, 71-111.

Bovet, M., Domahidy-Dami, C., & Sinclair, H. (1982). Conservation and *decalages.* *Human Development, 25*, 201-14.

Carey, S. (1988) Conceptual differences between children and adults. Mind and Language, 3, 167-181.

Carey, S. (1991) Knowledge acquisition: Enrichment or conceptual change? In S. Carey and R. Gelman (Eds.) *The epigenesis of mind: Essays on biology and cognition* (pp. 257 – 291). Hillsdale, NJ: Erlbaum.

Carey, S. (1999) Sources of conceptual change. In Schollnick, E., Nelson, K., Gelman, S., & Miller, P. (Eds.) *Conceptual development: Piaget's legacy* (pp. 293-326). Mahwah, NJ: Lawrence Erlbaum Associates.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind and Language*, *16* (1), 37-55.

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus*, 59-68.

Chi, M. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R.N. Giere & H. Feigl (Eds.), *Minnesota Studies in the Philosophy of Science*, (pp.129-186). Minneapolis: University of Minnesota Press.

Chinn, Clark A. and Brewer, William F. (1993). The role of anomalous date in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational. Research 63*: 1-49.

Clement, J. (1989). Learning via model construction and criticism: Protocol evidence on sources of creativity in science. In J. Glover, R. Ronning, & C. Reynolds (Eds.), *Handbook of creativity: Assessment, theory, and research* (pp. 341-381). New York, Plenum.

Clement, J. (1991). Nonformal reasoning in science: The use of analogies, extreme cases, and physical intuition. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Church, R.M. & Broadbent, H.A. (1990). Alternative representations of time, number, and rate. *Cognition, 37*, 55-81.

Dehaene, S. (1997). *The Number Sense*. Oxford: Oxford University Press.

Dickinson, D.K. (1987). The development of material kind. *Science education, 71*, 615-628.

diSessa, A. and Sherin, Bruce. (1998). What changes in conceptual change*? International Journal of Science Education, 20*, 1155-1191.

Duckworth, E. (1986). *Inventing density*. Grand Forks: University of North Dakota Center for Teaching and Learning. North Dakota Study Group on Evaluation.

Gallistel, R. & Gerlman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43-74.

Gelman, R. (1991) Epigenetic foundations of knowledge structures: Initial and transcendent constructions. In S. Carey and R. Gelman (Eds.) *The epigenesis of mind: Essays on biology and cognition* (pp 292 – 322). Hillsdale, NJ: Erlbaum.

Gennaro, E.D. (1966). Teaching density--A real problem. *School Science and Mathematics,* 66, 559-560.

Gentner, D., Brem, S., Ferguson, R., Markman, A., Levidow, B., Wolff, P., & Forbus, K. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. The

Journal of the Learning Sciences, 6, 3-40.

Gopnik, A. & Meletzoff, A. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Hartnett, P. & Gelman, R. (1998) Early understandings of numbers: Paths or barriers to the new construction of new understandings? *Learning and Instruction, 8*, 341-374.

Hesse, J. and Anderson, C. (1992) Students' conceptions of chemical change. *Journal of Research in Science Teaching, 29*, 277-299.

Hewson, M.G. (1986). The acquisition of scientific knowledge: analysis and representation of student conceptions concerning density. *Science Education, 70*, 159-170.

Johnson, P. (1998). Children's understanding of changes of state involving the gas state, Part 2: Evaporation and condensation below boiling point. *International Journal of Science Education, 20*, 695-709

Johnson, P. (2000) Children's understanding of substances, Part 1: recognising chemical change. *International Journal of Science Education, 22*, 719-737.

Johnson, P. (2002) Children's understanding of substances, Part 2: explaining chemical change. *International Journal of Science Education, 24*, 1037-1054.

Keil, Frank (1999). Conceptual change. In R.A. Wilson and Frank Keil (Eds.) *The MIT Encyclopedia of the Cognitive Sciences* (pp. 172-182). Cambridge, MA: MIT Press.

Kitcher, P. (1988) The child as parent of the scientist. *Mind and language, 3*, 217-228.

Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change*. Chicago: University of Chicago Press.

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, 249-270.

