

# The Inquiry Project: Final Report

Submitted to the National Science Foundation  
DR K-12 Program

Sue Doubler, TERC  
David Carraher, TERC  
Roger Tobin, Tufts University  
Jodi Asbell-Clarke, TERC

With substantial contributions by

Carol L. Smith, UMass-Boston  
Analúcia D. Schliemann, Tufts University

Sept. 8, 2011  
TERC, Cambridge, MA

The Inquiry Project: Final Report.....	1
Summary .....	5
I. Education and Research Activities .....	7
A. Education Activities in Grades 3-5 .....	8
1. Development of the Inquiry curriculum .....	8
2. Development of Formative Assessments.....	11
4. Professional development with teachers.....	12
B. Research Activities.....	12
1. Research overview .....	12
2. Methods of Coding & Analysis .....	14
II. Findings.....	15
A. Understanding Matter and Material Things.....	15
1-Identifying Materials: Is sawdust wood? Will it burn? .....	15
2-Invariance of Amount of Material and Weight When Shapes Change .....	18
3- Properties of Tiny (Visible and Invisible) Things.....	20
4-Sugar dissolving in water .....	23
5-Solids, Liquids, and Gases.....	24
6-Summary of Results: Matter and material things .....	28
B. Understanding Volume .....	28
1-Block Rearrangement task.....	31
2-Clay sphere vs. pancake task .....	34
3-The Sizing-Up (Volume Invariant) Task.....	37
4-Water Displacement Task.....	40
5-Area Task.....	42
6-Summary of volume results.....	44
C. Understanding Weight, Size, and “Heaviness” of Materials .....	45
1-Larger is not always heavier .....	45
2-“Heavier” vs. “Made from Heavier Material” .....	49
3-Infering materials from weight and size.....	53
4-Summary of Results for Weight, Size, and Heaviness of Materials.....	58
D. Granularity of Weight, Length, and Number .....	59
1-Overall Granularity Results .....	61
2- Granularity By Grade Level .....	62
3. The Repeated Halving of Numbers.....	66
4- Summary of results on granularity and repeated halving of numbers.....	69
E. Relating scientific concepts and quantification of components of a mixture (sugar and water).....	71
1-Same Amount of Sugar, Different Amounts of Water .....	73
2-Amount of Sugar Proportional to Amount of Water .....	74
3-Different Ratios .....	75
4-Different Simplified Ratios .....	77
5-Summary of Results.....	78
F. Conclusions.....	78
III. References & Reports .....	82

## TABLES

TABLE 1: NUMBER OF TREATMENT AND CONTROL STUDENTS INTERVIEWED.....	13
TABLE 2: RETENTION STATISTICS FOR THE TREATMENT AND CONTROL GROUPS.....	13
TABLE 3: INVARIANCE ACROSS GRINDING (WOOD, IRON) & MELTING (BUTTER): % OF CHILDREN SAYING YES FOR ALL THREE SUBSTANCES.....	16
TABLE 4: INVARIANCE ACROSS GRINDING: "SAWDUST BURNS" AND "IRON FILINGS ARE ATTRACTED TO A MAGNET".....	17
TABLE 5: AMOUNT OF CLAY, PLASTIC BOTH VIEWED AS INVARIANT ACROSS SHAPE CHANGE.....	19
TABLE 6: AMOUNT, WEIGHT, AND BALANCE AS INVARIANT ACROSS SHAPE CHANGE FOR CLAY, PLASTIC (PATTERN ACROSS 6 ITEMS).....	20
TABLE 7: "TINY (VISIBLE) PIECES TAKE UP SPACE AND HAVE WEIGHT" (2 ITEMS).....	21
TABLE 8: "TINY (INVISIBLE) PIECES TAKE UP SPACE AND HAVE WEIGHT" (2 ITEMS).....	22
TABLE 9: UNIVERSAL GENERALIZATION GIVEN AT LEAST ONE WEIGHT JUDGMENT.....	23
TABLE 10: DISSOLVED & UNDISSOLVED SUGAR: PERCENT WHO JUDGE BOTH SAME AMOUNT AND SAME WEIGHT...24	24
TABLE 11: PERCENT WITH DIFFERENT MATTER CLASSIFICATIONS.....	26
TABLE 12: MATTER JUSTIFICATIONS: COMMON PROPERTIES OR CHARACTERISTICS.....	27
TABLE 13: ON ALL VOLUME TASKS EXCEPT BLOCK ARRANGEMENT, THE TREATMENT GROUP SIGNIFICANTLY OUTPERFORMED THE CONTROL GROUP IN GRADES 3-5.....	31
TABLE 14: FOUR COMPARISONS USED TO DETERMINE HOW STUDENTS CONTENTED WITH WEIGHT-SIZE 'ANOMALIES' (PROBLEMS 2-4).....	47
TABLE 15: PERCENT OF CHILDREN WHO USE DIFFERENCE IN MATERIAL/HEAVINESS OF MATERIAL TO EXPLAIN WHY WEIGHT & SIZE DO NOT COVARY IN PROBLEMS 2 TO 4.....	48
TABLE 16: PERCENT OF CHILDREN WHO SYSTEMATICALLY DISTINGUISH HEAVINESS OF KIND OF MATERIAL FROM HEAVINESS OF OBJECT (ALL THREE ITEMS) (N=347, COMPLETE SAMPLE).....	53
TABLE 17: GRANULARITY OF NUMBER, WEIGHT AND LENGTH (ALL GRADES AND GROUPS).....	62
TABLE 18: ASSOCIATION BETWEEN ATTRIBUTION OF WEIGHT TO A TINY, INVISIBLE SPECK AND GRANULARITY OF WEIGHT.....	65
TABLE 19: GRANULARITY OF NUMBER, WEIGHT, AND LENGTH. REPRESENTATIVE VALUES OF "SOME". (ALL GRADES AND GROUPS).....	66
TABLE 20: MIXTURE TASKS, SHOWING QUESTIONS AND ALSO DIAGRAMS SHOWN TO STUDENTS.....	72

## FIGURES

FIGURE 1: BLOCK REARRANGEMENT TASK STIMULI.....	31
FIGURE 2: BLOCK REARRANGEMENT TASK (TREATMENT GROUP): DOES ONE ARRANGEMENT OF BLOCKS TAKES UP MORE SPACE?.....	33
FIGURE 3: BLOCK REARRANGEMENT TASK (CONTROL GROUP): DOES ONE ARRANGEMENT OF BLOCKS TAKES UP MORE SPACE?.....	34
FIGURE 4: CONSERVATION OF CLAY--"DO THESE OBJECTS TAKE UP THE SAME AMOUNT OF SPACE?".....	34
FIGURE 5: VOLUME OF CLAY TASK (TREATMENT GROUP).....	35
FIGURE 6: VOLUME OF CLAY TASK (CONTROL GROUP).....	36
FIGURE 7: VOLUME JUDGMENT--WHICH OF THE ABOVE BLOCKS IS BIGGER? THE PURPLE, OBLONG BLOCK MEASURES 2"x8"x1" (16 CU. IN.); THE WOODEN BLOCK MEASURES 2"x3"x3" (18 CU. IN.). STUDENTS WERE NOT TOLD THESE MEASURES. HOWEVER THEY JUDGED THE SIZES USING A SET OF SMALL CUBES (APPROXIMATELY 1" ON EACH SIDE).....	37
FIGURE 8: INVARIANTS USED IN JUDGMENTS OF SIZE OF OBJECTS (TREATMENT GROUP).....	39
FIGURE 9: INVARIANTS USED IN JUDGMENTS OF SIZE OF OBJECTS (CONTROL GROUP).....	39
FIGURE 10: PREDICTIONS AS TO WHETHER BRASS AND ALUMINUM CYLINDERS WOULD MAKE THE WATER LEVEL RISE TO THE SAME OR DIFFERENT LEVELS (TREATMENT GROUP).....	41
FIGURE 11: PREDICTIONS AS TO WHETHER BRASS AND ALUMINUM CYLINDERS WOULD MAKE THE WATER LEVEL RISE TO THE SAME OR DIFFERENT LEVELS (CONTROL GROUP).....	42
FIGURE 12: INVARIANTS USED IN JUDGING SIZE OF CARDS (TREATMENT GROUP).....	43

FIGURE 13: INVARIANTS USED IN JUDGING SIZE OF CARDS (CONTROL GROUP) .....	44
FIGURE 14: OBJECTS USED IN "HEAVIER MATERIAL TASK" .....	50
FIGURE 15: CRITICAL COMPARISON #1. "IS ONE OF THESE IS MADE OF THE HEAVIER MATERIAL OR NOT?" <u>NOTE: H AND I ARE SMALL DISCS OF ALUMINUM AND DELRIN, RESPECTIVELY; THEY ARE TOO LIGHT TO ANSWER THE QUESTION WITH CONFIDENCE BASED ON FELT WEIGHT.</u> .....	51
FIGURE 16: CRITICAL COMPARISON #2. "IS ONE OF THESE IS MADE OF HEAVIER MATERIAL OR NOT?" <u>NOTE: THE TINY DISC, J, IS BRASS; THE LARGE CYLINDER IS ALUMINUM.</u> .....	51
FIGURE 17: CRITICAL COMPARISON #3. "IS ONE OF THESE MADE OF THE HEAVIER MATERIAL OR NOT?" <u>NOTE: BOTH OBJECTS ARE ALUMINUM.</u> .....	52
FIGURE 18: OBJECTS FROM THE "INFERRING MATERIALS FROM WEIGHT AND SIZE TASK". <u>NOTE: THE THREE A CYLINDERS ARE MADE FROM BRASS; THE C CYLINDERS, AS WELL AS B, ARE ALUMINUM; D IS MADE OF A PLASTIC HAVING A DENSITY DIFFERENT FROM ALUMINUM, AND BRASS, AND DELRIN (G).</u> .....	54
FIGURE 19: CYLINDER B CORRECTLY JUDGED TO BE MADE OF ALUMINUM .....	55
FIGURE 20: CYLINDER D CORRECTLY JUDGED TO BE MADE OF ALUMINUM .....	56
FIGURE 21: CYLINDER A CORRECTLY JUDGED TO BE MADE OF BRASS .....	57
FIGURE 22: CYLINDER C CORRECTLY JUDGED TO BE MADE OF ALUMINUM (TREATMENT VS. CONTROL).....	58
FIGURE 23: GRANULARITY OF NUMBER--HOW MANY NUMBERS ARE THERE BETWEEN 4 AND 7? (4 AND 5?) .....	59
FIGURE 24: STIMULI FOR THE GRANULARITY OF WEIGHT TASK: "HOW MANY WEIGHTS DO YOU THINK COULD BE BETWEEN THAT OF THE RED BALL AND THE GREEN BALL?" .....	60
FIGURE 25: GRANULARITY OF LENGTH STIMULI. "HOW MANY LENGTHS COULD BE BETWEEN THAT OF LINE A AND THAT OF LINE B?" .....	61
FIGURE 26: GRANULARITY OF NUMBER BY GRADE LEVEL .....	63
FIGURE 27: GRANULARITY OF WEIGHT BY GRADE LEVEL.....	64
FIGURE 28: GRANULARITY OF LENGTH BY GRADE LEVEL.....	65
FIGURE 29: REPEATED HALVING QUESTIONING AND RELATION TO RESPONSE CATEGORIES .....	67
FIGURE 30: REPEATED HALVING OF NUMBER (TREATMENT GROUP).....	68
FIGURE 31: REPEATED HALVING OF NUMBER (CONTROL GROUP).....	69
FIGURE 32: (2 CUBES IN 4 CUPS) SWEETER THAN (2 CUBES IN 6 CUPS).....	73
FIGURE 33: (2 CUBES IN 6 CUPS) CLAIMED SWEETER THAN (2 CUBES IN 4 CUPS) BECAUSE "THERE IS MORE WATER" .....	74
FIGURE 34: (1 CUBE IN 3 CUPS) JUST AS SWEET AS (2 CUBES IN 6 CUPS) .....	75
FIGURE 35: (3 CUBES IN 8 CUPS) LESS SWEET THAN (2 CUBES IN 4 CUPS) .....	76
FIGURE 36: MAKING TWO MIXTURES EQUALLY SWEET (CONTROL) .....	77
FIGURE 37: MAKING TWO MIXTURES EQUALLY SWEET (TREATMENT) .....	77
FIGURE 38: (2 CUBES IN 3 CUPS) SWEETER THAN (1 CUBE IN 2 CUPS) .....	78

---

### Summary

TERC, Tufts University, and two schools in Greater Boston collaborated through the Inquiry Project to investigate the development of students' scientific understanding from 8 to 11 years of age (Grades 3-5), particularly with respect to their evolving concepts of material, weight, volume, density, and state of matter. As a curriculum development and teacher education initiative, it has tested ideas about the teaching and learning of these concepts. As a research endeavor, it has helped clarify how these concepts develop over time, how they can be successfully nurtured through instruction, and how they prepare students for later learning about the atomic-molecular theory of matter.

The project drew inspiration from a learning progression about matter and the atomic-molecular theory proposed by Smith, Wiser, Anderson and Krajcik (2006). It also drew on mathematics education research about quantitative reasoning, especially regarding ratio and proportion (e.g. Schliemann & Carraher, 1992).

The particulate model of matter rests on the concepts of mass and volume as distinct and measurable quantities. It relies on the understanding that matter can undergo change and involves processes that elude direct perception (e.g. the solution of sugar or evaporation of water). One of the basic issues of the work was to assess how young students can benefit from being introduced to these foundational ideas early on.

In this final report, we provide an overview of the project activities and findings across the longitudinal study, including the findings and activities of Year 5. Our work supports the following main conclusions:

- Grade 3 children begin with concepts of weight, volume, material, and matter, that are different from, and frequently at odds with, those of scientists. This supports our assumption that elementary school curricula may need to target these conceptual difficulties that are typically overlooked as important in traditional science curricula.
  - *Weight.* In the case of weight, young students initially rely heavily on their direct judgments of how heavy an object feels when hefted. Accordingly, they are reluctant to attribute weight to tiny objects when their weight cannot be directly felt.
  - *Volume.* Young students characteristically judge the size of objects by 'competing indices of size' such as length, perimeter, footprint, and surface area.
  - *Matter and material:* Overall, they young students do not have an overarching and coherent concept of matter as something that includes all solids and liquids, takes up space and has weight. They tend not to acknowledge gases as matter at all.
  - *Density.* Young students are unlikely to make clear and consistent distinctions between the 'heaviness' of an object (its weight) and the heaviness of materials from which objects are made (i.e. the relative weight of equal volumes of material) but this distinction may provide an

entry point into density prior to the introduction of formulas for computing density.

- *Quantities as dimensions.* Scientists regard basic physical quantities such as weight, volume, and area as measurable attributes that can be added, subtracted, multiplied and divided as well as ordered on quantity lines. Young students do not.
- Second, Treatment students (those who received the Inquiry Project curriculum) made clear progress from Grade 3 to 5 in understanding weight, volume, density, material, and matter. There was evidence that they were engaged in a productive, but drawn-out process of reconceptualizing the concepts in their matter network. Although they still showed difficulty on various items and were not at ceiling on all items even at the end of grade 5, the majority had made substantial progress in making changes that allowed them to meaningfully view matter as something that takes up space and has weight and that also includes gases as well as solids and liquids. More specifically:
  - By the end of Grade 3, over 90% understood that the identity of a material remained unchanged with decomposition into little pieces and that reshaping did not change the amount of material; almost 70% linked weight to amount of material and understood that even tiny things have weight and take up space; they also made significant progress in understanding that reshaping did not change an object's volume.
  - By the end of Grade 4, children over half of the children were systematically distinguishing the heaviness of objects from the heaviness of materials, and correctly inferring what material two "mystery" objects were made of using information about heaviness for size (not simply weight). Over half understood the amount of water displaced by a submerged object depended on its volume, not its weight, and more than a third correctly used the same cubes to judge the size of two rectilinear objects according to volume (rather than length or area.)
  - By the end of Grade 5, children were far more likely to have an explicit concept of matter and realize that even tiny things have weight. Over three-quarters classified all solids and liquids as matter and almost two-thirds were also including gases in these groupings. Over three-quarters understood that tiny pieces of clay take up space and have weight and that when sugar dissolves in water, the amount of sugar and its weight remain invariant. Many now explicitly said that all matter takes up space or has weight, and over 60% now systematically distinguished the weight of objects and the heaviness of materials. However, the 5<sup>th</sup> Grade Treatment students did not make further progress with volume and indeed showed some (slight) regressions on all volume tasks; this may have reflected the shift in curriculum focus to weight and matter. Disappointingly, fewer than a third of the treatment students judged size on the basis of volume (as opposed to area) at the end of Grade 5. In addition, although the vast majority of students understood that even tiny thing have weight, fewer

than a third treated weight as a continuous dimension (with lots or an infinite number of weights between the weight of two balls). These findings point to the need to keep volume front and center even after grade 4, and to have explicit discussions about the granularity of weight.

- Third, the progress among the Control students (those who had the standard curriculum) was much spottier. On some key tasks their progress was delayed relative to the Treatment students, although they eventually caught up (e.g., some of the conservation tasks). On many tasks, however, they either made no progress at all across grades 3 to 5 or very limited progress, so that by the end of Grade 5 there remained significant differences between the Treatment and Control students. For example, the majority of Grade 5 control students still did not think tiny pieces of clay took up space and had weight, that the amount and weight of sugar was conserved on dissolving, that the volume of clay was conserved on reshaping, or that volume, not weight was relevant on water displacement. The majority did not systematically distinguish the heaviness of materials from the heaviness of objects; they did not form systematic matter groups that included at least some gases as well as all solids and liquids; nor did they articulate the idea that all matter has weight or takes up space. On these tasks, their developmental trajectories were relatively flat and level of performance significantly lagged behind that of the Treatment students. It appears the Inquiry curriculum was more likely than the standard curriculum to help students develop a sound model of matter.
- Fourth, there were encouraging indications that young students can make use of quantity lines (number line where the units are in units of measure such as grams or cubic centimeters) to systematically order sets measurements and to think of physical quantities as dimensions. However, they tended to regard the distinct values on quantity lines and number lines as being “few and far between.” In principle, treating quantities and numbers as dimensions would seem to be consistent with the emergence of proportional reasoning. Both Treatment and Control students made some progress from Grades 3 to 5 in these regards, although in both cases the progress was limited and there were no differences between Treatment and Control students. Hence both proportional reasoning about physical quantities and understanding quantities as continua may need more direct nurturing.
- Finally, our data showed there were many similar relationships across tasks for Treatment and Control students, despite their different levels of performance and curricular histories. These relations may reflect conceptual constraints that need to be exploited in any curriculum based on a Learning Progression.

---

## I. Education and Research Activities

The Inquiry Project curriculum was designed to help elementary-school children build a sound understanding of matter and different materials in which the physical quantities of weight, volume, and density are inter-related and for which they recognize solids, liquids, and gases as distinct forms of matter. Our work has been guided by the view that

students' understanding of the atomic molecular theory rests on foundations built during the elementary and middle school years. The foundations consist of their *macroscopic understandings* of matter as well as their mastery of key, interrelated concepts. Research suggests that these understandings are achieved through deep and broad reconceptualizations of children's physical, mathematical, epistemological, and symbolic knowledge (Smith et al., op. cit.; Wiser & Smith, 2008).

---

## A. Education Activities in Grades 3-5

### 1. Development of the Inquiry curriculum

The Inquiry Project developed a coherent curriculum based on a learning progression about matter for Grades 3–5. The development process was iterative and collaborative. Together, developers, researchers, and teachers refined the Grade 3–5 learning progression and used the progression to develop a collection of cross-Grade learning experiences. These experiences were designed to deepen understanding of a network of core concepts central to understanding matter. Results emerging from the Inquiry research were taken into account, as were the realities of current school practice, e.g., time constraints, existing science curriculum, and the state and national standards for each Grade level. Key conceptual strands—the measurement of weight and volume, and the conceptual understanding of the nature of materials—continued to build across the three Grades culminating in Grade 5 with a careful transition to thinking about phenomena on a microscopic as well as macroscopic scale and an introduction to the particulate model of matter.