Lehrer, R. (2003). Developing Understanding of Measurement. In J. Kilpatrick, W.G. Martin, and D. Schifter (Eds.) *A research companion to Principles and Standards for School Mathematics* (pp. 179-192). Reston, VA: The National Council of Teachers of Mathematics.

Lehrer, R., Jaslow, L. and Curtis, C. (2003). Developing understanding of measurement in the elementary grades. In D. Clements and G. Bright (Eds*.) Learning and teaching measurement: 2003 Yearbook* (pp. 100-121). Reston, VA: The National Council of Teachers of Mathematics.

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.

Lehrer, R., Strom, D., and Confrey, J. (2002). Grounding metaphors and inscriptional resonance: Children's emerging understanding of mathematical similarity. *Cognition and Instruction, 20,* 359-398.

Nakhleh, M. B. & Samarapungavan, A. (1999). Elementary school children’s beliefs about matter. *Journal of Research in Science Teaching, 36*, 777-805

Nakhleh, M. B. Samarapungavan, A., & Saglam, Y. (2005). Middle school children’s beliefs about matter. *Journal of Research in Science Teaching, 42*, 581-612

Nersessian, N.J. (1989). Conceptual change in science and in science education. *Synthese, 80,* 163-183.

Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R.N. Giere & H. Feigl (Eds.), *Minnesota studies in the philosophy* *of science* (pp. 3-44). Minneapolis: University of Minnesota Press.

Piaget, J. & Inhelder, B. (1967). *The child's conception of space*. New York: W.W. Norton & Company.

Piaget, J. and Inhelder, B. (1974). *The child's construction of quantities.* London: Routledge and Kegan Paul.

Quine, W. (1960). *Word and object*. Cambridge, MA: MIT Press.

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

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

Smith, C., Snir, J., & Grosslight, L. (1992). Using conceptual models to facilitate conceptual change: The case of weight/density differentiation. *Cognition and Instruction*, *9*, 221-83.

Smith, C., Solomon, G., and Carey, S. (2005). Never getting to zero: Elementary school students' understanding of the infinite divisibility of matter and number. *Cognitive Psychology, 51*, 101-140.

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

Snir, J., Smith, C., and Raz, G. (2003) Linking Phenomena with Competing Underlying Models: A Software Tool for Introducing Students to the Particulate Model of Matter. *Science Education, 87*, 794-830.

Spelke, E. (1991) Physical knowledge in infancy: Reflections on Piaget's theory. In S. Carey and R. Gelman (Eds.) *The epigenesis of mind: Essays on biology and cognition* (pp. 257-291). Hillsdale, NJ: Erlbaum.

Stavy, R. (1995) Conceptual development of basic ideas in chemistry. In S. Glynn & R. Duit (Eds.), *Learning science in the schools: Research reforming practice* (pp. 131-154). Hillsdale, NJ: Lawrence Erlbaum.

Thagard, P. (1992). Conceptual revolutions. Princeton, NJ: Princeton University Press.

Wellman, H. & Gelman, S. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, *43*, 337-375.

Wiser, M. & Carey, S. (1983) When heat and temperature were one. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 267-297). Hillsdale, NJ: Lawrence Erlbaum Associates.

Yair, Y. and Yair, Y. (2004). "Everything comes to an end": An intuitive rule in physics and mathematics. *Wiley InterScience* (www.interscience.wiley.com), 594-609.

Author Note

This study was funded by National Science Foundation Grant MDR-8955381 awarded to Carol Smith and Susan Carey. I would like to thank Deborah Maclin, Carolyn Houghton, Helen Davis and Lorraine Grosslight who as Research Assistants helped in many aspects of this project. I also want to thank the students who took part in our study and their teacher, Sandra Smith, who so ably collaborated on the project and taught the innovative unit.

Address for Correspondence: Carol L. Smith, Department of Psychology, University of Massachusetts at Boston, Boston, MA 02125. Phone: 617-287-6359. Fax: 617-287-6336. Email address: Carol.Smith@umb.edu

Footnotes

[1] The focus here is on the development of children's understanding of the physical properties of matter and materials (e.g., weight, volume, density), not with their understandings of its chemical properties (e.g., chemical composition; explanations of macroscopic material properties such as malleability or fluidity). Neither MT1 or MT2 provides explanations of these chemical properties other than as (potentially) intrinsic characteristics of different materials or has resources that allow students to represent the distinction between physical and chemical change. In contrast, the atomic-molecular theory of matter is of much broader scope. It not only has resources that allow students to represent the distinction between physical and chemical change (Johnson, 2000, 2002), but also provides them with deeper explanations of both the physical and chemical properties of matter. In recent work (Nakhleh & Samarapungavan, 1999; Nakhleh, Samarapungavan, & Saglam, 2005), children's ideas about the underlying structure of particular materials were probed with attention to the issue of whether they might have intuitive and consistent macroscopic frameworks for thinking about these issues prior to learning about the atomic-molecular theory. (For example, did they consistently conceptualize materials as being made of one continuous substance or as being made of macroscopic bits and pieces of different shapes and sizes?) They were also interested in whether students' used these macro-continuous or macro-particulate ideas about the structure of materials in their explanations for macroscopic properties such as malleability, rigidity, fluidity and for processes like melting and dissolving. In general, they found that most young children's beliefs about the underlying structure of materials were consistently macroscopic although not necessarily consistently macro-continuous or macro-particulate. Thus they argued they did not "exhibit the levels of theoretical coherence that characterize explanatory frameworks." (Nakhleh & Samarapungavan,

1999, p. 800). Further, children typically could not provide deeper explanations of properties such as malleability or fluidity. Instead they treated them as intrinsic self-explanatory properties. Many aspects of their results are consistent with the claim investigated here: namely, that children have some initial (physical) matter theories of limited explanatory scope in which the (macroscopic) notions of matter and material kind themselves play an explanatory role.

[2] The development of an understanding of conservation of matter across many different kinds of transformation (including phase changes and chemical change) is a long, protracted process of theory building, not a simple matter of logic. Students need to understand what happens in a given type of transformation in physical or chemical terms, and based on that understanding, available evidence and relevant argument, come to conclusions about whether matter is conserved. Thus there would be no reason to assume that children would acquire some general belief in conservation of matter "across the board". To make the transition to Commonsense Matter Theory 2, they need to understand the conservation of matter across subdivision and spatial re-arrangement, as those are central to conceiving of amount of matter as an additive quantity (i.e., one that can be decomposed and recomposed). Understanding that matter is conserved across phase and chemical change is obviously more challenging, and would be greatly aided by theoretical knowledge of the atomic-molecular theory of matter (Johnson, 1998, 2002) as well as their ability to make sensitive measurements. Given the great difficulties students have in understanding atomic-molecular frameworks, it is not surprising that many high school students have not worked out a good understanding of conservation of matter in these situations (Hesse & Anderson, 1992; Stavy, 1995).

[3] Although Au (1994) concedes that there is developmental change in children's understanding of material kinds, she also argues that young children have more competence in understanding materials as constituents (in some limited contexts) that is drawn on in a "bootstrapping" process. We agree with many aspects of her analysis and note that MT1 depends on children's making a principled distinction between objects and materials. Her work is different from ours, however, in focusing on young children's concept of specific material kinds rather than matter itself and in not distinguishing between central and less central properties of materials. Further, she only presents data about whether individuals are above chance in responding across a diverse range of item types (which include items we would predict they should have no difficulty with) rather than whether they are systematically correct across all items. The fact that they increasingly make errors as grain size becomes smaller is consistent with their having an alternative way of thinking about the core properties of materials.

[4] See Lehrer (2003) and Lehrer et al. (2003) for a detailed conceptual analysis of what is involved in understanding the logic of measurement that goes well beyond the account given here, along with descriptions of types of instructional approaches that help elementary children construct these understandings. They make a strong case that the simple teaching of measurement *procedures* does not automatically translate into *conceptual understanding* of measurement, and in some ways can obscure aspects of the logic. In their work, it is central not only to involve students in metaconceptual discussion about measurement, but also in the initial construction of a unit prior to presenting more standard units for a given quantity.