The development work resulted in three curriculum units each comprised of 15 to 18 investigations. Opportunities for formative assessment were embedded throughout the materials

*Grade 3: What are things in my world made of? How much is there?*

The Grade 3 curriculum has four sections that help lay the foundation for a learning progression about matter. The first strand, Investigating Materials, helps students distinguish between objects and materials. Students build their understanding that objects in their daily lives are made of many different types of materials with different properties. The second strand, Investigating Weight, focuses on weight not only as a property of objects but also more fundamentally as a property of matter itself. Students make the transition from felt weight, perceived with their hands, to measured weight using a pan balance. The third strand, Investigating Standard Measures, has students share their measurements of weight with each other and introduces the need for a standard unit of measurement. The fourth strand, Investigating Volume, introduces volume as another important property of matter.

Embedded within these strands is the development of representations (for example, a number line used to represent weight).

*Grade 4: What's under our feet? Investigating earth materials*

Earth materials provide a rich context for students' ongoing study of the important cross–

Grade concepts of weight, volume, and density.

The unit includes five sections. In Section 1, Underfoot, students investigate a variety of earth materials, examining and describing their properties. They discover that rocks can contain many different minerals. Through further investigation, they find evidence that minerals have some properties that stay the same regardless of the size of the sample. In Section 2, Heavy for Size, students measure the weight of liquid and granular samples of material, and think about which materials are “heavy for size” (the term we use in place of density). In Section 3, Liquid Measures, they measure both volumes and weights of liquids and compare the properties of oil and water. They focus on Mineral Materials in Section 4, using water displacement to measure the volumes of rocks and of samples of granular material (such as sand). This gives rise to a discussion about how much space is between the particles, and what could fill those spaces. Finally, in the culminating Section 5, they consider Transformations, such as the natural grinding of earth materials, leading to consideration of the weathering processes. For example, when an object, such as a shell or a rock, is transformed, what changes and what stays the same?

*Grade 5: Water Transformations: What changes and what stays the same?*

Without using the terms molecules or atoms, the Grade 5 unit provides a set of experiences that helps students develop the understanding that particles too small to see can have weight, take up space, and can help us understand the story behind transformations such as puddles evaporating or drops condensing on a glass.

In Section 1, Water a Liquid, students are introduced to the mystery of a disappearing sea and the scale at which evaporation can occur. They build their own mini-lakes and review concepts of weight, volume, and heaviness for size. As students investigate salt dissolving in water, they are introduced to three concepts that are central to the goals of this curriculum: 1) the matter we encounter in everyday life at the visible level is composed of tiny particles that are visible only when they are clumped together, and become invisible when they spread apart; 2) these individual particles continue to have weight and take up space even when they are spread apart and are too small to see; and 3) these particles must exist in unimaginably large numbers in order to account for the measurable size and weight of objects in the macroscopic world.

In Section 2, Water to Vapor, evaporation and condensation are investigated. Students consider, “What happens to water when it evaporates? Does evaporation destroy water, or does water continue to exist after it evaporates?” As students explore evaporation within a closed system, they also start to see the relationship between condensation and temperature.

In Section 3, Water to Ice, students consider how ice and water compare. While at the visible scale most of the properties of ice and water are quite different, the individual particles are identical. Students cannot prove this, but they rely on a combination of data they have collected, scientific reasoning, and finally a computer model as they consider whether or not water and ice are the same material. Students are presented with scientists’ belief that all matter—not just salt or water—is composed of tiny particles too small to

see individually. The computer model provides the particle view of the solid-to-liquid transformation of water as well as the relationship between temperature and particle motion.

In Section 4, Air a Gas, students investigate air, a mixture of gases. The class collects data to establish that air has weight, takes up space, and has properties that can be explored and described. Air has much in common with the more tangible forms of matter—solids and liquids—but this form of matter is not visible. Students use the computer model to investigate the liquid-to-gas transformation of water at the particle level.

In Section 5, Two Scales, students draw the connections between transformations at the visible level and their growing understandings of matter at the particle level. They apply their new understandings to their experiences with their mini-lakes and with bodies of water in the real world.

In this third unit, the water cycle is investigated at a tabletop scale rather than a global scale so that students can work directly with evidence about the change and the constancy of matter. Investigations with evaporation and condensation provide a bridge between visible and non-visible phenomena. Weight now becomes an important indicator of whether or not mass is conserved. Student experiences that bridge the visible and non-visible place more emphasis on inference, and care is taken to support the development of inference skills. The particulate model is introduced slowly over multiple investigations in which students need to explain what happens when water “disappears.” The computer model helps to students visualize the “tiny particles” that make up samples of solids, liquids, and gases.

Thus, students’ introduction to the particulate model of matter is grounded in evidence. Capitalizing on their belief that weight is an inherent property of matter, students explore the conservation of material across physical changes, using weight constancy as evidence for it. The water cycle theme provides a context for students to explore dissolving, melting, and thawing. They are introduced to gases first in the context of evaporation and condensation, and then explore the materiality of air. Again, volume and weight play an evidential role in the material nature of gases. Students use a displacement argument for air as matter—just like a brick in water, when they enter a room, the air needs to move out of the way. The large difference in heaviness for size between gases on one hand and solids and liquids on the other also becomes apparent.

Over three years, the Inquiry Curriculum aims to: replace hefted weight with an objective, extensive concept of weight as an inherent property of matter; differentiate length, area, and volume as distinct meanings of “big”; and build a concept of volume that, like weight, is measured objectively. At the same time, the notion of material is also systematized. Specific properties distinguish materials: how heavy for their size are objects of different materials; state and material identity are two intersecting concepts; and material identity is preserved when a solid chunk is reduced to powder. The concept of matter is built by generalization of properties rather than via definition—focusing on *weight and volume as inherent properties of materials* creates an important bridge between solids and liquids on one hand, and gases on the other, which leads to a new network of conceptual understanding.

This careful conceptual building is intended to establish links between existing and new ideas so that the new ideas make sense, thereby avoiding the destabilization commonly created by simply telling students “solids, liquids and gases are forms of matter.” Students first associate weight and volume with solids and liquids, and, importantly, with the tiniest pieces of materials, as well as with their conservation across transformations in which material remains visible (grinding, melting, freezing). In fifth Grade, gases are introduced via evaporation and condensation in a two-bottle closed system in which water evaporates in one bottle and condenses in the second. This context evokes the previous transformations. Again, students can rely on weight as evidence that evaporated substances continue to exist. They also can rely on their belief that tiny pieces continue to exist even if they cannot see them, and have weight and volume, to start making sense of the gaseous state.

All units include investigations, brief essays from scientists and psychologists about children’s ideas, concept cartoons (i.e., assessments), embedded assessment opportunities, student notebooks, and guiding questions for discussion. The completed version of each unit is available for free on the Inquiry Project website ([inquiryproject.terc.edu](http://inquiryproject.terc.edu)).

## 2. Development of Formative Assessments

Opportunities for formative assessment are identified within the web-based version of the curriculum. The learning experience itself is the source of assessment data. Each assessment opportunity includes the goal of the assessment expressed as a question. For example, do students understand that when a ball of clay is reshaped, the weight and volume will stay the same? Goals may address science content or the inquiry process. The teacher then **collects evidence** of students’ progress towards the goal. In the classroom, there are three sources of evidence: what students say as they discuss their work, what they do as they carry out first-hand investigations, and what they write or draw. Based on a set of criteria, the teacher **interprets** and appraises what she sees or hears, and decides what **next steps** will help students make further progress towards the goal. Then, she sets a new goal and another formative assessment cycle begins.

Each curriculum unit also includes a set of Concept Cartoons<sup>®</sup>. Cartoon style assessment items benefit from information presented visually with minimal text or language load. Students enjoy directing their responses to the characters themselves and so are motivated to engage in some serious thinking, reasoning, and argumentation. For each concept cartoon, there is a description of the purpose of the assessment and a section called “Things to look for in student responses” to help teachers to interpret student responses and determine next steps in the learning. Note: Concept Cartoons<sup>®</sup> were created by Brenda Keogh and Stuart Naylor in 1991.

## 3. Implementation in Grades 3–5 classrooms

Early trials of core investigations were conducted with small groups of students at TERC and in non-research classrooms. A full pilot was then conducted in a local-area school.

The completed curriculum was implemented, revised, and implemented again in the research classrooms at Forestdale School, Malden, MA and Mason School, Roxbury,

MA. A co-teaching model was employed to ensure fidelity of implementation. Each investigation was videotaped in at least one of the five classrooms. Copies were made of all student notebooks for research purposes.

#### 4. Professional development with teachers

Grade 3–5 teachers from Forestdale School and Mason Schools met with the developers and researchers monthly throughout the project. Although the focus was on the curriculum at one Grade level each year, all teachers in Grades 3–5 took part in the meetings across all three years. The aim was to help teachers to understand how conceptual development evolved across multiple Grades.

At these meetings, teachers:

- “Tried out” investigations from the curriculum to become familiar with the materials they would teach. They also provided feedback to developers;
- Deepened their understanding of the science behind the investigations;
- Worked with formative assessment strategies, e.g., reviewing student notebooks and giving useful feedback, and analyzing student responses to concept cartoons to gain insight into children’s scientific ideas; and
- Became familiar with the rationale and strategies for productive science talk, particularly around working with data. (Within the curriculum, discussion is a critical component.)

A key purpose of the professional development was to ensure that the curriculum was faithfully implemented. Accordingly, time within the PD sessions was devoted to helping teachers to become familiar with the rationale underpinning the learning experiences. Viewing teachers as development partners was important as their insights helped us to bridge research and the reality of the school.

---

### **B. Research Activities**

#### 1. Research overview

Two groups of students were followed longitudinally from Grades 3 through 5 using repeated-measures, quasi-experimental design:

- **Treatment students** (those who received the Inquiry science curriculum for nine weeks in each of Grades 3, 4, and 5) were interviewed on four occasions over two and one-half years: (a) early Grade 3, before the Inquiry Curriculum; (b) end of Grade 3 after the first Inquiry Curriculum unit; (c) end of Grade 4 after the second Inquiry Curriculum unit; and (d) end of Grade five after the third Inquiry Curriculum unit. These students are from five classrooms in two different schools.
- **Control students** (students from the same school who received the standard science classroom instruction and the same teachers in Grades 3-5) are interviewed at three occasions: (a) end of Grade 3, (b) end of Grade 4, and (c) end of Grade 5.

The repeated measures design allows an isolation of the effects of maturation and age-related development from the quasi-experimental Treatment. Assuming there are no systematic initial or historical differences between the two groups—something supported by data from the first interview occasion, as well as standardized MCAS (Massachusetts Dept. of Education, 2007) test scores—we can assess the extent to which the implemented Inquiry Curriculum exerts a significant effect on learning. Once differences have been demonstrated that are not attributable to chance variation, we seek to interpret the findings as a function of the Treatment. As with all broad educational interventions, the distinction between the Treatment and Control conditions cannot be reduced to a single factor or variable.

We conducted 345 interviews with between 54-67 treatment and 35-38 Control students at each occasion of testing, interviewing the same students at each occasion as much as possible. Because some students left the school across the Grades or because some students did not return permission slips at a particular testing time, there was some attrition of the original sample across Grades. Attrition was greater for the Control Group than the Treatment group. Whereas permission slips returns diminished for the Control Group, there were increasing permission slips returns over time, which led to adding some new Treatment students to the sample (post Grade 3, and post Grade 5). Any treatment student added was, however, present for the entire Grade 3-5 curriculum intervention. To maintain the sample size in the Control Group, we also added new students among those who returned permission slips. In general, the new Control students added were eager to participate, and there were high levels of motivation to participate in the interview for both Treatment and Control students.

Table 1 displays the total numbers of Treatment and Control students interviewed at each point in time. Table 2 provides data on attrition of students in the two groups, and the number of students for whom we have complete longitudinal data across all time periods (henceforth referred to as the *Complete Longitudinal Subsample*). In both the Treatment and Control Groups, the majority of students were interviewed at least two or three times.

**Table 1: Number of Treatment and Control students interviewed**

Group	Pre-G3	Post-G3	Post-G4	Post-G5
Treatment	58	67	56	54
Control	No interview	38	37	35

**Table 2: Retention statistics for the Treatment and Control Groups**

Group	Pre-G3	Post-G3	Post-G4	Post-G5	All interviews
Treatment	58 (100%)	57 (99%)	46 (79%)	42 (72%)	40 (69%)
Control	No interview	38 (100%)	27 (71%)	19 (50%)	17 (45%)

69% of the original treatment group interviewed took part in all four treatment interviews; 45% of the original Control Group took part in all three Control interviews. At the end of Grade 4 the treatment group subjects were more likely to take part in the interviews than the Control subjects (79% vs. 71%). At the end of Grade 5 the difference is greater (72% vs. 50%).

This trend is not particularly surprising. The treatment students stood to benefit from the Inquiry instruction; this may have inclined their parents to grant permission to interview. In addition, the Inquiry teachers may have been somewhat more persistent than Control teachers about getting their students to return signed permissions forms.

If familiarity with the interview tasks were a factor in performance, it should not systematically favor either of the longitudinal groups (treatment and Control subjects who took were interviewed throughout the three years). It would possibly show up as differences in performance as a function of the number of times a subject was interviewed, or more progress in the longitudinal vs. complete sample. One way we addressed this issue was by comparing the progress in longitudinal vs. complete sample. We found no evidence of differences in this regard.

## 2. Methods of Coding & Analysis

Most of our analyses make use of the data from the entire sample (where both between-group and within-group Grade changes were tested using Pearson Chi square tests). For longitudinal analyses we drew from the subsample of students for whom we had interview data for all phases of the investigation. In such cases group differences were typically tested using Chi-square, whereas within-group changes relied on Wilcoxon Matched Pair Tests. Overall, findings were highly consistent in terms of percentage of students showing a given understanding at a given Grade, evidence of Treatment/Control differences, and tests of whether students made progress across time periods, increasing our confidence in our results based on our complete sample. Because the complete sample involves larger numbers at each time period (and hence greater statistical power for detecting Treatment/Control differences), we present these results in the summary Tables and analyses that follow. However, in presenting findings about student progress from Grade to Grade, we will present findings both from within-student analyses of progress (assessed in the longitudinal subsample) as well as tests of progress from the complete sample, as the within-student analyses provide a more sensitive (and controlled) measure of progress.

The principal sources of research data include (a) structured individual clinical interviews conducted throughout Grades 3–5 with students in the Treatment and in the Control Group and (b) video footage and student work from teaching experiment lessons. Here we will focus on the out-of-class interview activities and data. The classroom activities and data are described in the Education Activities sections of this report.

Each student received the same two-hour interview on multiple occasions. The interview consisted of 10-multi-part tasks that probed children's concepts of matter, amount of material, weight, volume, density, number, fraction, division, ratio, and proportion. Sub-tasks were structured so that the interviewer could adjust the questioning to the thinking of the student.

The research team's activities during Year 5 fall into the following main categories:

- Completion of coding and data analysis,
- Discussions with the full team about the relations between findings and implications for design of curriculum, and
- Write-up and presentation of findings at two National Conferences (AERA, Jean Piaget Society Meetings).

---

## II. Findings

If students are to recognize the explanatory power of the particulate model of matter and begin to accept it as a reasonable and useful way of thinking about the world, they need a firm understanding of the nature of matter and materials, the concepts of weight and volume, and skill in inferential and hypothetical reasoning. They also need quantitative reasoning skills linked to these concepts. Our results suggest that significant progress can be made in laying this foundation in Grades 3-5 with activities that are highly engaging for the students, not overly time-consuming, and compatible with existing curricular constraints. Progress due to the Treatment was more pronounced in some areas than others, as the following sections will show.

---

### A. Understanding Matter and Material Things

What do children need to know in order to have a meaningful concept of matter? How can we assess what children understand about matter, given that many children may not yet know the word “matter”?

In this section, we discuss four ways we probed implicitly for relevant precursor understandings, and then consider a task that probed children's explicit ideas about the meaning of the word “matter” itself, considering both the similarities and differences between Treatment and Control students in their performance on these tasks. We argue there is much more to developing a meaningful concept of matter than learning the word: indeed, many understandings need to be in place before children are ready to develop a meaningful general concept of matter. Central to LP work is identifying those useful precursor understandings as well as understanding key conceptual obstacles.

#### 1-Identifying Materials: Is sawdust wood? Will it burn?

Well before they have learned the word “matter”, preschoolers have begun learning the names of specific materials (e.g., clay, sand, sugar, wood, water), some of their typical uses and properties (clay can be used to make things, sugar makes things sweet, water is a drink). They also are beginning to understand that the same objects can be made of different materials (e.g., one can have a metal, plastic, or wooden spoon or baseball bat) and are beginning to think about what materials different objects can be made of.

But how abstractly do children think of these materials? Do they think of materials as “formless stuff” from which many different objects can be made or fashioned, but which nonetheless has some enduring and characteristic properties? Do they distinguish the properties of the material from the property of the object, and realize that some properties of objects depend upon the kind of material of which they are made? For example, that

rubber balls are bouncy *because* they are made of rubber, but metal balls are not, although both balls roll because they are round. Do they understand that a material can maintain its identity and some (objective) properties across changes in its state (e.g., decomposition into little pieces, melting) despite changes in surface appearance? An important part of developing a concept of matter is distinguishing objects from materials, recognizing, for instance, that materials can maintain remain invariant despite changes in the identity of the object.

To assess whether children understand that materials maintain their identity and some properties across changes in state and outward appearance, we presented them with three different cases: (a) a rectangular piece of wood before and after being sanded down to a fine powder; (b) a piece of iron before and after being transformed into tiny black filings; and (c) a pat of yellow butter before and after being melted to form a clear runny liquid.

We asked whether the sawdust, black filings, runny liquid were still wood, iron, and butter, respectively. We also asked whether the sawdust would burn and the filings would be attracted to a magnet.

The results suggest that elementary school children in both Treatment and Control Groups are beginning to conceptualize materials in this more abstract way, although they are more confident about the invariance of material identity than about material properties.

At the beginning of 3rd Grade, 63% of the pre-Treatment students consistently judged that the sawdust was still wood, the black filings still iron, and the runny liquid still butter. Children who affirmed the continued identity of materials focused on the fact that it *came from* a chunk of that material (arguments based on its historical continuity with the previous stuff) and/or that it was just cut into little pieces or ground up or melted (arguments based on the irrelevance of the specific transformation for changing material identity). In contrast, those who said the identity had changed focused on changes in appearance (it doesn't look or feel the same) or form (it's powdery, it's runny, it's little pieces) as reasons for thinking it were no longer the same material.

**Table 3: Invariance Across Grinding (Wood, Iron) & Melting (Butter): % of Children Saying Yes for All Three Substances**

	Grade 3 Pre	Grade 3 Post	Grade 4 Post	Grade 5 Post
<b>Treatment</b>	<b>63%</b>	<b>91%*</b>	<b>88%</b>	<b>93%</b>
<b>Control</b>		<b>74%</b>	<b>84%</b>	<b>91%</b>

Treatment/Control difference Significant Pearson Chi Square, \*p < .05

Treatment students made significant progress in understanding the invariance of materials by the end of Grade 3 ( $\chi^2$  (1, N=126) = 12.98, p < .001 for complete sample): at that point, fully 91% of the Treatment children maintained that sawdust was still wood, the filings still iron, and the runny liquid still butter. This may reflect the impact of the 3<sup>rd</sup> Grade Inquiry Curriculum more than the effects of repeated testing, as performance of the Grade 3 Treatment children was significantly better than the Grade 3 Control ( $\chi^2$  (1, N=104) = 4.39, p < .05), the progress children made from pretreatment Grade 3 to post

Grade 3 was significant, while the progress of Control children from post Grade 3 to post Grade 4 was not (based on both Chi Square and Wilcoxon tests of progress), and helping children distinguish the properties of objects and materials was one goal of the 3<sup>rd</sup> Grade Inquiry curriculum. More specifically, children in the Inquiry Curriculum discussed whether something was a property of a material or an object and were told one way to decide whether something is a property of a material is to imagine whether the description would be true of a tiny piece as well as a large piece.

Nonetheless, understanding that materials maintain their identity with grinding and melting are already common in the Grade 3 Control Group as well, although not yet as strong as Grade 3 Post Treatment. By Grade 5, children in the Control Group show the same level of performance as the Treatment students did in post Grade 3. So clearly these ideas are developing based on experiences both groups have with materials.