[5] The students were given the Weight/Density Interview task battery described more completely in Smith et al. (1997) which involved the manipulation and exploration of a variety of materials. The Clay tasks came about half way through the interview after students had

ordered objects by their size, weight, and density and created visual models designed to give information about the objects size, weight, and density. The Styrofoam tasks came at the very end of the interview, and were preceded by questions about Thermal Expansion of alcohol and sinking and floating of objects in different liquids. Thus, the Styrofoam questions did not immediately follow the Clay task.

[6] This portion of their middle school curriculum was concerned with simply introducing the scientist's notion of mass as a measure of amount of matter rather than with distinguishing between the scientist's notions of mass and weight or with providing distinct ways of quantifying these two quantities. Our pretest data indicated that many students thought mass was more related to the amount of space an object occupied than to an object's weight. Thus, in the teaching unit, students learned that mass is related to an object's weight, learned to measure mass using balance scales, and determined that even small object's had measurable mass.

[7] Our interview also included tasks that probed for students' differentiation of weight and density in multiple ways as the full teaching unit was designed to facilitate weight/density differentiation as well. Consistent with prior results (Smith et al., 1997), there were strong relations between students' patterns of responses on the Styrofoam/clay tasks and on tasks that assessed weight/density differentiation. In particular, none of the students with Hard Core Felt Weight patterns differentiated weight and density on any of our tasks; in contrast, the majority of those with Weight as a Property of Matter patterns clearly differentiated weight and density on at least some tasks. Because of the complexity of the density unit and the issues involved in that part of the study, those findings will be presented in a separate paper.

Table 1

Contrasting Concepts in Matter Theory 1 and 2

Concept	Matter Theory 1: Core Properties of Concepts	Concept	Matter Theory 2: Core Properties of Concepts
Matter	Objective stuff; Perceptually accessible (can see, feel, touch)	Matter	Fundamental constituent that is conserved (not created/destroyed); Takes up space; Has weight
Air/nothing	Penetrable; empty space	Air	Gaseous phase of matter
		Vacuum	Empty space
Material Kind	Characteristic perceptual properties: color, texture, smell, taste, capability of burning	Material Kind	Underlying properties (e.g., density) more central than surface appearance (e.g., color)
Size	Perceived spatial extent Undifferentiated global bigness Analog magnitude representation	Occupied Space	Extensive quantity; Property of matter Measured as 3-dimensional quantity: cm^3 is unit Additive & continuous quantity, mapped to number
		Length	1-dimensional quantity: cm is unit Additive & continuous quantity, mapped to number
		Area	2-dimensional quantity: cm^2 is unit Additive & continuous quantity, mapped to number
Weight/ Density	Felt weight (conflates extensive & intensive senses of weight: heavy and heavy for size) Analog magnitude representation	Mass/ Weight	Extensive quantity; property of matter Measured on balance scales; gm is unit of weight Additive & continuous quantity, mapped to number
		Density	Intensive property of material kinds (heavy for size) Quantified as weight (gm) per unit volume

Table 2

Overview of Tasks used to Assess Different Aspects of Concepts

Concept:	Aspect probed	Task	Specific Questions
Weight	Felt weight vs. property of matter (continuous quantity)	Styrofoam: Weight (Interview item)	Does this weigh a lot, a tiny bit, or nothing at all? (for a cracker-size and a BB-size piece) (For a piece too small to see) Do it have any weight?
	Felt weight vs. property of matter (continuous quantity)	Clay : Weight (Interview item)	Does the clay ball weigh more, less or the same? (after adding a BB size piece of clay to the ball)
	Felt weight vs. property of matter (continuous quantity)	Sugar Task (Written test item)	Does 10 grains/1grain of sugar weigh a lot, a tiny bit or nothing at all?
	Sum of gram units in balance on scale (additive quantity)	Weight measurement (Written test item)	How much does this object weigh? (asked of object in balance with 9 gram pieces, see Figure 1)
Size	Global bigness vs. property of matter (continuous quantity)	Styrofoam: Space (Interview)	Can there be a piece of Styrofoam too small to see? Does it take up any space?
	Sum of cm3 units or L x W x H (additive quantity)	Volume Measurement (Written test item)	What is the volume of this object? (asked of an object that was divided into cubes, see Figure 1)
Matter	Amount of matter: not clearly defined vs. additive & continuous quantity	Styrofoam: Amount of Matter (Interview item)	Does this has a lot, a tiny bit, or no matter at all? (asked of cracker and a BB-size piece)
	Amount of matter: not clearly defined vs. additive & continuous quantity	Clay: Amount of Matter (Interview item)	Does the clay ball have more, less, or the same amount of clay as before? (after adding a BB size piece)
	Matter: not defined at micro level vs. Infinitely divisible underlying constituent	Styrofoam: Thought Experiment (Interview item)	If we repeatedly cut this Styrofoam in half, would we ever reach a point where there was no more matter left to divide?
	Entities that are matter (e.g., Solid, liquid, powder, gas)	Matter Judgments (Written Test item)	Is this matter/not matter? (rock, table, dog, tree, grain of sugar, particle of chalk, water, air, heat, echo, shadow, wish, dream)
	Properties of matter (e.g., see, feel, touch vs. takes up space, has weight)	Matter Justifications (Written Test item)	What are the properties of matter?