Both Treatment and Control children also made significant progress in making inferences that some (non-perceptual) properties of materials would remain the same across grinding, although their understanding of the invariance of these properties lagged behind their understanding of the invariance of materials. On this measure there were no differences between Treatment and Control students at any Grade (see Table 4).

**Table 4:** Invariance Across Grinding: “Sawdust Burns” and “Iron Filings are attracted to a Magnet”

	<b>Grade 3 Pre</b>	<b>Grade 3 Post</b>	<b>Grade 4 Post</b>	<b>Grade 5 Post</b>
<b>Treatment</b>	<b>42%</b>	<b>64%</b>	<b>70%</b>	<b>78%</b>
<b>Control</b>	<b>--</b>	<b>55%</b>	<b>65%</b>	<b>80%</b>

The fact that children had more doubts about the invariance of these properties than material identity is not surprising. In fact, it is an empirical question, which properties of large chunks are properties of that material maintained at a very small grain size. For example, glass, sugar, and salt are all transparent in large chunks but white as powders. Understandably some children wondered whether particles of wood needed to be a certain size in order to burn or to be attracted to a magnet. This points up an important conceptual challenge for students. For it to be meaningful for students to say something is still the same material, they need to know of *some* properties of that material are preserved in the decomposition rather than to think all of its properties have changed. Given that many perceptual properties change with decomposition, it is important that children have knowledge of objective physical properties that are distinguishing of different materials at a small grain size. Thus, an important part of developing children’s understanding of materials is developing their understanding of those properties that are invariant across a broad range of grain sizes and hence that can be used in tracing material identity across various transformations. In fact, many of the most important properties scientists use to identify materials (e.g., electrical and thermal conductivity,

density, solubility, boiling points, etc.) are not ones that are initially most salient to or known by students.

The fact that there were no differences between Treatment and Control students in making inferences about these properties is also not surprising, as they were not discussed in the Inquiry Curriculum. (See later section C where we discuss the development of children's understanding of density as a property of materials where there was more evidence for Treatment Control differences.) These properties were chosen because they were ones we thought children might associate with these materials based on their everyday experience (e.g., children commonly see wood logs burn in fireplaces or campfires and have experiences with iron nails or other things being attracted to magnets). At the same time, we thought children might not have direct experience with seeing whether sawdust burns or filings are attracted to magnets. (We found from student comments in the interviews, however, that some had experimented with iron filings in other science classes or with toys). In this way, they required a genuine inference based on their conceptual knowledge. We did not allow children to test out their ideas about whether small pieces have these properties in the interview because we wanted the question to remain undecided across repeated interviews.

## 2-Invariance of Amount of Material and Weight When Shapes Change

Hand in hand with the development of a concept of material<sup>1</sup> is the development of a concept of "amount of material." Children's early concepts of physical amounts (such as big and heavy) are perceptually centered quantities judged directly by looking or hefting, and hence subject to change amid the flux of everyday sensory experience. "Amount of material," in contrast, is a model-centered quantity inferred based on a pattern of relations and invariant across changes in perceptual appearances. For example, children can infer that a clay ball has a certain (unspecified) "amount of clay," and then reason if that object is broken into pieces or reshaped that amount remains invariant across changes in perceptual appearance because no further material has been added or taken away. This new concept of amount of material can then be used as a basis for reconceptualizing size and weight: they become objective physical quantities, rather than perceptually centered quantities, as each is linked to amount and assessed using various physical phenomena (e.g., balance scales, water displacement).

We used two tasks to probe children's developing understanding of amount of material as invariant across shape change and to assess whether they linked both weight and volume to this invariant.

One was an extension of the standard Piagetian "conservation of clay" task. Children were presented with two clay balls identical in size and shape that were shown to be in balance on a balance scale. Then one ball was flattened into a pancake, and children were asked of the ball and pancake: (a) do they have the same amount of clay, or does

---

<sup>1</sup> Children's initial concept of material is a precursor of the later more abstract concept of matter, which is not needed until children are ready to explore what stays the same across transformations such as phase change. Similarly, children's initial concept of "amount of material" is a precursor for the later concept of mass. Developing an explicit concept of mass differentiated from and inter-related with weight was not a goal of the Inquiry Curriculum.

one have more? (b) Do they weigh the same, or does one weigh more? (c) Do they take up the same amount of space (have the same volume), or does one take up more space? (d) Will they balance on the scale, or will one side go down? And (e) if placed completely under water, will they make the water level go up the same amount or will one make the water rise higher?

In the second task, children were presented with two shapes made of 12 equal size clear plastic cubes: one in a block shape (2 x 3 x 2) and the other a longer “snake” shape (2 x 6 x 1), and were asked the same 5 questions, but in a different order: (a) Do these two shapes have the same volume (i.e., take up the same amount of space?); (b) Do they weigh the same? (c) Do they have the same amount of plastic?; (d) Will they balance on a balance scale?; and (e) Will they make the water level rise the same amount? In both cases, children had to answer without using any measuring instruments, and were given no feedback on their answers.

Table 5 shows the percentage of students who judged that the *amount of clay* and the *amount of plastic* were the same across the shape changes (ball vs. pancake; block vs. snake). At the beginning of Grade 3, 64% of the children consistently judged the amount of material to be invariant; the remaining children were less consistent, typically judging it as invariant for one but not both tasks. (The two tasks were of comparable difficulty.) By the end of Grade 3 both Treatment and Control students were near ceiling on both questions and performance remained at this high level across Grades 4 and 5. There were no Treatment or Control differences. This result is hardly surprising; indeed it is consistent with much past developmental research, and affirms that the idea of “amount of material” as an invariant is an important resource on which curricula can build.

**Table 5: Amount of Clay, Plastic Both Viewed as Invariant Across Shape Change**

	Grade 3 Pre	Grade 3 Post	Grade 4 Post	Grade 5 Post
<b>Treatment</b>	<b>64%</b>	<b>88%</b>	<b>91%</b>	<b>89%</b>
<b>Control</b>	<b>--</b>	<b>80%</b>	<b>84%</b>	<b>86%</b>

Table 6 shows, however, that although it is fairly obvious to children that the amount of material is invariant across shape change, it is not equally obvious that the weight of the objects is invariant, or that the two objects will balance on a scale. Instead, we found that children often deal with these as different questions, which they answer in different ways. Thus, at the start of Grade 3, only 41% of the children consistently judged that the objects not only would have the same amount of material, but also would weigh the same, and would balance on a balance scale.

However, for the Treatment students, there is a big improvement by the end of Grade 3 in judging amount, weight, and balance as invariant across shape change. This improvement may be attributable to the Inquiry Curriculum, as performance of the Control students lagged behind, only reaching Grade 3 post treatment levels by the end of Grade 5. An important goal of the Grade 3 curriculum was helping students develop an objective, measurable concept of weight, linked to amount and the balance scale. Both

Treatment and Control Groups do make progress across Grade, which probably reflects common curricular experiences with learning about weight measurement.

**Table 6: Amount, Weight, and Balance as Invariant Across Shape Change for Clay, Plastic (pattern across 6 items)**

	Grade 3 Pre	Grade 3 Post	Grade 4 Post	Grade 5 Post
<b>Treatment</b>	<b>41%</b>	<b>69% **</b>	<b>80% **</b>	<b>78%</b>
<b>Control</b>	<b>--</b>	<b>43%</b>	<b>51%</b>	<b>69%</b>

Treatment/Control difference Significant, \*\* $p < .01$ , Pearson Chi Square

Children's understanding of the invariance of volume across shape change lagged behind their understanding of the invariance of amount and weight, although again was enhanced for Treatment students relative to Control. The difficulty of those questions is most likely related to constructing an understanding of volume; hence those findings will be presented in the section on volume (next main section). In the absence of a concept of volume, children focus on other meanings of taking up space (how long it is, or how much space it covers) that are not invariant across shape change.

### 3- Properties of Tiny (Visible and Invisible) Things

A more direct way to probe whether children think of taking up space and having weight as either perceptually grounded or objective quantities is to ask them simple questions about the properties of tiny pieces of clay. At issue is whether children think that even tiny pieces of clay take up space, have weight, and maintain their existence and these properties with repeated division. Previous research has found that most young children think tiny pieces weigh nothing at all and take up no space, because “they are too tiny to weigh anything or take up any space”—a key indicator that the properties are still perceptually centered quantities.

Children are first handed a large ball of clay (about the size of a squash ball) and asked: *Do you think this weighs something?* We chose a piece that has palpable weight (120 grams) so that children would readily agree that it does, to set the stage for the next questions. Children then watch as a tiny piece (about 2 mm in diameter) is broken off. The piece is handed to them, so they have the opportunity to both see and feel the piece, and they are asked: *Do you think this weighs a tiny bit or nothing at all? How do you know? Do you think this piece takes up any space? How do you know?*

Children are next asked about the possibility of an invisible piece of clay: *Could there ever be a piece of clay so tiny that you couldn't see it?* If students agree that there could be such a piece, they are then asked: *Would that tiny piece, so small you can't see it, take up any space? How do you know? What that tiny (invisible) piece have any weight? How do you know?*

We have found these very simple questions are a powerful means of eliciting some of students' most basic presuppositions about matter, weight, taking up space, and amounts, without requiring any technical vocabulary.

Table 7 shows that at the beginning of Grade 3, only 7 % of the pre-Treatment students said that the speck of clay both took up space and weighed a tiny bit.<sup>2</sup> The majority of Grade 3 to Grade 5 Control students also failed to attribute both taking up space and having weight to the visible little pieces. The fact that problems arise for *visible pieces* is instructive, as it shows problems emerge well before the microscopic scale. The fact that this problem persists for the Control students at all time periods suggests this conceptual difficulty is not adequately addressed with typical elementary school curricula. We believe success on these questions calls for model-mediated reasoning about weight. As the pieces are too tiny to feel that they have any weight or could affect a balance scale, children need to appeal to some theoretical beliefs (such as that everything must weigh something) in order to conclude that they do.

Strikingly, the Treatment students, but not the Control, made dramatic progress on these questions by the end of Grade 3. Now the majority asserts that the tiny piece both takes up space and has weight. These new understandings seem directly attributable to the Inquiry Curriculum, as there are highly significant differences between the Treatment and Control Groups. These differences between Treatment and Control Groups are maintained through the end of Grade 5.

**Table 7: “Tiny (Visible) Pieces Take Up Space and Have Weight” (2 items)**

	<b>Grade 3 Pre</b>	<b>Grade 3 Post</b>	<b>Grade 4 Post</b>	<b>Grade 5 Post</b>
<b>Treatment</b>	<b>7%</b>	<b>67% ***</b>	<b>57% *</b>	<b>76% ***</b>
<b>Control</b>	<b>--</b>	<b>10%</b>	<b>30%</b>	<b>31%</b>

Treatment/Control difference Significant, \*  $p < .05$ , \*\*\* $p < .01$ , Pearson Chi Square

The third-Grade Inquiry Curriculum provided direct support for thinking about this issue in multiple ways (see Wiser, Smith, Asbell-Clarke, & Doubler, 2009 for more details). Briefly, children used their hands to order a series of objects by weight, compared their “felt weight” order with the weight obtained from using a balance scale, and explicitly discussed the strengths and limits of their senses. They also learned to use measurements to compare the weights of objects and to use weight line representations. They explored the additivity of weight in the 10-10-10-10 challenge (constructing new objects from 10 grams of Styrofoam, 10 grams of aluminum, 10 grams of wood, and 10 grams of clay) and predicting the weight of the diverse objects created by class members. Finally, they used their weight line representations to explicitly consider what happens to the weight of a 4 gm. object (clay piece) that is repeatedly divided in half. Would you ever get to zero? They also put the pieces together to confirm that they summed to 4 gm. The data suggest these discussions were highly effective and that children can develop an understanding of these issues as early as third Grade.

---

<sup>2</sup> The three most common patterns were to (a) deny that the pieces took up space or had weight; (b) agree the piece took up space, but deny it had any weight; or (c) agree that it both took up space and had weight. In general, more children thought tiny pieces took up space than had weight.

Although children continued to investigate the weight and volume of materials in the Grade 4 Inquiry Curriculum (extending their investigations to liquid and granular materials), children only returned to the issue of the weight of tiny things in the Grade 5 Inquiry Curriculum, this time in the context of investigating the weight of a drop of water and of gases. (The 4<sup>th</sup> Grade curriculum focused more on developing the idea of volume and coordinating weight and volume by developing a notion that materials varied in their heaviness for size<sup>3</sup>.) This fact may explain the slight drop in children's performance in Grade 4 and resurgence again in Grade 5 when the issue was revisited. Certainly the work children did in Grade 3 prepared the ground for their benefiting from the new lessons in Grade 5.

Further evidence that the Grade 3-5 Treatment students had acquired new beliefs about what things take up space and have weight come from their responses to the questions about *invisible pieces* and an analysis of their *justifications* of their judgments (see Tables 8 and 9). Again, we found striking differences between the Treatment and Control students, that were present by the end of Grade 3 and continued through to the end of Grade 5.

**Table 8: “Tiny (Invisible) Pieces Take Up Space and Have Weight” (2 items)**

	Grade 3 Pre	Grade 3 Post	Grade 4 Post	Grade 5 Post
<b>Treatment</b>	7%	45% ***	41% *	54% *
<b>Control</b>	--	5%	16%	26%

Treatment/Control difference Significant, \*  $p < .05$ , \*\*\* $p < .01$ , Pearson Chi Square

Table 8 shows the percentage of children who think invisible pieces exist, take up space, and have weight. At the beginning of Grade 3 we found three-quarters of the students were willing to accept that invisible pieces could exist, but almost none thought they both took up space and had weight. Thus, accepting that invisible pieces exist was not nearly so problematic for them as accepting that invisible pieces take up space and have weight. By the end of Grade 3, almost half of the Treatment children thought invisible pieces take up space and have weight—a number that slightly increased by the end of Grade 5, and remained significantly different from the Control Group at all three Grades.<sup>4</sup>

Table 9 shows the percentage of children who justified their judgments that a small piece of clay had weight by offering a “Universal” generalizations about weight—direct

<sup>3</sup> In the Inquiry Curriculum the terms *heavy* and *light* were taken as referring to the weight of an object. *Heaviness*, or heaviness for size was generally used to convey the relative weight of one type of material vis-à-vis an equal amount (volume) of another material.

<sup>4</sup> Although some Treatment students who had agreed that visible pieces take up space and have weight denied that invisible pieces do, most commonly (in about half the cases) this was because they weren't sure that Invisible pieces exist. Thus, these children weren't being inconsistent in their reasoning about the properties of visible and invisible pieces so much as having difficulty in accepting the premise itself.

evidence that children's changed judgments were a product of the development of new, explicit beliefs. These justifications took a variety of forms, most typically "Everything has to have weight," "All things have weight" or "If it exists it must weigh something," but sometimes "All solids have weight," "All objects have weight" or even "All matter has weight." Again, these explicit Universal Generalizations emerged at the end of Grade 3 for the Treatment students, and were much more prevalent for Treatment than Control students at all three Grades. However, many of the Control students who judged that tiny things have weight, also appealed to Universal Generalizations as justifications.

**Table 9: Universal Generalization Given At Least One Weight Judgment**

	<b>Grade 3 Pre</b>	<b>Grade 3 Post</b>	<b>Grade 4 Post</b>	<b>Grade 5 Post</b>
<b>Treatment</b>	<b>2%</b>	<b>60% ***</b>	<b>43%*</b>	<b>59%***</b>
<b>Control</b>	<b>--</b>	<b>3%</b>	<b>22%</b>	<b>17%</b>

Treatment/Control difference Significant, \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ , Pearson Chi Square

Although similar numbers of post Grade 3 and 5 Treatment students appealed to universal generalizations about weight, there was a change in wording used. By the end of Grade 5, 35% of the Treatment students explicitly said "All matter has weight" compared to only 3% at post Grade 3, where students used the more generic "Everything has weight" or "All objects have weight." This undoubtedly was related to the introduction of the general idea of "matter" in the Grade 5 Inquiry Curriculum. No Control student appealed to generalizations about matter in this task, at any time period.

Taken together, these data show that during Grade 3, the Treatment students made considerable advances in the tendency to attribute weight and space to tiny yet visible objects as well as to invisibly tiny pieces of matter in part because they were developing explicit beliefs that "everything" must weigh something and take up space. In contrast, the Control students, who had their regular science curriculum, made much less progress on these questions. This shows that young children are certainly ready to engage with these issues, but they need instructional support to do so—these ideas do not simply "come for free" with development.

There was also evidence that developing an understanding that tiny things have weight and take up space was an important watershed idea that contributed to the development of other important ideas about matter and materials. In subsequent sections, we will provide evidence that it goes hand in hand with (a) recognizing commonalities among solids and liquids (contributing to the early development of a concept of matter) and (b) differentiating the heaviness of objects and the heaviness of materials. It also prepares the way for developing the belief that all matter has weight and takes up space.

#### 4-Sugar dissolving in water

Another way that we probed children's understanding that even tiny (invisible) things have weight was in the context of sugar dissolving into water. In this task, children were shown two beakers containing the same amount of hot water, and two equal-size sugar

cubes. One of the sugar cubes was placed in Beaker B and children watched as the sugar dissolved. After the children were asked to explain what was happening, they were asked two key questions: (a) Was the amount of stuff in Beaker B (the sugar-water mixture) the same as the amount of stuff when one considered Beaker A and the sugar cube together (Beaker B was placed in one circle and Beaker A and the cube were placed in another), or does one have more stuff? (b) If one puts Beaker B on one side of the scale, and Beaker A and the sugar cube on the other side, would they balance or would one side go down more?

Table 10 shows that Treatment children made steady progress in understanding that there was the same amount of stuff before and after dissolving and that the dissolved water and sugar would still balance with the water and sugar cube, with the greatest progress between Grade 4 and Grade 5 (based on both Chi square and Wilcoxon tests of progress). In contrast, much more limited progress was made by Control students (not significant with either Chi square or Wilcoxon tests). At the end of Grades 4 and 5, the differences between the Treatment and Control were significant.

**Table 10: Dissolved & Undissolved Sugar: Percent Who Judge Both Same Amount and Same Weight**

	Grade 3 Pre	Grade 3 Post	Grade 4 Post	Grade 5 Post
<b>Treatment</b>	<b>44%</b>	<b>52%</b>	<b>61%*</b>	<b>85%***</b>
<b>Control</b>	<b>--</b>	<b>35%</b>	<b>38%</b>	<b>46%</b>

\*Chi Square,  $p < .05$ , \*\*\* Chi Square,  $p < .001$

The fact that greater progress occurred between Grade 4 and Grade 5 is probably due to the fact that children explicitly investigated what happens to weight when salt is dissolved in the 5<sup>th</sup> Grade Inquiry Curriculum. However, the fact that there were already Treatment and Control differences by 4<sup>th</sup> Grade (before the topic of dissolving was discussed in the Inquiry curriculum) suggests that earlier discussions of the weight of tiny things in Grade 3 was already preparing the ground for their understanding of this issue.

### 5-Solids, Liquids, and Gases

A fifth task directly probed children's ideas about what entities were "matter" and "not matter." At issue was what classification they would make, and how they would justify them. Prior research has shown that children have great difficulty identifying gases as matter; hence, we expected that children would not see gases as belonging in the same class as solids and liquids, although there are limited prior data on classification patterns for elementary school children.

To begin this task and to help children who might not yet know the scientific word "matter," we told all children:

*Now we are going to talk about what things are matter, are made of stuff, and what things are not. For example, this rock is made of stuff, it is matter. This piece of cloth is made of stuff too, it is also matter. This (plastic) horse is made of stuff, and real horses are also stuff, they are matter. But when you feel happy*

*about something, your happiness is not made of stuff, it is not matter, It is not like the rock or the horse. And ideas are not made of stuff either, they are not matter.*

*On these cards I have the names of more things. I would like you to put these cards into piles—things that are matter, that are some kind of stuff; things that are not matter; and things you are not sure about.*

There were 14 cards naming the following things: a piece of wood, an ice cube, a dog, a mosquito, a pile of sand, water, milk, air, smoke, steam, heat, light, shadow, and dream. After sorting into piles, children were asked why they had put some things in the matter pile, if there was a way to decide if something was matter, and if there was any way all the things in the matter pile were alike. They were then asked why they had put some things in the matter pile, if there was a way to decide if something was not matter, and if there was any way “not matter” things were all alike. Finally, they were asked about the items in their unsure pile, including what made them unsure and which pile they would put them in if forced to choose. Throughout the task, children were free to switch their ideas about what pile things belonged in. Our analyses were based on their final classifications of all the items when they were forced to choose.

Table 11 shows the results when children’s classifications are analyzed this way.

Surprisingly, at the beginning of Grade 3, only 37% picked out all the solids and liquids as matter (i.e., piece of wood, ice cube, dog, mosquito, sand, water, milk)—certainly the “easiest” items to recognize as matter. The low percentage of children who grouped all these items together shows children’s difficulties with the category of matter begin well before their need to assimilate gases to this category—a fact not highlighted in the prior literature. Thus, coming to deeply appreciate the commonalities of diverse solids and liquids is itself an achievement. In keeping with the well-known difficulty of viewing gases and gas-like mixtures (air, smoke, steam) as matter, we found only 25% of the Grade 3 Pre-Treatment students not only picked out all solids and liquids, but also included some gases. Only 8% picked out all solids, liquids, and some gases without over-extending to immaterial items. Only one child picked out all and only material entities.

The remaining classifications by pre-treatment 3<sup>rd</sup> Graders were quite idiosyncratic and did not indicate even a “precursor” idea of matter. For example, many simultaneously excluded some solid or liquid material entities (such as water, sand, or ice cube) while including some non-material entities (such as heat, light or shadow). Their justifications showed they were often focusing on a variety of irrelevant attributes, such as whether something is important, natural or man-made or whether they knew what it came from (e.g., water isn’t matter because it doesn’t come from anything else) or how it was made (e.g., shadows are matter because they are made from light and people). Others made only under-extension errors, but without respecting “natural” categories (e.g., mosquito would be excluded but not dog, or water would be excluded but not milk, or some solids would be left out while liquids were included, etc.) Again, they showed little evidence of a clear precursor concept.

Importantly, Table 11 also shows that Treatment students made substantial progress in developing these more systematic and coherent classifications, with the greatest progress occurring between the end of Grade 4 and Grade 5. Thus, by the end of Grade 5, the

majority (76%) systematically included ALL the solid and liquid materials in their classifications; the majority (63%) also extended these classifications to include some gases. A substantial group (37%) even included all solids, liquids, and some gases without overextending to immaterial entities, although only 6 students (9%) included all and only material entities.

The fact that the biggest improvement occurred between Grades 4 and 5 makes sense, as an explicit concept of matter was only introduced in the Grade 5 Inquiry curriculum in the context of studying water transformations and phases of matter. It was assumed that the foundation laid in the previous two Grades, studying the properties of various solid, liquid, and granular materials, including developing measures of weight and volume and investigating the material's heaviness for size would have prepared them for taking the step of tracing materials across phase change and recognizing some fundamental commonalities between solids and liquids. The data confirm that expectation. It should be noted at no point in the Inquiry curriculum did students engage in a "matter sorting" task of the sort involved here or explicitly engage in contrasting material and non-material entities.

In contrast, there was no significant change in classification patterns among the Control students from Grades 3 to 5, as revealed by either Chi square tests (with the whole sample) or Wilcoxon tests (with the longitudinal sample). This is instructive because Control students also had an explicit unit on transformations of water in the water cycle (in one school in Grade 3, and in the other school at all three Grades). This fact highlights the need to work with students to develop the concept of "matter" itself (and even that water is matter) in order for the topic of the water cycle to make sense as illustrating "phases of matter."

**Table 11: Percent with Different Matter Classifications**

	<b>Matter Classification Includes:</b>	<b>Grade 3 Pre</b>	<b>Grade 3 Post</b>	<b>Grade 4 Post</b>	<b>Grade 5 Post</b>
<b>Treatment</b>	<b>1. All solids &amp; liquids</b>	<b>37%</b>	<b>51%</b>	<b>56%</b>	<b>76%*</b>
	<b>2. All solids, liquids &amp; some gases</b>	<b>25%</b>	<b>36%</b>	<b>36%</b>	<b>63%***</b>
	<b>3. All solids, liquids &amp; some gases while excluding heat, light, shadow</b>	<b>8%</b>	<b>12%</b>	<b>5%</b>	<b>37%***</b>
<b>Control</b>	<b>1. All solids &amp; liquids</b>	<b>--</b>	<b>44%</b>	<b>51%</b>	<b>54%</b>
	<b>2. All solids, liquids &amp; some gases</b>	<b>--</b>	<b>26%</b>	<b>26%</b>	<b>29%</b>
	<b>3. All solids, liquids &amp; some gases while excluding heat, light, shadow</b>	<b>--</b>	<b>8%</b>	<b>8%</b>	<b>6%</b>

Treatment/Control difference Significant, \*  $p < .05$ , \*\*\* $p < .01$ , Pearson Chi Square

We also analyzed what children said as they justified their classifications. At issue was whether children would mention any common properties, and if so, whether those properties would be ones that would allow them to only see commonalities between solids and liquids (hold, touch) or whether they would be aware of properties that not only unite solids and liquids but gases as well (takes up space, has weight.)

Table 12 shows that at the start of Grade 3 children rarely articulated any common properties (19% said matter was something you could hold or touch; 14% mentioned it was something that took up space; no one mentioned weight.) This finding fits with what we observed with children's classification patterns, where most had idiosyncratic classification patterns that didn't show any awareness of a precursor matter concept of something that you can touch, see, and hold. Those who did mention commonalities almost always had more systematic classification patterns.

Among the Treatment and Control students, there was a small but significant increase in the number of students mentioning "hold and touch" as common properties (based on Chi Square analyses with the complete sample and Wilcoxon analyses with the longitudinal subsample). Among the Treatment students (but not the Control) there was also a large and highly significant increase in the number mentioning that matter was something that takes up space or has weight (from 14% who mention taking up space or having weight in early Grade 3 to 50% who mention either taking up space or having weight at the end of Grade 5). When Grade 5 Treatment children's justifications are considered across the properties of tiny things and matter sorting tasks, fully 48% explicitly stated all matter has weight, 43% that matter takes up space, and 60% that matter has weight or takes up space. Such beliefs are relatively uncommon among Control students across both tasks.

**Table 12: Matter Justifications: Common Properties or Characteristics**

		<b>Grade 3 Pre</b>	<b>Grade 3 Post</b>	<b>Grade 4 Post</b>	<b>Grade 5 Post</b>
<b>Treatment</b>	<b>1. Hold, Touch</b>	<b>19%</b>	<b>34%</b>	<b>25%</b>	<b>43%</b>
	<b>2. Take up space or have weight</b>	<b>14%</b>	<b>10%</b>	<b>5%</b>	<b>50%***</b>
	<b>3. Solid, liquid, or gas</b>	<b>5%</b>	<b>4%</b>	<b>9%</b>	<b>9%</b>
	<b>4. Made of Particles, Atoms</b>	<b>0</b>	<b>1%</b>	<b>5%</b>	<b>9%</b>
<b>Control</b>	<b>1. Hold, Touch, Solid</b>	<b>--</b>	<b>23%</b>	<b>24%</b>	<b>43%</b>
	<b>2. Take up space or have weight</b>	<b>--</b>	<b>10%</b>	<b>5%</b>	<b>9%</b>
	<b>3. Solid, liquid, or gas</b>	<b>--</b>	<b>3%</b>	<b>5%</b>	<b>6%</b>
	<b>4. Made of Particles, Atoms</b>	<b>--</b>	<b>5%</b>	<b>3%</b>	<b>0</b>

Treatment/Control difference Significant, \*\*\* $p < .001$ , Pearson Chi Square

### 6-Summary of Results: Matter and material things

In summary, children in the Treatment group made steady progress across five different tasks that probed different facets of a developing concept of matter. Taken together, these results show children in the Treatment group were productively developing an inter-related set of understandings that supported their construction and use of an explicit macroscopic concept of matter in Grade 5. Although their concept of matter is still quite different from the scientists' concept, we believe it provides a foundation on which they can continue to build in the middle school years, as they develop further understanding of the particulate nature of matter and begin to differentiate mass and weight.

In contrast, the progress of the Control students was spottier. Although the majority made progress on the two easier tasks, there was less evidence of progress on the three harder tasks that probed for reconceptualization of weight as linked to tiny pieces of materials and an emerging concept of matter. We believe this may be because existing “topic” driven curricula, often do not address foundational presuppositions and do not provide opportunities for sustained concept building and reorganization over time.

Significantly, we found that the development of understanding that tiny visible pieces have weight and take up space was strongly linked with recognizing the commonalities among solids and liquids and developing the foundations of an abstract concept of matter. This relation held at all ages, in both Treatment and Control Groups. Such linkages need to be exploited in LP curricula.

---

### **B. Understanding Volume**

We noted earlier that students in Grades 3–5 are still working to understand weight, but thoughtful instruction can lead to significant gains. In general terms, progress demands a diminished reliance on subjective judgments (how heavy an object feels) and a growing reliance than on the outcomes of an instrument-mediated process of measurement. Units of measure open up the possibility of operating on (adding, subtracting, multiplying and dividing weights) measures of weight and quantifying ratios of weights (“This object is seven times as heavy as that one”). By objectifying weight, students can more naturally extend the dimension to encompass cases outside the range of human sensitivity. More generally, by quantifying weight, they can systematically investigate the relation of weight to other variables.

One wonders whether something similar may occur in the case of volume.

But volume presents its own set of issues. First, it is more challenging to understand what the relevant invariant is. Research has firmly established (Piaget & Inhelder, 1941) that an understanding of volume follows and most likely builds on the conservation of substance (“After flattening this ball of clay into a pancake, is there as much clay as before?”) and conservation of weight (“Does the clay pancake weigh as much as the ball of clay?”). Volume poses subtleties that are not fully captured in an expression such as “the amount of space an object takes up” or in a formula such as  $v = l \times w \times h$ . Volume combines information related to three spatial dimensions, yet in treating volume as a measurable magnitude, one regards it as one-dimensional: measures of volume can be unambiguously ordered on a number line. Sometimes volume corresponds to space occupied; at other times it corresponds to a void. In still other cases (displacement) it

corresponds to a region of space that other objects or substances are pushed out of or prevented from occupying.

Volume is a measure of size of three-dimensional objects, but not the only one. Consider, for example, how we typically judge the size of a house (floor space) or the size of a desktop computer (footprint). Furthermore the “object” of interest may itself be elusive. What object corresponds to the volume of water displaced by a ship floating in the sea, to space between grains of sand, or to the space between the nucleus of an atom and the innermost electron shell? At times, volume seems to correspond roughly to the amount of stuff. Yet one needs to distinguish volume from amount of material, mass, and weight.

Because of these subtle issues, we assessed students’ understanding of volume through several measures. Each measure would provide some insight into the acquisition of the concept. The data are helping us to determine which aspects are more amenable to influence by instruction and help clarify several inroads to development that could be exploited in future studies.

We used the following tasks to assess students understanding of volume at each point in our longitudinal investigation:

- **Block Rearrangement task:** The student judges whether two arrangements of transparent cubes (each having the same number of cubes but arranged in different shapes—a 2 x 3 x 2 “block” and a 2 x 6 x 1 “snake”) take up the same or different amounts of space. At issue is what invariant (number of blocks, length or thickness of objects, number of faces of blocks etc.) they focus on in making their judgments. We assumed this would be the easiest volume task. Because the objects are already “visually structured” as composed of equal units, children do not need to impose an “unseen units structure” on the objects.
- **Clay Conservation of Volume task:** Children are shown two identical clay spheres and watch as one is flattened into a pancake. The student is asked whether the sphere and the pancake have the same amount of clay, whether they weigh the same, and whether they take up equal amounts of space. This task is more challenging because the invariant of volume is less salient than other spatial dimensions (area, length) and cannot be judged directly. At issue is whether children have some emerging theoretical beliefs linking the volume of the object to the amount of material in the object, and reason that because the object has only been reshaped, with nothing added or taken away, the volume is unchanged.
- **Sizing-Up Task (Volume Invariant Task):** Students are asked to compare the sizes of two three-dimensional rectilinear objects of different shapes (a wooden block 2 x 3 x 2 inches and a purple fine-textured foam block 2 x 8 x 1 inches). Then they are asked to measure the blocks using a set of 1-inch cubes (20 are provided). The purpose was to determine whether students used the cubes to measure the volume of the blocks and, if not, what invariants (properties) they were basing their judgments on. This task is much more challenging than the Block Rearrangement task in that there are many competing invariants they could attend to in this situation (such as object length, perimeter, area of a face, or even surface area.)

- **Water Displacement Task:** Students are shown and invited to handle a brass and an aluminum cylinder of the same size and shape, and of clearly different weights. The student predicts and attempts to explain what would happen if each cylinder were completely submerged in a beaker of water. (Only the aluminum cylinder is actually placed in water.) At issue is what invariant the student thinks is relevant to water displacement: object size or object weight. If students are going to use water displacement as a measure of object volume, they first need to understand (theoretically) that it is object volume, not object weight, that is relevant to water displacement when one has objects that can be fully submerged in water. Prior research suggests that students initially think weight rather than object size or volume is relevant to water displacement, further evidence that children need to develop and revise relevant theoretical beliefs about volume as part of the process of learning to measure object volume.
- **Area Task:** Students are shown two rectangular cards of different shapes (a 2x4 inch blue rectangle and a 3x3 inch orange rectangle) and are asked, “Which card is bigger? (... covers more space on the table)” Then they are asked to “... measure how much space each card covers...” and invited to use any objects on the table, including a set of 1x1 inch tiles. At issue is how children spontaneously use the tiles in this task, and hence what invariant they focus on: Do they use the tiles to cover each card or do they use the tiles to measure a length or perimeter? We expected that measuring area would be much easier than measuring volume (because the relevant “invariant” is more directly given perceptually and does not call for as complex a structuring of an array), and that children who were able to focus on volume in the Sizing-Up Task would have appropriately measured the area of the cards.

Table 13 (below) provides a broad overview of the results. Pre-G3 results are included as a baseline. Significance tests correspond to comparisons of Control Group responses to Treatment group responses. Two independent judges who were blind to the student’s group coded the volume and area invariant tasks. In all tasks<sup>5</sup>, an interjudge reliability of at least 85% was obtained. Disagreements were settled by either a third judge or by agreement of the original two judges. The Blocks, Clay and Water Displacement task represent children’s final judgment on each task after being probed for their reason. In these tasks, children’s judgments were almost always accompanied by a relevant justification (over 95% of the time) and different judgments were accompanied by qualitatively different justifications. For this reason, objective judgment data were deemed good indicators that children were attending to different invariants.

The percentages are to be interpreted as follows. In the upper leftmost cell, the value 32% indicates that prior to the Inquiry Project intervention, at the beginning of Grade 3, 32% of the students judged that the two arrangements of transparent cubes took up the same amount of space. Collapsing the results from the end of Grades 3-5, 54% of the Control students responded correctly on the same task, whereas 60% of the treatment students responded correctly over the same Grade levels. This difference of 4% was not

---

<sup>5</sup> Similar standards of interjudge reliability were used throughout the study unless otherwise indicated.

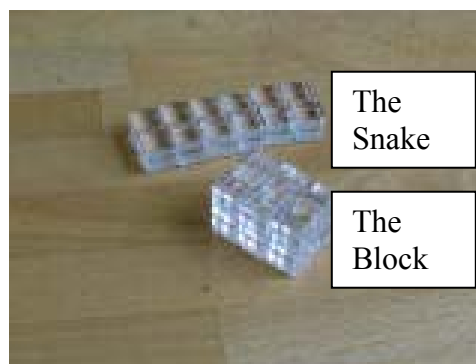
statistically significant. On each of the remaining four tasks, the Treatment Group significantly outperformed the Control Group. We will now inspect the results more closely.

**Table 13: On all Volume Tasks except Block Arrangement, the Treatment Group significantly outperformed the Control Group in Grades 3–5.**

	<b>Block Rearrangement</b>	<b>Clay Volume</b>	<b>Sizing Up (volume)</b>	<b>Water Displacement</b>	<b>Area</b>
<b>Percent Correct</b>					
Pre G3	32%	25%	5%	10%	46%
Control (G3-G5)	54%	20%	15%	13%	61%
Treatment (G3-G5)	61%	46%	29%	40%	77%
<b>G3-5 Treatment vs. Control</b>					
Pearson Chi-Square Test	--	20.59	7.95	23.73	7.97
Significance	n.s.	p< .0001	p=. 005	p < .0001	p=. 005

### 1-Block Rearrangement task

The two objects in the blocks in the Rearrangement Task were composed of the same number of transparent blocks of fixed size, just arranged in different shapes (see Figure 1). They were brought out fully assembled and children were asked: Do these two objects take up the same amount of space (i.e., have the same volume) or not? Figures 2 and 3 show the changing judgments by Grade separately for the Treatment and Control students.



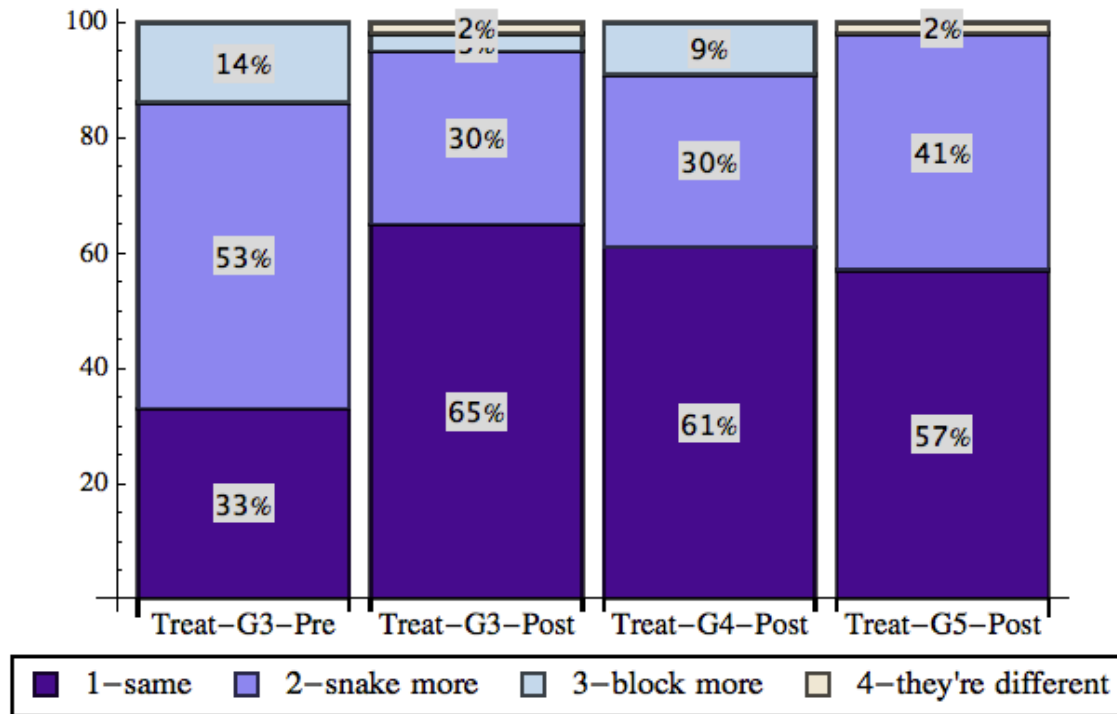
**Figure 1: Block Rearrangement task stimuli**

Quite strikingly, at the time of the pretest for the Treatment students (early Grade 3), only 32% of the children judged the objects had the same volume. Those who gave a “same volume” judgment argued either they had the “same number of blocks” or that one of the

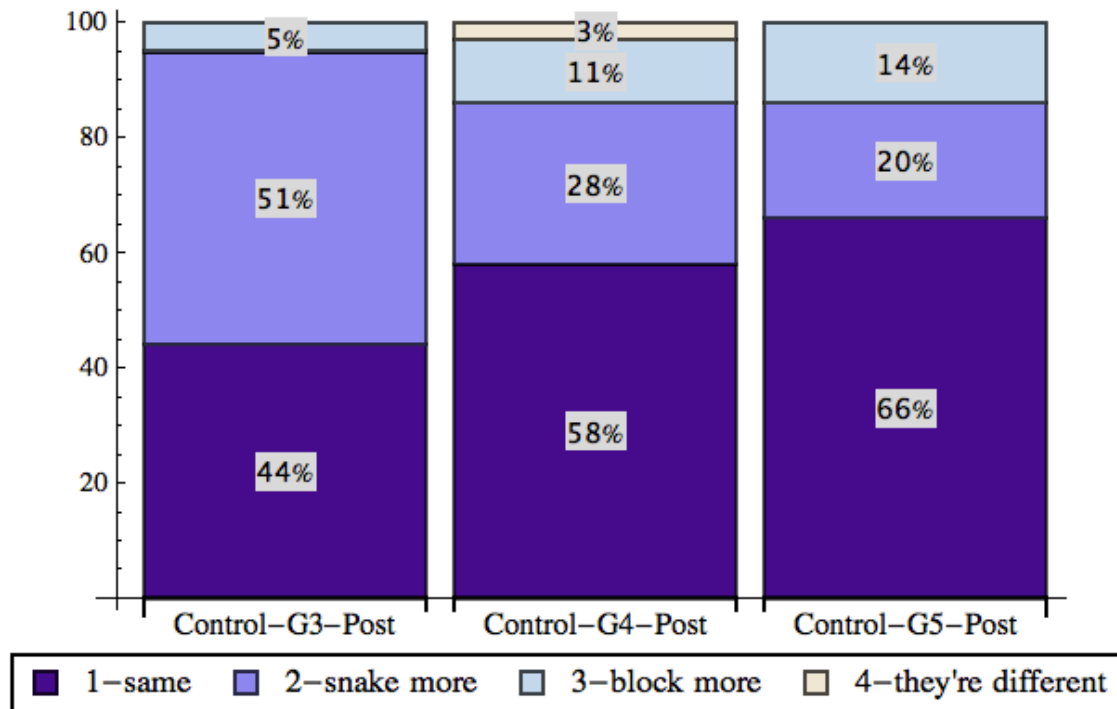
shapes could be folded to make the other. Thus, even though the shapes were obviously composed of unit cubes, most children at the beginning of Grade 3 didn't focus on this invariant as relevant to judging taking up space (many more focused on it as relevant to judging the amount of plastic). Instead, over half of the children judged that the longer object (the snake) took up more space, typically focusing on its greater length. By the end of Grade 3, the judgment pattern had dramatically changed for the Treatment students: now over 60% judged the two shapes had the same volume (a significant increase). The percent of students making this judgment remained essentially the same across Grade 4 and 5 as well.