Table 3

Relation Between Weight Judgments on the Styrofoam and the Sugar Tasks (Pretest)

	Sugar Task	
Styrofoam Task	1 Grain Weighs nothing at all	1 Grain Weighs something
Small or invisible piece weighs nothing	20	3
All matter has weight	1	18

Note: N = 42.

Table 4

Relations Among Weight Judgment Patterns and Matter/Space Judgments (Pretest)

		Matter/Space Judgments			
Weight Pattern	n	% who judge Even invisible Styrofoam Takes up space	% who judge Visible Styrofoam/clay has some matter	% who judge Continued existence of matter with repeated division	% with all three understandings
Hard Core Felt Weight	11	28 %	36 %	36 %	18 %
Transitional Felt Weight	12	50 %	50 %	33 %	17 %
Property of Matter	19	100 %	95 %	79 %	79 %

Table 5

Relation Between Weight Judgment Pattern and Matter Justifications (Pretest)

	Matter Justification (Matter Categorization Task)	
Weight Pattern (Styrofoam/Clay)	See, Feel, Touch, Blank or Other	Made of Material, Solid/Liquid/Gas Exists, Takes up space, Has Weight (without see, feel, touch)
Felt Weight	20	3
Property of Matter	3	16

Note: N = 42.

Table 6

Relation Between Weight Pattern and Understanding of Weight and Volume Measurement (Pretest)

		Measurement Understanding		
Weight Pattern (Styrofoam/Clay)	n	% Correct Weight Measurement	% Correct Volume Measurement	% Both Correct Measurements
Hard Core Felt Weight	11	9 %	0 %	0 %
Transitional Felt Weight	12	42 %	8 %	8 %
Property of Matter	19	58 %	42 %	42 %

Table 7

Full and Partial[a] Correlations Between Three Summary Measures at Pre and Posttest (Pearson)

Measures Correlated	Pretest Measures		Posttest Measures	
	Full	Partial	Full	Partial
Matter/Space & Weight	.62**	.51**	.69**	.52**
Weight & Measurement	.54**	.39*	.63**	.38*
Matter/Space & Measurement	.43*	.14	.58**	.26

****** significant at or beyond .001 level, 1-tailed
* significant at .01 level, 1-tailed

[a] Partial correlations are computed with the effects of the third measure controlled

Table 8

Changes in Weight Patterns (Styrofoam/Clay task) of Individuals from Pretest to Posttest

	Post Weight Pattern		
Pre Weight Pattern	Hard Core Felt Weight	Transitional Felt Weight	Property of Matter
Hard Core Felt Weight	3	4	4
Transitional Felt Weight	1	3	8
Property of Matter	0	0	19

Note: N = 42

Table 9

Coherency of Change: Relation Between Pre to Posttest Change on Styrofoam Weight Task and Number of Other Tasks on Which Individuals Change to MT2 Patterns

	Pattern of Change: Other Tasks[a]			
	Number of Other Tasks Change to MT2 Patterns			
Pattern of Change: Styrofoam Weight	0	1	2	3
Maintain Felt Weight (n = 11)	8	3	-	-
Change to Property of Matter (n = 12)	-	2	8	2

[a] The other tasks were: (a) Stryrofoam Space & Amount of Matter; (b) Weight Measurement; and (c) Matter Categorization Task Justifications.