The improvement at the end of Grade 3 by the Treatment students may reflect the impact of the Grade 3 Inquiry Curriculum, as the performance of Grade 3 Treatment students was significantly better than Control (65% vs. 44%,  $\chi^2(1, N=99) = 4.78, p = .03$ ). This interpretation makes sense given that a *cubing method* (making models of three dimensional objects with unit cubes) was introduced in the Grade 3 curriculum as a means of determining the volume of objects.

Surprisingly, the Control students improved from Grade 3 to Grade 5, but the Treatment students did not make further improvement, so that by Grades 4 and 5, the difference between Treatment and Control was no longer significant. The fact that the Treatment students did not continue to improve after Grade 3 on this task was somewhat surprising, as volume continued to be an important target in the Grade 4 Inquiry curriculum as well, where children learned to measure the volume of liquids and use water displacement to compare the volumes of irregularly shaped solids. Overall, the fact that almost 40% of Grade 5 Treatment and Control students focused on other spatial attributes (e.g. length, height, area) rather than volume highlights the challenges in helping students construct a robust concept of volume that is differentiated from other spatial attributes.



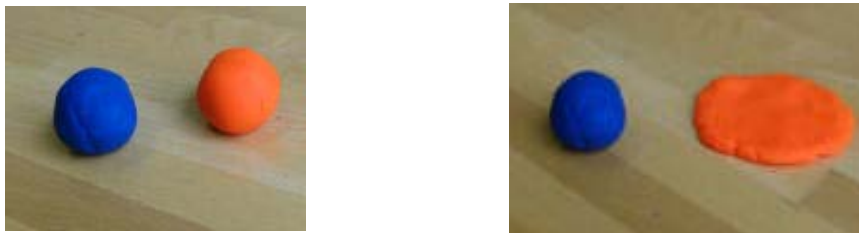
**Figure 2: Block Rearrangement Task (Treatment Group): Does one arrangement of blocks takes up more space?**



**Figure 3: Block Rearrangement Task (Control Group): Does one arrangement of blocks takes up more space?**

2-Clay sphere vs. pancake task

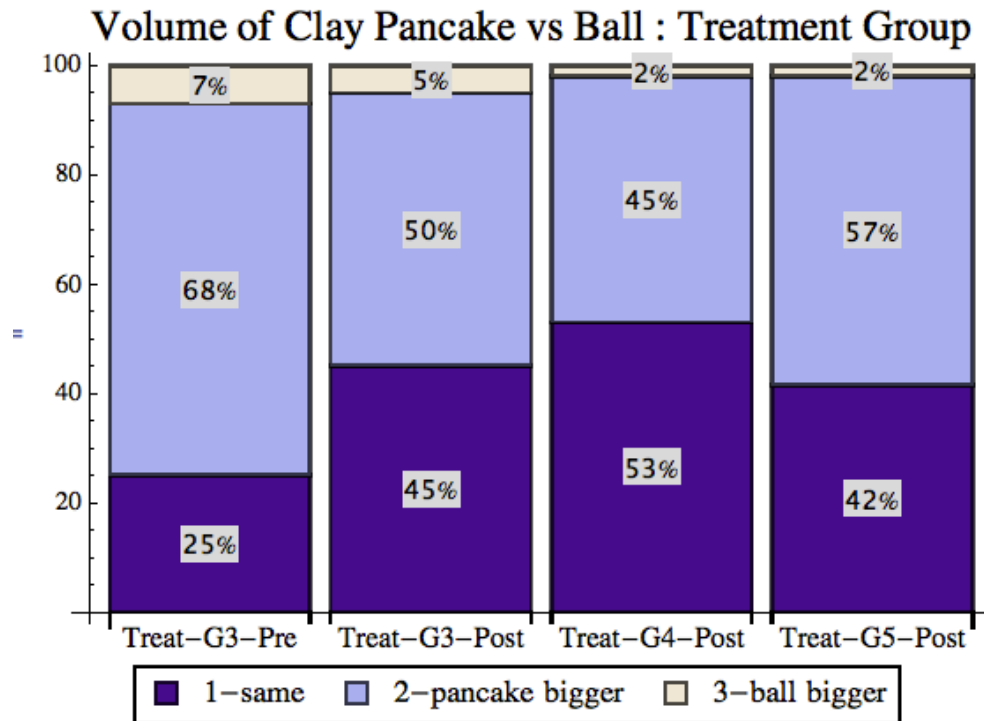
A second volume task used deformable material, clay (see Figure 4). Children were asked, “Do these objects take up the same amount of space (that is, have the same volume) or not?” They were also asked whether they have the same amount of clay, the same weight, and would balance on a scale; results of those questions about amount of weight were reported in the previous section. We expected this task to be far more challenging than the Block Rearrangement Task since no explicit units are given so students could not count to determine the volumes of the two objects. Without explicit units, the attribute of volume would presumably be less salient than area or shape. Children do, however, actually witness the transformation of a sphere into a pancake, so if they envision one is simply spatially re-arranging a fixed amount of stuff, they might reason that the ball and pancake must still take up the same amount of space because they both have the “same amount of material.” (It would have been useful if we had administered the Block Rearrangement Task in a way that would have allowed the student to witness the transformation of the “snake” into the “block.”)



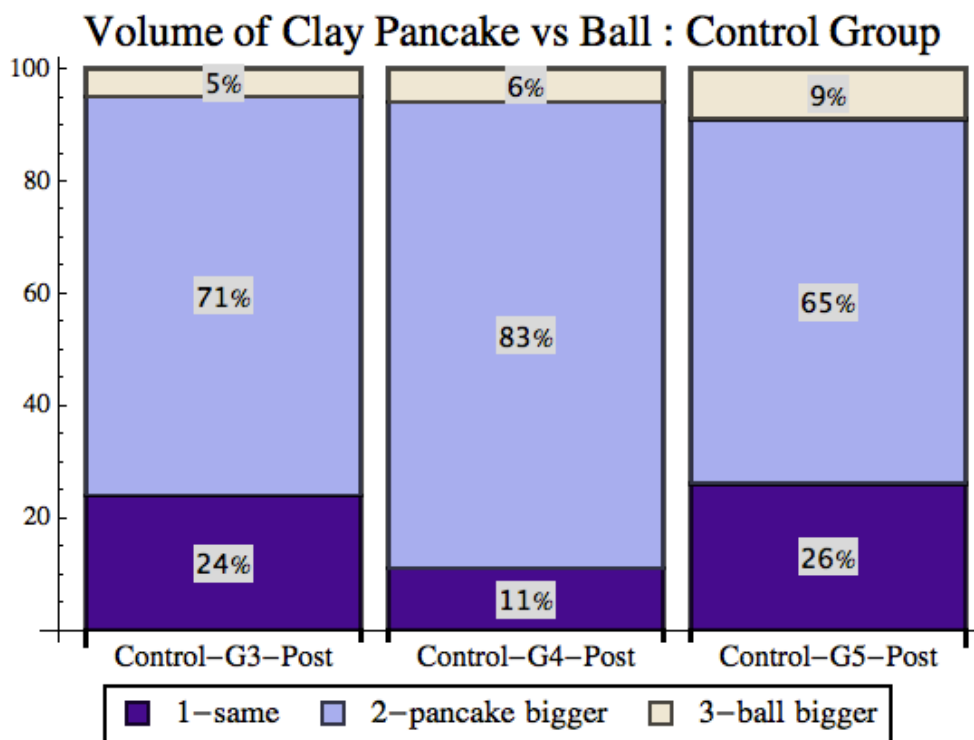
**Figure 4:** Conservation of Clay--“Do these objects take up the same amount of space?”

Figures 5 and 6 show the responses by Grade level for the Treatment and Control students. At the beginning of Grade 3 (pre-Treatment), most children judge that the pancake takes up more space than the ball; indeed, only about a quarter of the students say that they take up the same amount of space. For most, this is not because they don’t understand they still have the same amount of clay, as 78% of the students judge that they have the same amount of material. The difficulty, therefore, lies in knowing which spatial invariant we are talking about (volume rather than area or length), and in having the relevant theoretical beliefs that link volume to amount of material. Treatment students made significant progress in realizing the shape change did not affect object volume by the end of Grade 3. By the end of Grade 4, over half the students (53%) judged that the ball and pancake took up the same amount of space. This is very close to the percent (60%) that judged the block and snake had the same volume in the Blocks Rearrangement task, but still much less than the 88% who judge that the ball and pancake have the same amount of material, same weight, and will balance. Clearly, Treatment children are coming to link volume and amount of material, although distinguishing

volume from length or area remains a problem for some children. Indeed, chi square tests showed making “same volume” judgments on the two tasks was highly related for the Grade 3-5 children ( $\chi^2(1, N=174) = 41.53, p < .0001$ ).



**Figure 5:** Volume of Clay Task (Treatment group)



**Figure 6:** Volume of Clay Task (Control Group)

Treatment students did not continue to improve in Grade 5. We wonder whether this may be due to the fact that the activities in Grade 5 turned away from a focus on the volume and density of solids and liquids to a focus on matter transformations and the material nature of gases; indeed where volume was introduced it was often in ways that undercut its relation to amount of material (e.g., volume changes in water with freezing and melting, despite the fact that the amount of material stays the same; ideas of compression and packing of particles.)<sup>6</sup>

The pattern observed among the Control students was quite different. Here there was no systematic change from Grade 3 to Grade 5 in children's understanding that the ball and pancake, although different shapes, had the same volume; indeed their levels of "same volume" responding remained at the very low level we had observed among the early Grade 3 (pre-treatment) students. Thus, although Control students made progress in focusing on the invariant of volume in the Block Rearrangement task, they did not carry that over to less obvious situation with the clay, where they focused primarily on the greater area covered by the pancake.

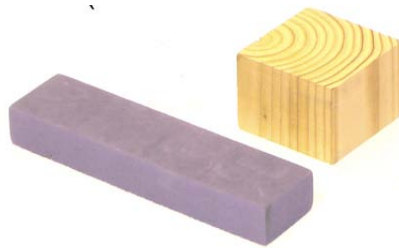
<sup>6</sup> In developing the Grade 5 curriculum we found ourselves at a cross-roads, with two separate routes to take: one on a more mathematical front that would have consolidated volume and density and the other on a more scientific front that explored phase change and matter. The Grade 3 and 4 curricula had prepared students to take either step next, but, in part because we were unsure of the mathematical sophistication of our students and were not simultaneously intervening with their math curriculum, we decided to take the latter route. Both are important, but couldn't be done at the same time.

Treatment students may have improved more than the Control students on this task because the Grade 3 Inquiry curriculum promoted an explicit belief linking the volume of an object to the amount of material. Among the Treatment students, the most common reason for concluding the volume was the same was that they have the same amount of clay; students also made other Piaget arguments as well, such as compensation (it's flatter but wider) and reversibility (you could roll it back into a ball) arguments.

### 3-The Sizing-Up (Volume Invariant) Task

What invariant do students focus on when measuring the size of objects? We created a simple third task that would allow us to identify the properties students used to judge the size of objects from their explanations and actions as they engaged in working with a variety of tools in devising a measurement. We devised the task in such a way that children could be successful on this task without knowledge of technical vocabulary or formulas.

The student was asked to compare the “amount of space” filled or taken up by two solid blocks (Fig. 7). The interviewer used sweeping gestures to emphasize three-dimensional space as opposed to space on the table surface. In the task, the wooden block measures 2”x3”x3”, thus taking up 18 cubic inches; the purple foam block measures 2”x8”x1” and so takes up only 16 cubic inches. So the block with the greater volume has a smaller surface area and a shorter length. (The student was not given these measures.)



**Figure 7: Volume Judgment**--Which of the above blocks is bigger? The purple, oblong block measures 2”x8”x1” (16 cu. in.); the wooden block measures 2”x3”x3” (18 cu. in.). Students were not told these measures. However they judged the sizes using a set of small cubes (approximately 1” on each side).

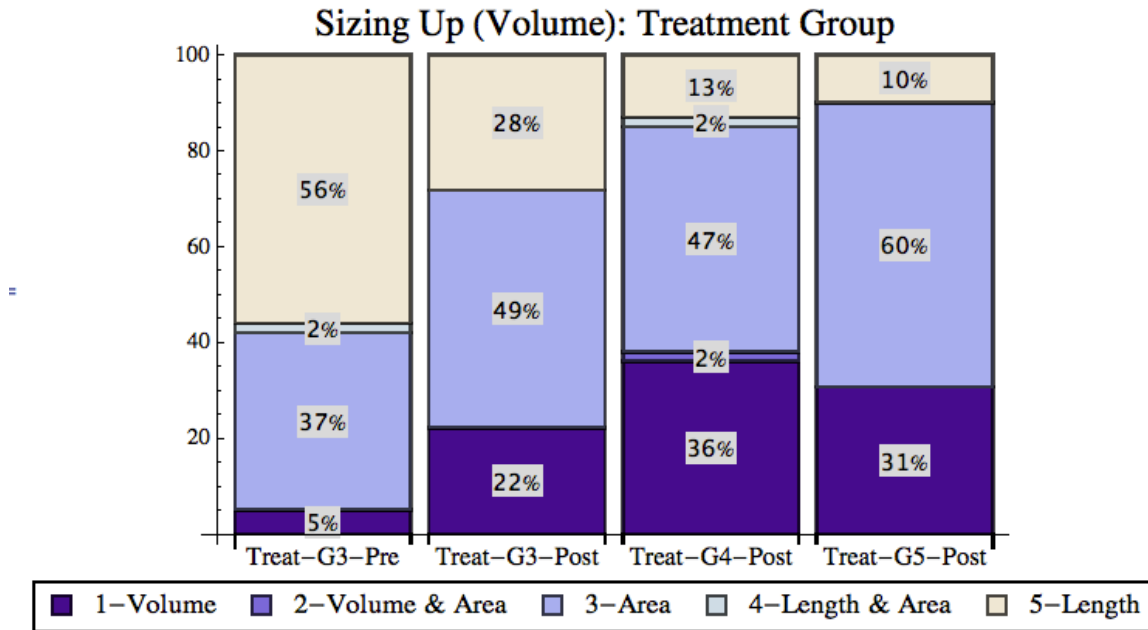
First, we asked them to measure the objects’ size, using methods and instruments of their choosing. A variety of materials were available, such as a broken tape measure, tiles, different size paper clips, and small (approximately 1 inch) cubes. Then, all students who did not spontaneously use a small cube to measure the blocks were prompted to do so. This task was much more demanding than the Block Rearrangement task, as students had to figure out a way to use the cubes to make a measurement. Further, if they chose to use

the cubes to make a little replica of the object and count the cubes, they would have to create a separate object with the cubes, rather than act on the object directly (such as when they use the cubes to cover one face of the object or surround the object.) From their explanations and actions in the use of a cube as tool we inferred the properties they attended to.

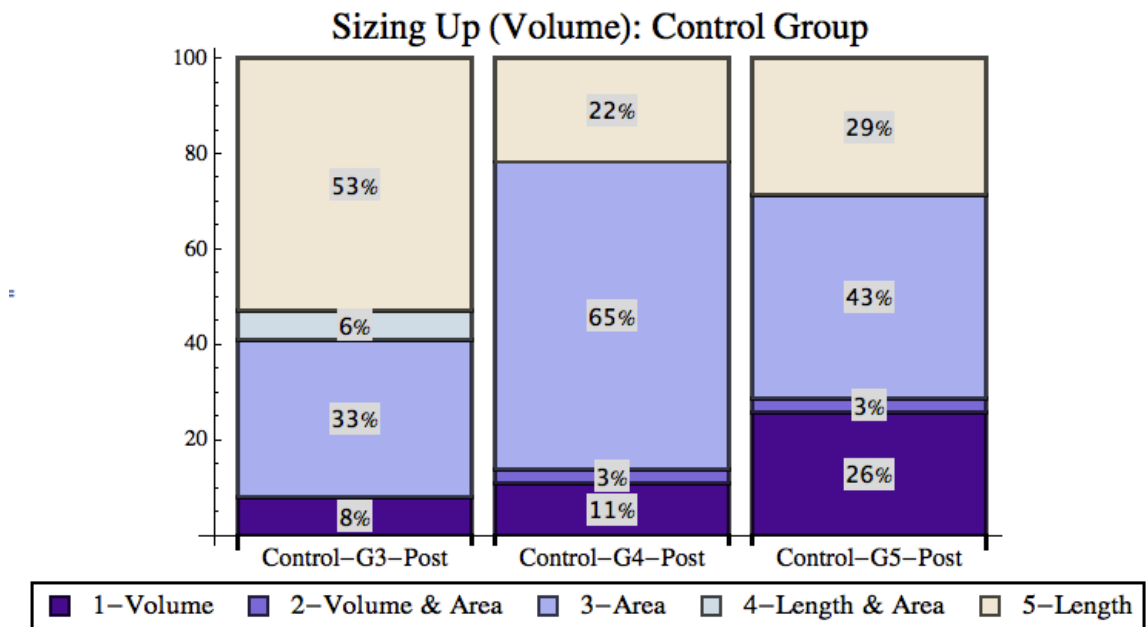
One might expect that three-dimensional cubes might induce students to measure the size of the blocks by the sum of the volumes of the cubes. However this was not the case. Some students used the cube edges to measure the length or perimeter of the blocks. Others overlaid cubes on one or more facets of the blocks, thereby showing that they were attending to the area of one or more surfaces of the blocks. (There was admittedly some doubt about whether some students were judging the purple foam block by area or by volume, because it was approximately one cube high. However, it was easy to dispel such doubts when we considered their judgments for the wooden block, which measured approximately 3x3x2 inches.)

The bar charts below (Figure 8, 9) show the overall results on the Volume task for students by Grade level for Treatment and Control students separately. The bottom (purple) segment corresponds to the proportion of students whose judgments were based on volume. We counted responses as based on volume if students made replicas of each object using the cubes, and counted the cubes (one of the most common approaches coded “volume by decomposition”); used the cubes to find the length of each side, and then multiplied the three lengths, or used the cubes to cover a face and then multiplied by the height (a less common method, called “volume by multiplication”). Both these methods resulted in the correct judgment that the wood block in fact took up more space than the purple, as well as (roughly) correct numeric values for the two volumes. In contrast, children who focused on length or area most commonly judged that the purple block took up the most space. (Children coded as “Volume & Area” or “Length & Area” did not consistently focus on the same invariant for each block: rather, they switched from “Volume” to “Area” or from “Area” to “Length.”)

The Inquiry Treatment students improved significantly and steadily from the beginning of Grade 3 (when only 5% made correct volume measurement) to the end of Grade 3 and Grade 4 when 22% and 36% made correct measurements. They also performed better than Control students at the end of Grade 3 and Grade 4, both Grades where volume was given special attention in the Inquiry curriculum. For students in the Control Group only 8% or 11% focused on the volume of the objects (as opposed to their areas or lengths) when measuring their sizes. The difference between Treatment and Control approached significance at the end of Grade 3 ( $\chi^2(1, N=67) = 3.4, p = .06$ ) and was highly significant by the end of Grade 4 ( $\chi^2(1, N=55) = 7.49, p = .006$ ). However, Treatment students had lost this advantage over the Control Group by the end of Grade 5. In other words, the experimental effect observed in Grade 4 was not sustained through Grade 5. This may reflect the fact that the Grade 5 Inquiry curriculum placed less emphasis on volume as a 3-dimensional quantity (e.g. using uni-dimensional measures of volume of liquid in graduated cylinders and 2D animations of particle motion), and did not explicitly contrast volume and area. Clearly most students (both Treatment and Control) were still focusing on other spatial invariants than the intended one—evidence of the challenges students face in constructing an understanding of volume.



**Figure 8:** Invariants used in judgments of size of objects (Treatment Group)



**Figure 9:** Invariants used in judgments of size of objects (Control Group)

As expected, using the cubes to make a volume measurement of the two blocks on the Sizing Up task was much more demanding than simply comparing the volume of two objects composed of equal size cubes in the Blocks Rearrangement task. Whereas approximately 60% of Treatment students succeeded on that task by the end of Grade 3, only 22% were able to make a volume measurement on the Sizing Up task by then, and only a little over a third ever succeeded at making a correct volume measurement. Performance on the two tasks (Block Rearrangement and Sizing Up task) were highly related for both Treatment and Control students, however, in that: (a) students who did not focus on Volume in the Blocks Rearrangement task, almost never focused on volume in the Sizing Up Task; and (b) students who focus on volume in the Sizing Up task, almost always had focused on volume in the Blocks Rearrangement task.

#### 4-Water Displacement Task

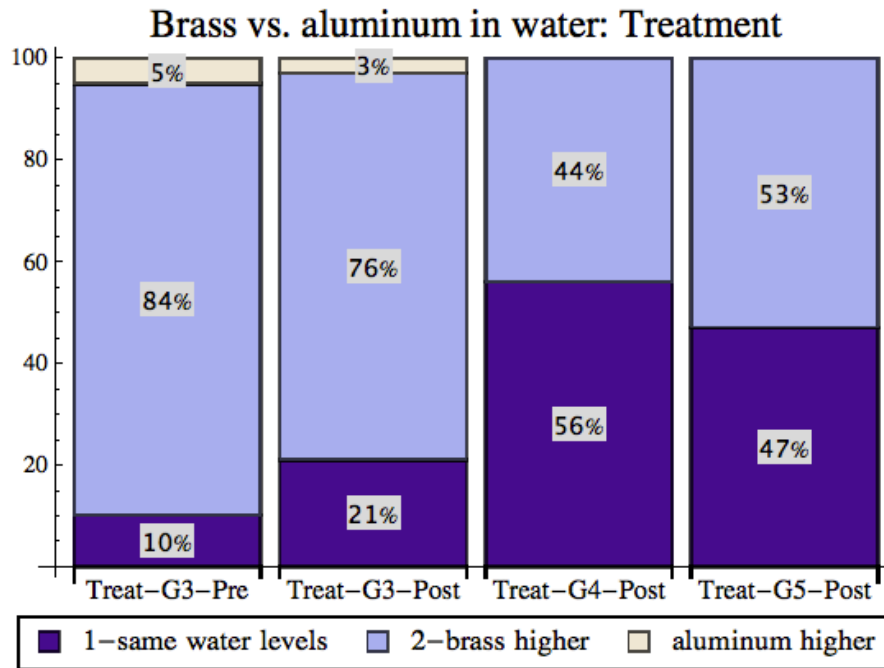
Another way to assess children's understanding of volume is to assess children's understanding of physical contexts in which volume is relevant. Water displacement is one such context. In the Water Displacement task we assessed whether young children knew that object size (not object weight) affected how much water would rise when an object was completely submerged in water.

Children were shown two identical cups of water and shown and invited to handle two cylinders of equal diameters and heights, but of quite different weights: one was made of brass and the other of aluminum. The children were first asked to predict what would happen (i.e., the water level would go up) when the aluminum cylinder was completely submerged in water. Then the interviewer placed the aluminum in water, marked the new water level, and asked what would happen if the brass cylinder were placed in the other cup of water: Would the level rise higher than it had with the aluminum, to the same level, or lower?

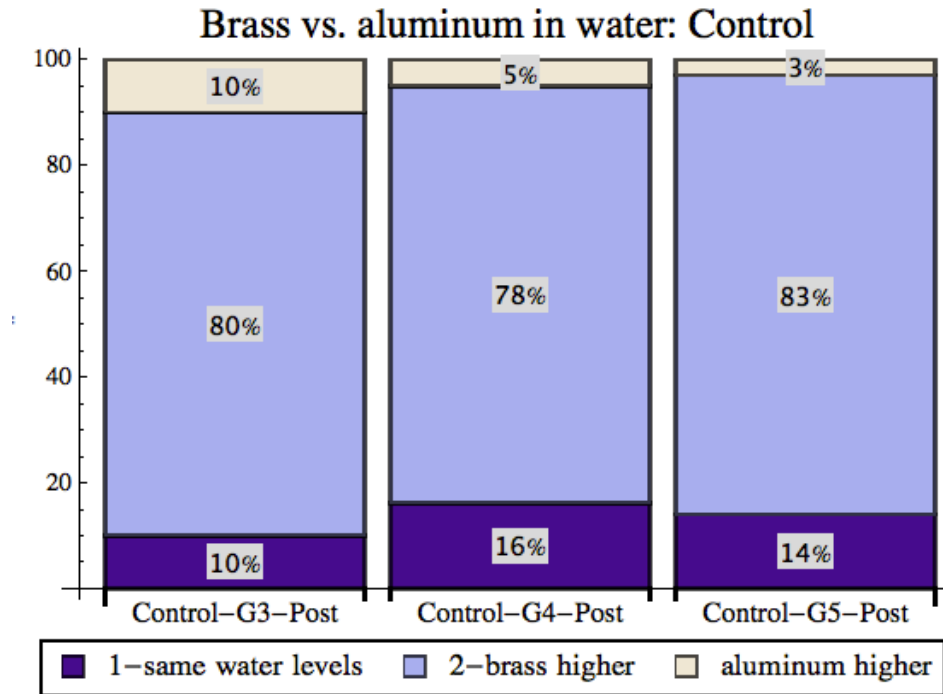
Figure 10 shows predictions made by the Treatment students as a function of Grade. In early Grade 3, prior to the start of the Inquiry Curriculum, only 10 % predicted the two cylinders would raise the water by the same amount because the objects were the same size. Almost 90% of children predicted that the immersed brass would result in a higher water level than the aluminum would, because of its greater weight. That is, most children focused on weight (or the amount of force) pushing down on the water as the relevant variable. As part of the Inquiry Curriculum in Grade 4, children directly investigated what variable was relevant to water displacement and by the end of Grade 4 the majority (56%) now focused on object size rather than object weight as the relevant variable. The new insight that object volume rather than object weight predicts the amount of water level rise was generally maintained by Treatment children (47%), even though this issue had not been discussed in the Grade 5 curriculum.

In contrast, Figure 11 shows that Control children had little insight about this fundamental issue, with their responses essentially remaining unchanged across Grades 3-5. That is, only a small percent (10%, 16%, 14%) judged that the two cylinders would raise the water levels by the same amount. By Grade 4, the Treatment students were

doing significantly better than the Controls ( $\chi^2(1, N=92) = 13.64, p < .0002$ ), an advantage maintained in Grade 5 as well ( $\chi^2(1, N=89) = 9.74, p = .002$ ).



**Figure 10:** Predictions as to whether brass and aluminum cylinders would make the water level rise to the same or different levels (Treatment Group)



**Figure 11:** Predictions as to whether brass and aluminum cylinders would make the water level rise to the same or different levels (Control Group)

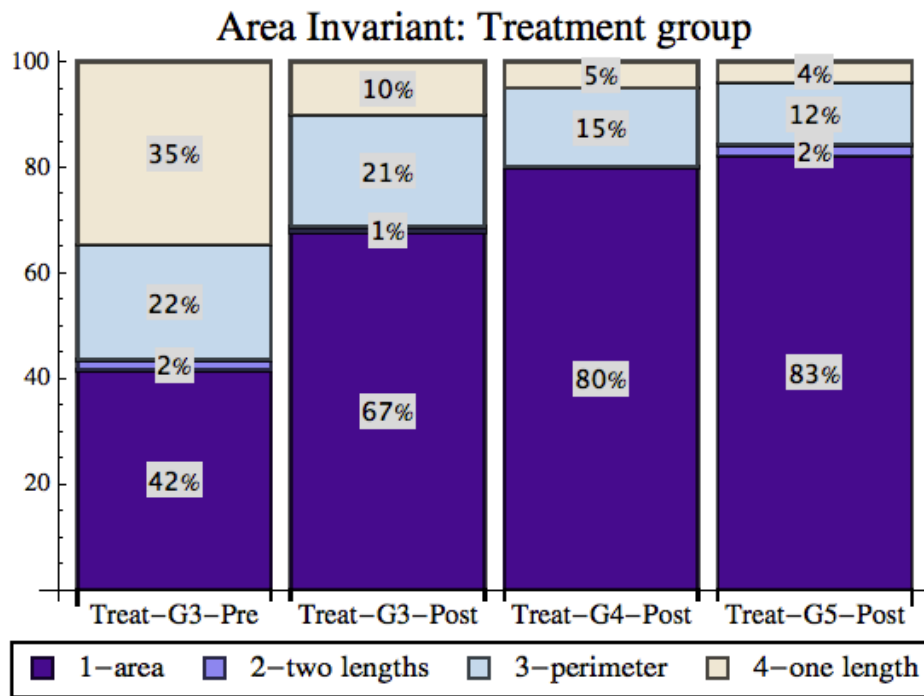
### 5-Area Task

A final task probed children's ability to use tiles to measure the areas of two rectangular cards of different shapes (2 x 4 inch blue card and a 3 x 3 inch orange card). From their actions and explanations we inferred what invariant they focused on when asked to determine which card took up more space on the table. Some children used the tiles to measure the length of a side; others used the tiles to measure a full or partial perimeter. Still others showed they focused on area, by either (a) entirely covering the two cards with tiles and counting the tiles, or (b) making a 2 x 4 array and a 3 x 3 array and multiplying.

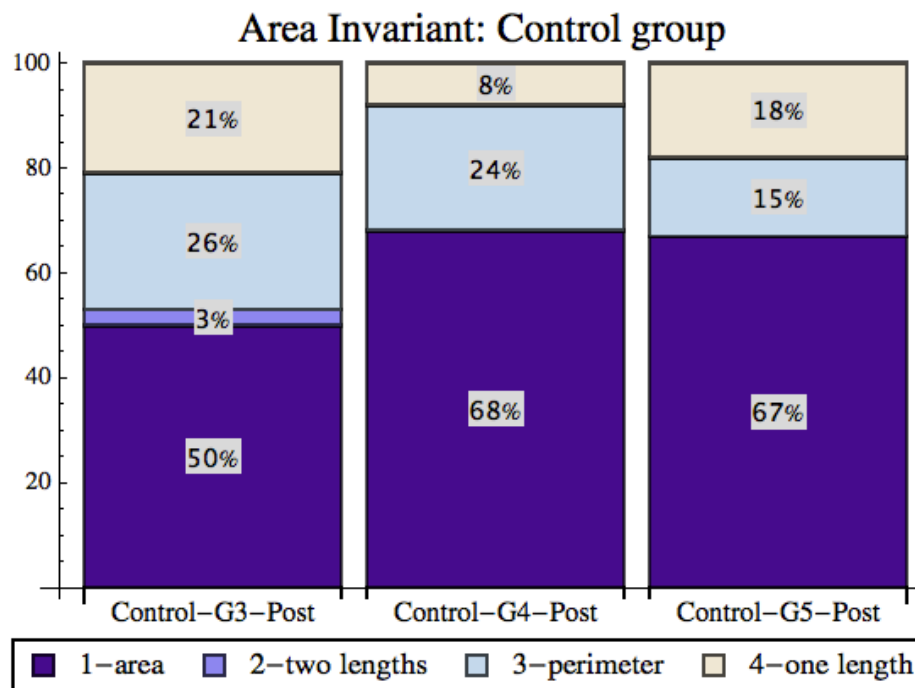
Figures 12 and 13 show the overall results on the Area Measurement task for students by Grade level for Treatment and Control students separately. The bottom (purple) segment corresponds to the proportion of students whose judgments were based on area. In general, children found it much easier to focus on area than volume, although at the start of Grade 3 only 42% of children compared the area of the two cards when asked which took up the most space; the others focused on comparing the length of sides or the perimeters of the two cards. Although the Inquiry curriculum didn't focus on measuring areas, Treatment children improved dramatically in making area judgments across Grades 3-5, so by Grade 5 83% were using the tiles to measure the area of the cards.

In contrast, the Control students made more limited progress. At the end of Grade 3, they looked more like the early Grade 3 Treatment students, with 50% measuring area; by the end of Grade 5 67% were measuring area. The overall difference between Treatment and Control students on the Area Invariant task was significant across Grades 3-5.

Why might the Treatment children have had an advantage on area, given that it was not a focus of the curriculum? It may be that their practice using cubes to measure volume (which as we've seen they often misinterpreted as measuring area) enhanced their understanding of area. Interestingly, we found highly significant relations between the invariants children attended to on the Area Invariant task and the Volume task. 90% of the children who focused on length or perimeter in the Volume task also focused on length or perimeter in the Area task. 90% of the children who focused on area in the Volume task also focused on Area in the Area task. Finally all but 2 children who correctly used cubes to measure volume had correctly used tiles to measure area. Thus, there was a progression from students who focused only on length measures of spatial extent, to those who distinguished length and area measures, to those who distinguished length, areas, and volume measures.



**Figure 12:** Invariants used in judging size of cards (Treatment Group)



**Figure 13:** Invariants used in judging size of cards (Control Group)

#### 6-Summary of volume results

In summary, our results revealed that both Treatment and Control students face numerous challenges in learning about volume and volume measure that center on coming to focus on the relevant invariant. When we discuss “taking up” or “filling up” space, it is not obvious to students that we are referring to a 3D sense of volume rather than a 2D sense of how much space a surface “covers.” Indeed children find the spatial attributes of length or area much more salient than volume, perhaps because these attributes are more accessible perceptually. Further, when children think about water displacement, they think simply of the water being pushed up by the object, with the amount of water pushed related to the force or weight of the object, rather than seeing the water as being displaced by immersing the object in the water, with the amount of water displaced proportional to object volume.

The Inquiry Curriculum did help students make significant progress with these issues and with understanding multiple aspects of volume: across all four of our volume tasks, the progress made by the Treatment group from Grade 3 to Grade 5 was highly significant. In contrast, the Control Group’s progress was more limited and spotty, showing significant progress on only the Blocks Rearrangement and Sizing Up tasks) and making no progress at all on the Volume of Clay and Water Displacement tasks. Further, Treatment students did show enhanced understanding of volume in comparison with Control students on three of the four volume tasks, as well as enhanced understanding of area (see Table 13). At the same time, our results suggest that only about a third of the Treatment students (33%) had a robust understanding of volume at the end of Grade 5, consistently focusing on volume rather than some other invariant across three or four of the tasks. Although this is significantly more than the Control Group, where only 8% had

such a robust understanding, there obviously is considerable room for improvement, especially in helping students distinguish volume from area. The Inquiry Curriculum could be improved by discussing these contrasts more explicitly with students throughout the curriculum.

---

### **C. Understanding Weight, Size, and “Heaviness” of Materials**

Central to a sound macroscopic understanding of matter is an understanding that the weight of a solid object made of a single material depends both on the volume of the material and on the particular kind of material. Consequently, we developed three tasks to probe how students deploy their knowledge of weight and volume in their emerging understanding of density. More specifically, the tasks were designed to determine:

- How children explained the fact that some smaller objects are heavier than larger ones and that objects of different sizes could be the same weight.
- Whether children could consistently distinguish between the weight of an object and the heaviness of the material from which it was made.
- Whether they could infer, from the size and weight of two solid, homogeneous objects, whether those objects might be made of the same material.

None of the tasks required knowledge of formulas or technical terms. Each, however, called for use of the idea that objects made of different materials vary in their heaviness for size in tasks that called for explanation, inference, and problem solving, and that were quite different from any curricular activities used in the Inquiry Curriculum.

At the start of Grade 3, children were inclined to confuse weight and density on each of the tasks. This is consistent with our past research and that of others. However between Grade 3 and 5, the Treatment students, but not Controls, made significant and steady progress on each of the tasks, suggesting that the Inquiry Curriculum was successful in promoting these understandings.

#### 1-Larger is not always heavier

In Part 1 children are shown cylinders of equal diameters, each made of a solid piece of a single material and wrapped in white paper. Two (A & C) are short; the other two (B & D) are three times as tall. B & C are made of aluminum, A is made of brass, and D is made of a polypropylene, a light plastic (see Table 14 for a description of the four pairs used in the four problems). At issue is how children contend with the fact that size does not always serve as a good predictor of the weight of objects: bigger does not always signify heavier. We wanted to explore how they link size to weight and also see whether they refer to the kind of material in seeking to account for size-weight ‘anomalies’, cases where size differences did not accord with weight differences. These tasks are a means of probing students’ intuitions about density before they had learned terminology or formulas for density.

In Problem 1, (see Table 14 for each of the four problems) cylinders B and C are both made of aluminum, so the larger object is the heavier one. This is fully consistent with the notion that larger objects are heavier and, not surprisingly, students almost always explain the difference in weight by referring to their differences in size. In the next three comparisons, size and weight do not co-vary: (a) in Problem 2, one object is heavier although they are both the same size; (b) in Problem 3, the smaller object is much heavier than a much larger object; and (c) in Problem 4, both objects weigh the same, although they are quite different in size. Problems 2 and 3 are useful in giving evidence that children are developing generalizations linking “heaviness” to kind of material. However, Problems 2, 3, and 4 require students to differentiate the weight of the object from the heaviness of the kind of material. Hence, children’s explanations on this problem are the most critical to determining whether they are beginning to differentiate weight and a precursor of density.

**Table 14: Four comparisons used to determine how students contended with weight-size 'anomalies' (problems 2-4)**



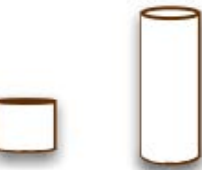

 <p>B Tall alum    C short alum</p>	<p>In <u>Problem 1</u>, the student is shown a large covered object (B) and a much smaller one (C) and asked: (1) <i>Are these the same size or different?</i> (2) <i>Are they the same weight or different?</i> After noting that B is both bigger and heavier than C, and comparing the weights on the balance scale, the student is asked to account for the difference in weight. Most students refer to size to account for the difference in weight.</p>
 <p>A Short brass    C short alum</p>	<p>In <u>Problem 2</u>, children are shown two equal-sized cylinders (A and C) that differ markedly in weight (A is much heavier than C). After comparing their sizes and weights, the child is asked: <i>How come they are the same size but don't weigh the same?</i> Should the child suggest that cylinder is hollowed out, he is reminded that they both are solid pieces.</p>
 <p>A Short brass    D tall Poly</p>	<p><u>Problem 3</u> concerns objects of different sizes (A and D) where the smaller cylinder is much heavier than the other. The student is asked, <i>How can the smaller one be heavier?</i></p>
 <p>A Short brass    B Tall alum</p>	<p>In <u>Problem 4</u>, children are shown two different size objects (A and B) that have the <i>same weight</i>. After predicting what will happen on the balance scale and then checking their prediction, children are given the final puzzle: <i>How can they be such different sizes yet weigh the same?</i></p>

Table 15 presents the percentage of children who consider that the two objects might be made of different materials, or that one object is made of a heavier kind of material than the other, in explaining why size and weight do not covary in Problems 2–4. In addition, it shows the percentage of children who coordinate the factors of being made of different materials and having different sizes in explaining how the two objects way the same in Problem 4. This material/size coordination (in their explanations) is good evidence that they are beginning to differentiate weight and density.