Table 10

Relations Among Weight Patterns and Matter/Space Judgments (Posttest)

		Matter/Space Judgments			
Weight Pattern	n	% who judge Even invisible Styrofoam Takes up space	% who judge Visible Styrofoam/clay has some matter	% who judge Continued existence of matter with repeated division	% with All three understandings
Felt Weight	11	18 %	55 %	36 %	9 %
Property of Matter	31	100 %	100 %	84 %	84 %

Note: In this Table the Hard Core and Transitional were collapsed because there were no differences between these two groups in percentage correct on Matter/Space Tasks.

Table 11

Relation Between Weight Pattern and Matter Justifications (Posttest)

	Matter Justification (Matter Categorization Task)	
Weight Pattern (Styrofoam/Clay)	See, Feel, Touch, Blank or Other	Takes up Space or Has Weight (without see, feel, or touch)
Felt Weight	8	3
Property of Matter	6	25

<u>Note</u>: N = 42

Table 12

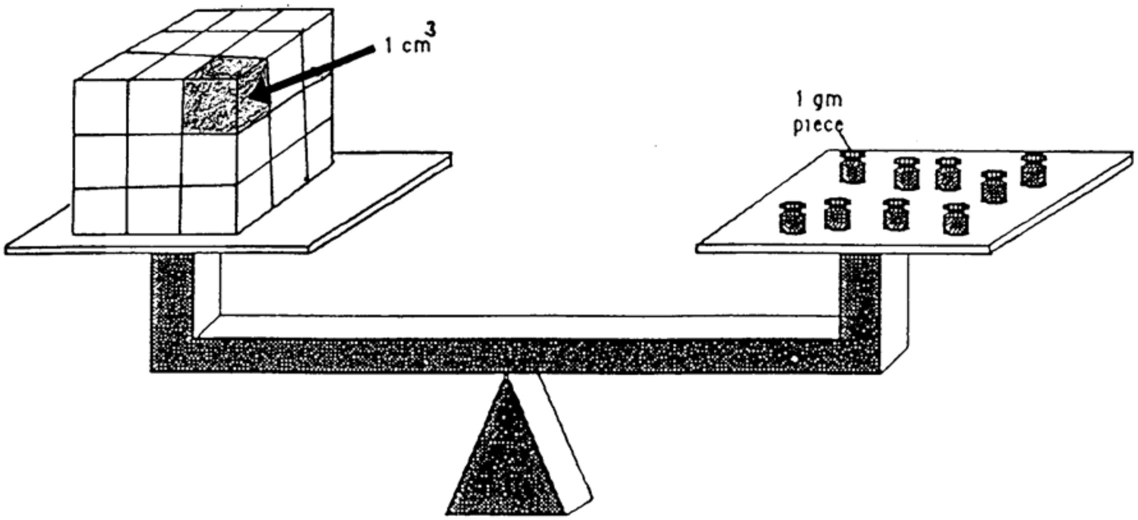
Relation Between Weight Pattern and Understanding of Weight and Volume Measurement (Posttest)