**Table 15: Percent of Children Who Use Difference in Material/Heaviness of Material to Explain Why Weight & Size Do Not Covary in Problems 2 to 4**

	<b>Problem &amp; Explanation</b>	<b>Grade 3 Pre</b>	<b>Grade 3 Post</b>	<b>Grade 4 Post</b>	<b>Grade 5 Post</b>
<b>Treatment</b>	<b>P2: Made of different materials; or one is heavier material</b>	<b>25%</b>	<b>66% ***</b>	<b>80% **</b>	<b>81% *</b>
	<b>P3: Made of different materials; or smaller one is heavier material</b>	<b>20%</b>	<b>72% ***</b>	<b>78% *</b>	<b>80% **</b>
	<b>P4: Made of different materials; or one is heavier material</b>	<b>10%</b>	<b>25%</b>	<b>44% **</b>	<b>52% **</b>
	<b>P4: Made of different materials; or one is heavier material</b>	<b>0</b>	<b>16%</b>	<b>33% *</b>	<b>41%</b>
	<b>P4: Coordination of material &amp; size in explanation</b>				
<b>Control</b>	<b>P2: Made of different materials; or one is heavier material</b>		<b>30%</b>	<b>54%</b>	<b>60%</b>
	<b>P3: Made of different materials; or smaller one is heavier material</b>		<b>33%</b>	<b>50%</b>	<b>54%</b>
	<b>P4: Made of different materials; or one is heavier material</b>		<b>20%</b>	<b>19%</b>	<b>26%</b>
	<b>P4: Made of different materials; or one is heavier material</b>		<b>10%</b>	<b>11%</b>	<b>23%</b>
	<b>P4: Co-ordinate difference in material &amp; size in explanation</b>				

At the start of Grade 3, although about 80% of Grade 3 pre-Treatment students understood that bigger objects are often heavier, and appealed to the size differences of objects B and C in explaining why B was heavier, they generally did NOT note that the kind of material an object was made of affected its weight (in the cases where the objects were the same size but different weights, or the smaller was heavier). Some appealed to what might be “inside” the objects, whether one might be hollow or filled with more objects—these seemed to be “object level” rather than “material level” explanations. Only 20–25% appealed to the fact that the cylinders might be made of different materials in Problems 2 and 3, and none considered this explanation for Problem 4.

There was dramatic improvement for the treatment students by the end of Grade 3 in referring to differences in material or heaviness of material in their explanations for Problems 2, 3, but not yet in Problem 4, where improvement was more gradual and many students made the error of concluding that the different size objects weighed the same because they were the “same material.” Moreover, at the end of Grade 3, children in the Treatment group also tended to over-generalize and used “material” as a justification for

why a larger object was heavier than a small one, without considering that size matters. This is evidence that as children form initial links between weight and material in the Grade 3 Inquiry Curriculum, they use their existing “weight” concept rather than a distinct and differentiated notion of “heavy for size”. Through use of the nine density cubes<sup>7</sup>, the third-Grade Inquiry curricular unit provided these children with a wealth of opportunities to learn about different materials and to see the relevance of kind of material for weight differences.

However, the Grade 3 Inquiry unit did not stress the difference between weight and density (children worked with same size comparisons, in which weight and density are correlated) or fully develop a concept of volume, so we did not expect children would make much progress on differentiating weight and density and that was what we found. Children made more progress in differentiating weight and density after the 4<sup>th</sup> Grade curriculum (which focused on developing a more robust concept of volume and the idea of heaviness for size). Providing an explicit explanation of how the different size cylinders could weigh the same by coordinating the factors of type of material and size was still difficult even at the end of Grade 5. At that point, 41% of the students were able to explain that the smaller cylinder was made of a heavier kind of material than the larger one, but there was more of the lighter material, so the two cylinders ended up weighing the same.

Overall, the within-child improvement is highly significant for Treatment students, from pre-Treatment to the end of Grade 5 for all four measures (based on both Chi Square and Wilcoxon tests,  $p < .001$ ), and from post Grade 3 to post Grade 5 for three of the four measures.

In contrast, the Control students made no statistically significant progress in their explanations of problem 4 from Grade 3 to Grade 5 by either Chi Square or Wilcoxon measures, although they did make progress in focusing on difference in material in their explanations for Problems 2 and 3. This is consistent with the assumption that Treatment students, but not Controls, are making progress in developing a concept of heaviness of kind of material that is becoming differentiated from weight.

Direct comparisons with Treatment students at the end of Grade 4 also show Treatment students performed significantly better than Controls on all four comparisons, and on three of the four comparisons at the end of Grade 5 (with the fourth in the right direction, and approaching but not reaching significance).

## 2-“Heavier” vs. “Made from Heavier Material”

In Part 2 we probed for a precursor of density by asking children to make judgments about whether one object in a pair was made of a “heavier kind of material.

In an initial instructional phase uncovered cylinders of different sizes and materials are introduced. The student is told that the objects are made of Delrin (a particular kind of hard plastic), aluminum, and brass. The student is asked to compare the weights of the like-sized cylinders made of the three different materials and asked which object is made

---

<sup>7</sup> *Density cubes* refers to a set of nine 1-inch cubes of nine different materials, including, wood, plastic, and both light and heavy metals. They are used for exploring relationships among weight, volume, density, and kind of material.

of the heaviest material of these three (Figure 14). This gives children a chance to learn how brass, aluminum and Delrin differ in “heaviness”.

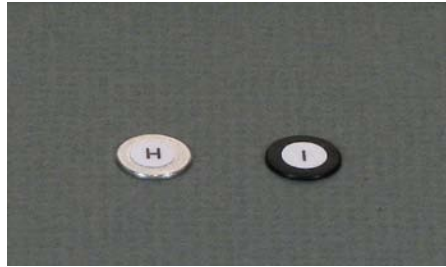
Following this, the student makes three critical comparisons that show whether she distinguishes an object’s density (more specifically, the heaviness of its material) from its weight. Unlike in the instructional phase, the child now must disregard the weight of the object in order to answer correctly, using information from the weight differences of equal size comparisons. The child is given a tiny discs of aluminum and Delrin, which both feel light and weightless (Figure 15); they are also told what materials each is made of, and are asked: “Is one made of a heavier kind of material or not? How do you know?”

They are also shown a small light sliver of brass and a large heavy piece of aluminum, and are asked which object is made of a heavier kind of material (Figure 16). Again, they need to ignore the fact that the large aluminum is heavy, and reason that the brass is a heavier material, and is simply light because it is so small.

In a final comparison, they are shown a small sliver of aluminum and a big piece, and are asked if one is made of a heavier kind of material (Figure 17). In this comparison, they need to reason they are both made of the same material, so although one is larger and heavier, neither is made of a heavier kind of material.



**Figure 14:** Objects used in “Heavier Material Task”.



**Figure 15: Critical comparison #1.** “Is one of these is made of the heavier material or not?” Note: H and I are small discs of aluminum and Delrin, respectively; they are too light to answer the question with confidence based on felt weight.



**Figure 16: Critical comparison #2.** “Is one of these is made of heavier material or not?” Note: The tiny disc, J, is brass; the large cylinder is aluminum.

Children’s judgments and explanations for these three critical items provides the second line of evidence that the Treatment children, but not the Controls, were developing a notion of heavier kind of material, which they distinguished from the felt weight of

objects came from analysis of their pattern of judgments in Part 2 of Task 3. Table 16 shows the percentage of students who ignored the misleading information of felt weight and were systematically correct across these three judgments: judging the aluminum disc to be made of a heavier kind of material than the Delrin sliver; the brass sliver to be made of a heavier kind of material than the large aluminum piece, and judging the two aluminum pieces to be made the same material. Correct judgments were accompanied by qualitatively different justifications. Children who made correct judgments focused on what they learned from the earlier (equal size comparisons) and what they knew about the different materials, while those who made incorrect judgments focused on making direct size and weight comparisons of the pairs.



**Figure 17: Critical comparison #3.** “Is one of these made of the heavier material or not?” Note: Both objects are aluminum.

**Table 16: Percent of Children Who Systematically Distinguish Heaviness of Kind of Material from Heaviness of Object (All Three Items)**

	<b>Grade 3 Pre</b>	<b>Grade 3 Post</b>	<b>Grade 4 Post</b>	<b>Grade 5 Post</b>
<b>Treatment</b>	<b>17%</b>	<b>33%</b>	<b>53%</b>	<b>63%**</b>
<b>Control</b>		<b>23%</b>	<b>35%</b>	<b>34%</b>

\*\* Pearson Chi Square,  $p < .001$ , Treatment Control Difference Grade 5

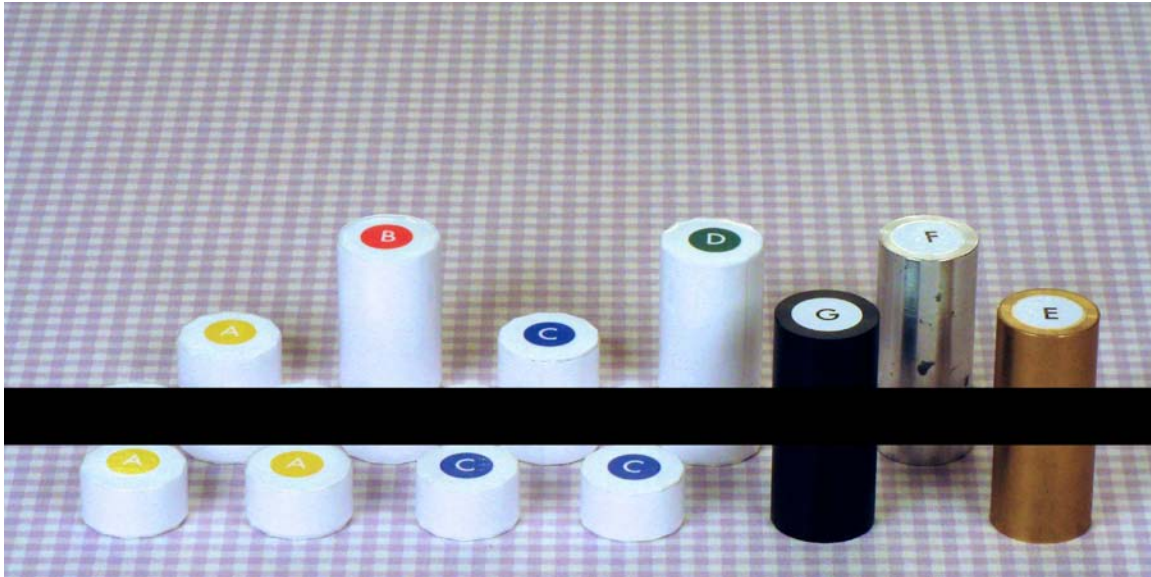
Treatment students showed significant improvement ( $p < .01$ ) from Grade 3 Pre to Grade 3 Post, and from Grade 3 Post to Grade 4 and Grade 5 post (as assessed by within child comparisons using the Wilcoxon matched pairs test for the longitudinal subsample and between child comparisons for the complete sample using Chi square). In contrast, there was no significant improvement from Grade 3 to Grade 5 for the Control students, either for the complete sample or for the longitudinal subsample. By the end of Grade 5, over 60% of the Treatment students were systematically correct, while the performance of the Controls remained at about 34%, a difference between Treatment and Control that was statistically significant (Chi square= 6.99, d.f.=1,  $p < .01$ ).

Significantly, we found some of the same patterns of relations across tasks in the two groups. These similarities are also of importance for LP work in suggesting important “inherent constraints” among concepts that any curriculum needs to be responsive to. For example, for both Treatment and Control students and for students at every grade level, systematically distinguishing heaviness of kind of material from heaviness of objects was strongly associated with understanding that tiny pieces take up space and have weight was strongly associated with systematically (Chi Square,  $p < .01$ ). The Inquiry curriculum was highly effective in enhancing both understandings, dramatically enhancing students’ understanding that tiny things take up space and have weight by the end of Grade 3, and more gradually and steadily enhancing their ability to distinguish the heaviness of materials and objects across Grades 3 to 5.

### 3-Infering materials from weight and size

In Part 3 the student tries to infer the possible materials of covered cylinders by comparing covered and uncovered cylinders (Figure 18). A balance scale and ruler are available for this. One solves the task by taking into account, in one way or another, the ratio of weight to volume (or volume to weight). Because this ratio is fixed for each material, one can eliminate materials that do not have the same weight-to-volume ratio as the target cylinder.

B is the same size and weight as F (aluminum). D is lighter than G (Delrin), and is in fact made of something else (a different kind of plastic). A is the same weight as F (aluminum) but one-third the size; if children think to stack the “extra copies” of A, they can determine that 3 A’s match E in size and weight. C is not equal in weight to E, F, and G; however, 3 C’s match F in height and weight.



**Figure 18:** Objects from the “Inferring Materials from Weight and Size Task”. **Note:** The three A cylinders are made from brass; the C cylinders, as well as B, are aluminum; D is made of a plastic having a density different from aluminum, and brass, and Delrin (G).

At issue are the features of the objects children attend to and what tools they use in making their inferences about materials: Do they focus simply on weight, or consider the relative size and weight of the objects? Do they not only lift the objects, but also make use the balance scale? (It is hard to distinguish the weights of G and D, but G is clearly heavier when put on the balance scale.)

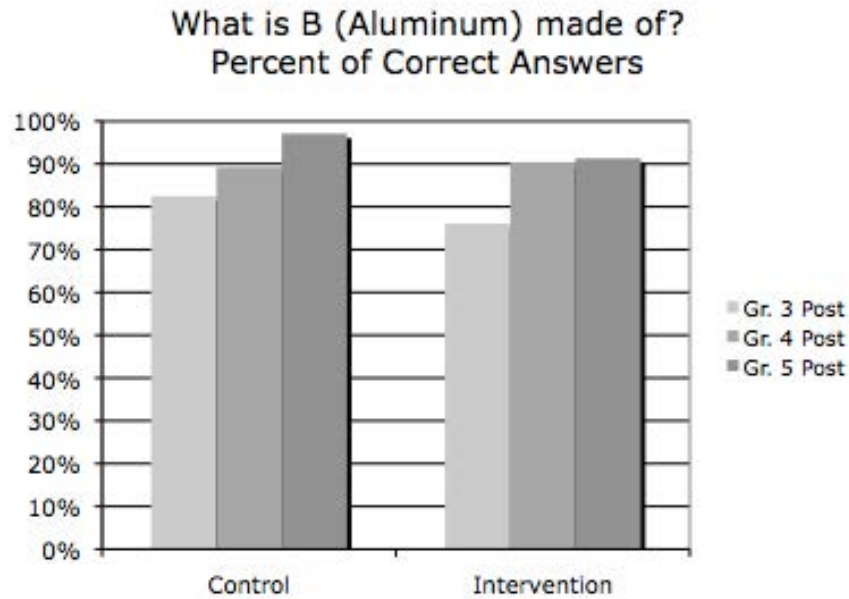
Of special interest is whether the student attempts to compensate for the differences in size by stacking small cylinders.

In inferring the material of covered cylinders, Grade 3 pre-treatment children generally reasoned based on direct comparisons of the weights of objects without taking size into account. They often simply hefted the individual objects without using the balance scale. Many children who found that the small covered (brass) cylinder and the large aluminum cylinder balanced on the scale concluded that they were made of the same material!

As they progressed through Grades 3 to 5, as we will see next, the Treatment students showed better results than the Control Group in sub-tasks that required compensating for the differences in size by stacking small cylinders.

### **Comparing Cylinders of Same Size**

*Sub-Task 1: What is B made of?* [Correct answer: aluminum] Children in all Grades and in both groups showed their best performance when they were asked to determine the material of B, a tall covered cylinder made of aluminum. Although there was an increase in correct answers from third to fifth Grade, no significant differences were found between groups or across Grades.



**Figure 19:** Cylinder B correctly judged to be made of aluminum

Here the percentage of correct answers ranges from 76 to 97%. Since B was the same size as the comparison cylinders, children only had to compare the weight of B with the weight of E, F, and G to find out that B could be made of the aluminum. Use of the scale, as opposed to holding cylinders in hands to compare their weights, did not differ across Grades or groups. Overall, when participants used the scale, 94% of their answers were correct; when they used only their hands they arrived at correct answers in 78% of cases.

*Sub-Task 2: What is D made of?* [Correct answer: some other material] Although D also was the same size as E, F and G, the children's performance regarding what cylinder D was made of was lower than in the case of cylinder B, for all Grades and for both groups, with correct answers ranging from 38% to 65%. There was a noticeable increase in the percentage of correct answers from third to fourth Grade in both groups, and a slight decrease from fourth to fifth Grade.

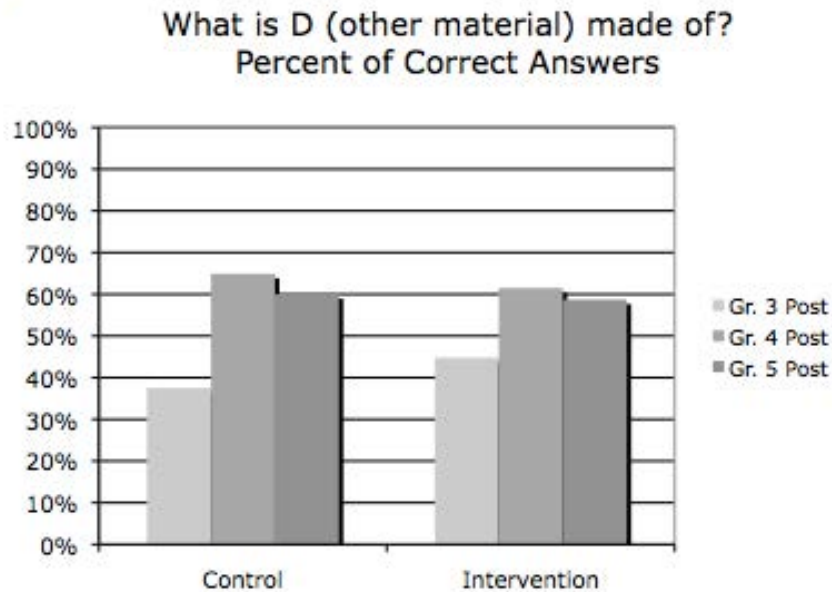


Figure 20: Cylinder D correctly judged to be made of other material

In this case, the students had to compare the target tall cylinder D with E, F, and G and realize that no comparison cylinder weighed the same as D. The increase in correct answers after third Grade was related to use of the scale, which happened more often in fourth and fifth Grades, for both groups. When the scale was used, the students arrived at the correct answer in 86% of the cases. When only hands were used the percentage of correct answers dropped to only 14%, with 78% of the students wrongly stating that cylinder D was made of Delrin.