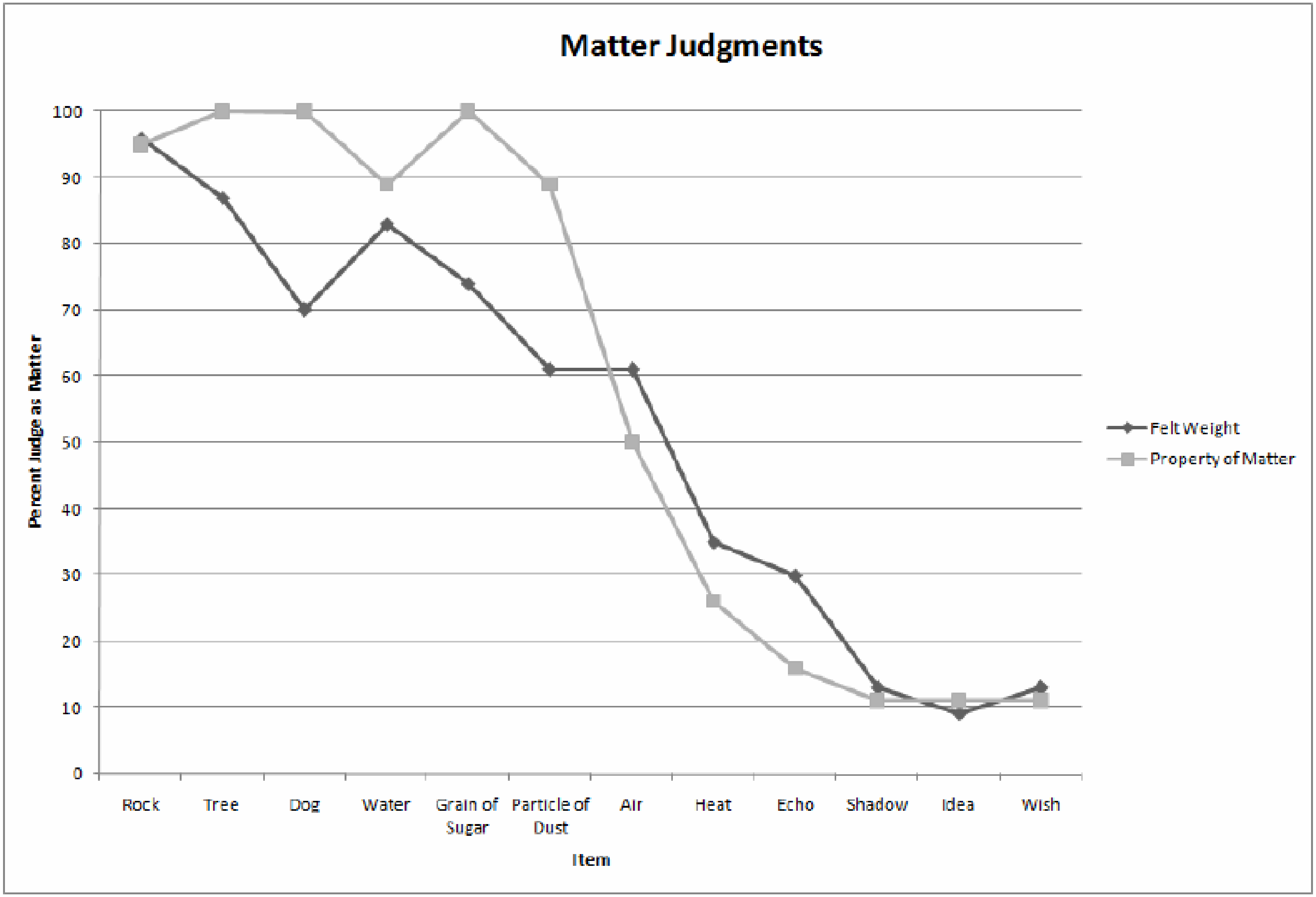
		Measurement Understanding		
Weight Pattern (Styrofoam/Clay)	n	% Correct Weight Measurement	% Correct Volume Measurement	Both Correct Measurements
Hard Core Felt Weight	4	25%	0%	0%
Transitional Felt Weight	7	43%	43%	29%
Property of Matter	31	90%	77%	71%

Figure Captions

Figure 1. The Measurement task included the following questions about the diagram (above). Here is an object in balance on a scale. Please answer the following questions with appropriate numbers and units. (a) What is the volume of the object? How can you tell? (b) What is the weight of the object? How can you tell? (c) What is the mass of the object? How can you tell?

Figure 2. Percent of students who judge an item to be matter (Matter Categorization task) for students with Felt Weight and Property of Matter patterns (Styrofoam/Clay task)


1 cm^3
1 gm
piece


Matter Judgments
Percent Judge as Matter
100
90
80
70
60
50
40
30
20
10
0
Rock
Tree
Dog
Water
Grain of Sugar
Particle of Dust
Air
Heat
Echo
Shadow
Idea
Wish
Item
Felt Weight
Property of Matter