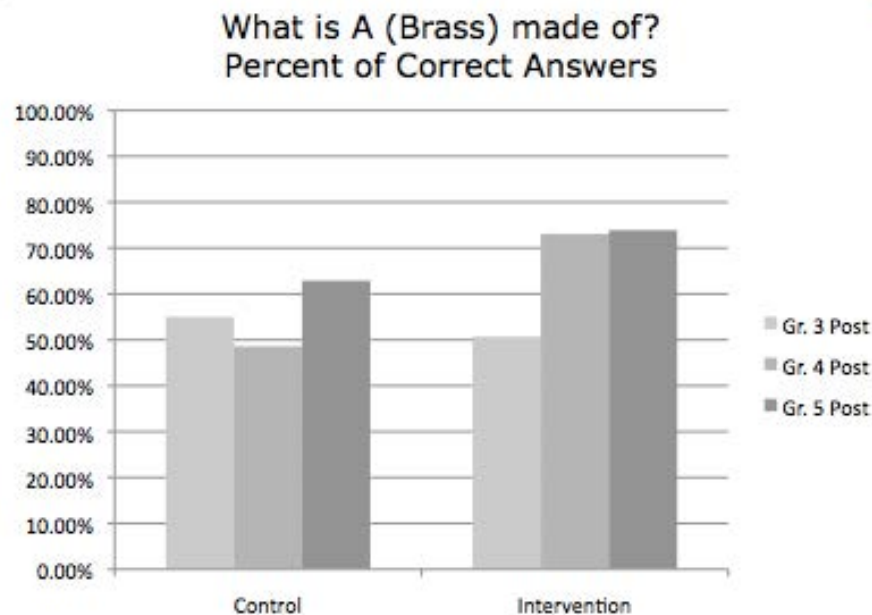
### ***Comparing Cylinders of Different Sizes***

For the last two tasks the target cylinders A and C are one-third the size of the comparison cylinders E, F, and G. As we will see next, this led to a difference in correct answers in finding both what A (brass) and C (aluminum) were made of, compared to answers for what B (aluminum) was made of in Problem 1 (where direct same size comparisons were available), although the drop for A (brass) was less than for C. More importantly to our analysis of the effect of the classroom intervention, in comparison to the Control Group, the Treatment group showed a higher percentage of correct answers in these two tasks and more progress from Grades 3 to 5.

In both of these tasks, precise correct answers would require determining the relationship between the heights of the two compared cylinders. The availability of copies of the small cylinders allowed children to stack three small target cylinders, thus obtaining a composed cylinder that could then be compared on the two-pan balance scale to the tall cylinders E, F, and G. As we will see, however, participants also reached correct answers using other means, mainly in determining what A (brass) was made of.

*Sub-Task 3: What is A made of?* [Correct answer: brass] Here, the target short cylinder A, made of brass, was heavier than cylinder G, had the same weight as F, and was lighter than E. The correct answer for this task was that cylinder A was made of the same kind of material as cylinder E.

The Intervention students made clear progress from Grade 3 to Grade 5 in judging that A was made of brass (Figure 21), with 51% of correct answers at Grade 3 and 74% at Grade 5. A Fisher exact probability test for the frequency of correct versus incorrect answers by Grades three and five showed a significant association between type of response and Grade level for the Intervention group ( $p = .01$ ). In contrast, no significant association between the same variables was found for the Control Group.  $p = .38$ ). Further, in comparison to the Control Group (Figure 22), in Grades 4 and 5 the Intervention group showed a higher percentage of correct answers (on average, 73%) than the Control Group (on average, 56%). Chi-square results showed no significant association between group and type of response in Grade 3, with Treatment students if anything a little worse, a significant association in Grade 4 ( $p < .05$ ), and no significant association in Grade 5, probably due to the fact that the progress made by Treatment students in Grade 4 became stable by Grade 5, while students in the Control Group showed a delayed progress in Grade 5.

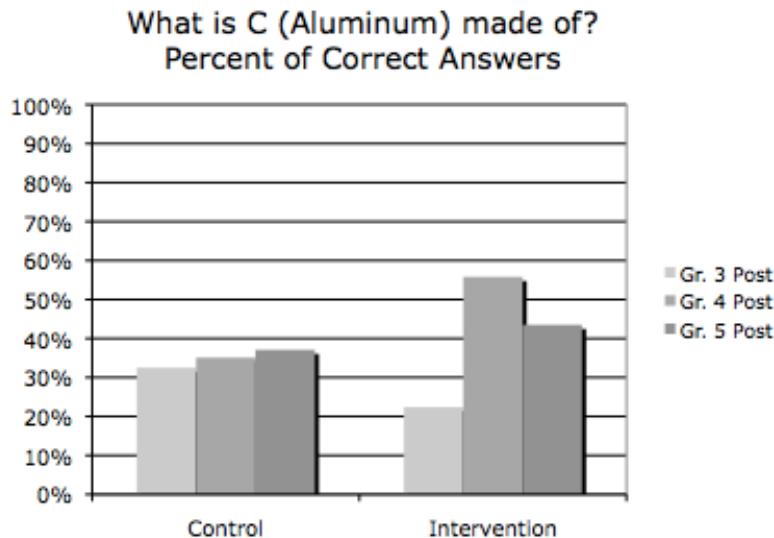


**Figure 21:** Cylinder A correctly judged to be made of brass

The increase in the treatment group's performance can be attributed to the fact that more children began to stack three copies of the small target cylinder and then compared the stack's weight to the weight of the tall cylinders. Of the children who used stacking 95% concluded correctly that cylinder A was made of brass. However, 44% of the children who did not use stacking also gave a correct answer. These correct answers were usually accompanied by comparison of the short and the tall brass cylinders by holding one in each hand and by statements that the small and the tall cylinders felt the same. Such

statements suggest intuition of some property related to the density of the material, or to its “heaviness.”

*Sub-Task 4: What is C made of?* [Correct answer: aluminum] Although this task is similar to the previous one, with a target cylinder that is one-third the size of the comparison cylinders, children’s performance is much lower, with percentages of correct answers ranging from 33 to 37 for the Control Group and from 22 to 56 for the Treatment group (see Figure 22).



**Figure 22:** Cylinder C correctly judged to be made of aluminum

Similar to the results of the previous task, the increase in performance from third to fifth Grade was more pronounced in the Treatment group. A Fisher exact probability test for the frequency of correct versus incorrect answers by Grades three and five showed a significant association between type of response and Grade level for the Treatment group ( $p < .05$ ). No significant association between the same variables was found for the Control Group.

Again, about one third of the participants stacked the cylinders to find the answer. Of the children who used stacking 80% concluded correctly that cylinder C was made of aluminum. Among those who did not use stacking, only 14% gave the correct answer.

The difference in children’s performance here, in comparison to the previous task, may be due to the fact that the compared cylinders did not give a clear feeling of “heaviness” or some intuition of density, as was the case for the brass cylinders. Here, correct answers required the coordination of size and weight and the relative quantification of size by stacking cylinders.

#### 4-Summary of Results for Weight, Size, and Heaviness of Materials

Students in the Treatment group made notable progress on three different tasks assessing their differentiation of the weight of objects and the heaviness of materials and realizing objects made of some materials are heavier for their size than others. They

systematically distinguished the questions of whether an object was heavier from whether it was made of a heavier kind of material. They appealed to differences in heaviness of materials to explain why smaller objects can be heavier or the same weight as larger objects. Finally, they recognized the need to compare weights of same-size cylinders, a novel task for them, in order to make inferences about what kind of material something was made from. In contrast, the Control students made little progress on these questions, so that by the time of Grades 4 and 5, the differences between Treatment and Control students were significant across most of these tasks.

---

#### D. Granularity of Weight, Length, and Number

Earlier we looked at the issue of whether young children *acknowledged infinitesimal objects*—objects too small to be perceived yet understood as having weight and taking up space.

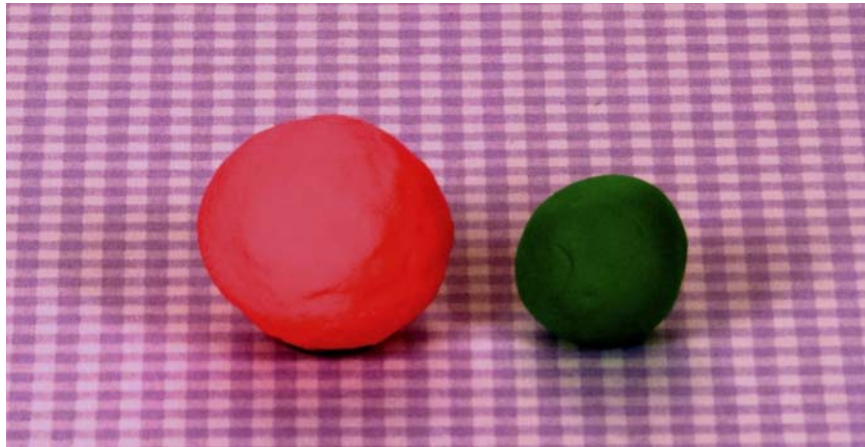
We now turn to the matter of *infinitesimal differences in value*: **how many distinct values of some property are assumed to lie between the values of two examples?** The notion of infinitesimal differences permeates the number line as a representational system for real numbers. Every real number occupies a unique position on the real line and all positions on the line are assumed to be occupied by a real number. This metaphor carries over into science, where quantities such as weight, speed, and position are dimensionalized, that is, conceived as orderable on quantity lines (number lines with units of measure). These ideas are illustrated by the granularity tasks used in the interviews.

**Granularity of number** was assessed by asking students to express how many numbers lie in the intervals (4, 7) and (4, 5)—see Figure 23. When asked about the interval from 4 to 7, students frequently assumed the interviewer was referring to counting numbers (for example, they would state there were two numbers, namely, 5 and 6). To move beyond the counting numbers, the interviewer then focused on the interval between 4 and 5 (“Are there any numbers between 4 and 5?”). If the student denied that there were numbers between 4 and 5, the interviewer asked, “Is four and one-half a number?” and encouraged the child to consider fractions.

- \* How many numbers are there between 4 and 7?
- \* How many numbers are there between 4 and 5?
- \* Is “four and one-half” a number?

**Figure 23: Granularity of number--How many numbers are there between 4 and 7? (4 and 5?)**

**Granularity of weight** was assessed by asking the student to speculate about the number of weights that might lie between the weights of two given objects. The student was shown and allowed to handle the two balls of clay shown in Figure 24). The interviewer then asked, “How many weights do you think there could be between that of the red ball and the green ball?” If the student appeared to find the question puzzling—students occasionally interpreted the question as referring to the number of balls that could literally fit *between* the two balls—she explained that a ball with a *weight between* referred to a ball that was *heavier than the green ball and lighter than the red ball*.



**Figure 24: Stimuli for the granularity of weight task: “How many weights do you think could be between that of the red ball and the green ball?”**

**Granularity of length** was assessed by asking the student to consider the number of lengths that might fall between two lines drawn on a sheet of paper (see Figure 25). Line A is approximately 7 inches long; Line B is roughly 5 inches long, but no measures are given and no notches or partitions are displayed. The key question was: “How many lengths could be between that of Line A and Line B?” If the student expressed doubt or hesitation about the meaning of between, the interviewer explained that a line with a *length between* was shorter than Line A but longer than Line B.

Line A

---

Line B

---

**Figure 25: Granularity of length stimuli. "How many lengths could be between that of Line A and that of Line B?"**

We did not expect most students to have explicit theories about infinitesimals. However, they could be expected to have intuitions about how many different weights, lengths, and numbers might exist between two given examples.

Table 17 gives an overview of the granularity results for (all Grades and groups). Two judges coded students' responses into four categories employed for granularity of number, weight, and length.

- *Infinite number*: Infinite, "as many as you want", "thousands and thousands", "hundreds and hundreds".
- *Lots*: 13-100.
- *Some*: one to 12.
- *None*

We used an admittedly liberal criterion for "infinite" because most young students could not be expected to know the word. If they expressed a very, very large number, we considered the response as *infinite*. *Some* was taken to include all values between 1 and 12. (Very few children mentioned amounts between 13 and 50.)

### 1-Overall Granularity Results

*Number Before* (Fig. 17, col 1) refers to students' answers before the interviewer mentioned the example of four and one half. In other words, *Number Before* represents the spontaneous, unprompted answer of the student; 56% of students initially said there were no numbers between 4 and 5. After the prompt about four and one-half (*Number After*) only 12.6% of the children continued to insist that there were no numbers between 4 and 5. But, remarkably, the great majority of responses (61.4%) reflected the view that

there were only some (one to 12) numbers between 4 and 5. Whereas we (the researchers and most adults) take for granted that the numbers on a number line are densely packed (perhaps even continuous insofar as no positions are left unoccupied by numbers), the students tended to think of the interval between two consecutive integers as inhabited by very few numbers.

Children also tended to view weight and length as very coarsely graduated dimensions. 78% believed there were fewer than 13 distinct weights that could lie between the samples. In keeping with their initial reliance on felt weight (which has little granularity), the vast majority of children (approximately 4 out of 5 children) assert there are only very few weights in between the shown examples (generally only 1, 2, or 3). Some children (about 13%) begin to suspect there might be more than a dozen, but still are imagining a limited number (15, 20, 50, etc.). But only exceptionally do young students (approximately 9% of the students) suggest that the number of possible distinct weights in the interval is enormous (“lots and lots,” “infinite,” “goes on and on”).

73.9% of the students thought there would exist fewer than 13 distinct lengths between the displayed lines. Clearly, these students do not think of length as a continuum or even a highly packed dimension. Rather, there are relatively few lengths in a given interval. Somewhat to our surprise, there was no indication that length was more finely grained than weight. We were surprised because much younger children are known to successfully seriate sticks much more finely graduated. Apparently it is one thing to perceive differences in a set of stimuli and another thing to mentally represent fine gradations in length in the absence of concrete stimuli.

It should be recalled that the differences between the sample stimuli were substantial: one line was 40% longer than the other, and one ball had four times the weight of the other!

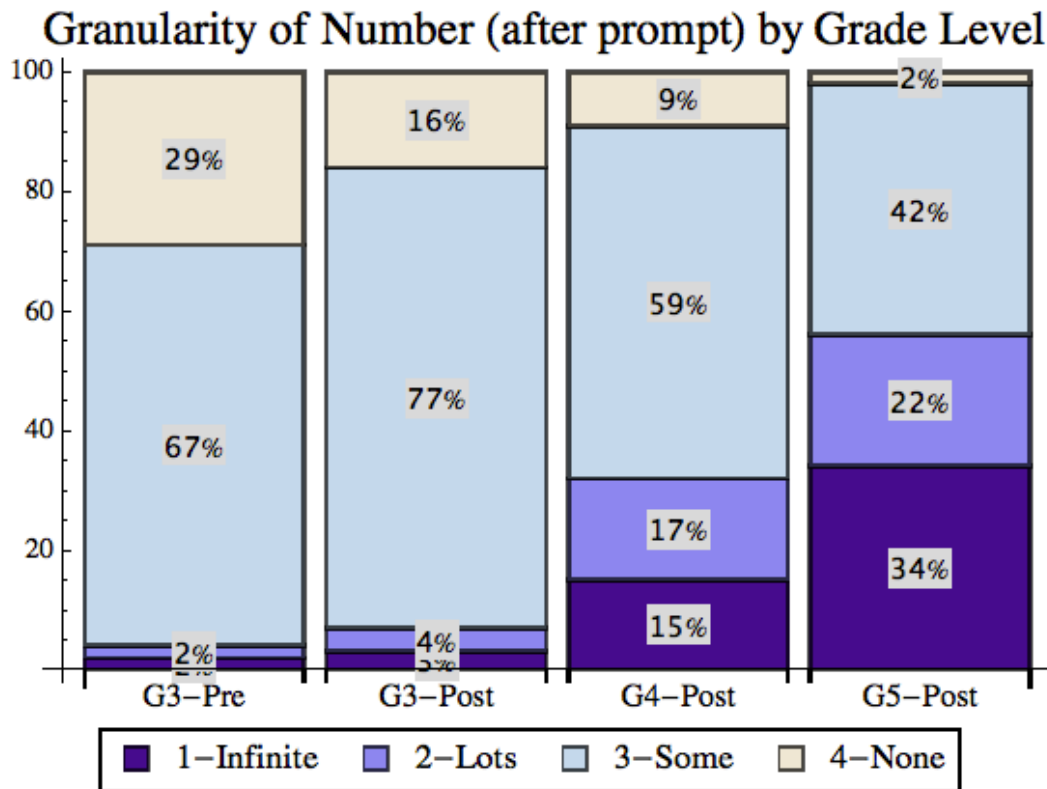
**Table 17: Granularity of Number, Weight and Length (all Grades and groups)**

	Number Before	Number After	Weight	Length
“Infinite”	12.9%	14.2%	9.0%	6.1%
Lots	10.0%	11.8%	13.0%	20.0%
Some (1-12)	20.8%	61.4%	69.9%	67.2%
None	56.3%	12.6%	8.1%	6.7%

## 2- Granularity By Grade Level

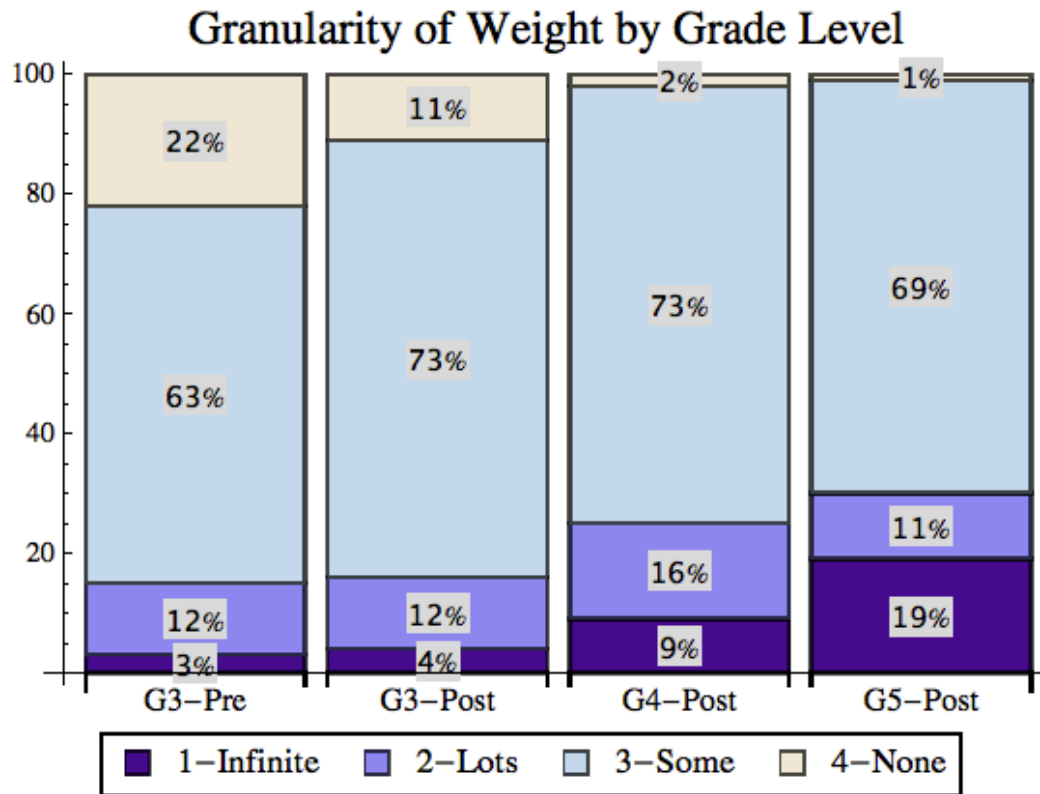
Students made substantial progress in Number Granularity throughout the investigation (see Figure 26). Whereas at the beginning of the study almost 1/3 of the students thought there were no numbers at all between 4 and 5—they either denied that 4 and  $\frac{1}{2}$  was a number or said that it did not lie between 4 and 5—by the end of the study, fewer than one student in twenty did so. Also by the end of Grade 5, more than half of the students believed that there were many, many numbers (Lots or Infinite) numbers between 4 and

5, as opposed to fewer than 5% at the beginning of the study. However there were no differences between the Treatment and Control Groups on this measure.



**Figure 26: Granularity of Number by Grade Level**

Students also made substantial progress in Granularity of Weight throughout the investigation (see Figure 27) although, once again, there were no differences between the Treatment and Control Groups. At the end of Grade 5, only 30% of the students believed there were very many (Lots or Infinite) possible weights between the two balls of clay they were shown.



**Figure 27: Granularity of Weight by Grade Level**

The reader will recall that the Treatment students had made substantially greater progress than the Control Students in acknowledging that tiny, invisible objects exist, have weight and take up space. Table 18 shows there was a relation between the Granularity of Weight responses and the Attribution of Weight to a Tiny Speck (all Grade levels and conditions) ( $\chi^2(1, N=345)=30.67, p<.0001$ ). Those students who acknowledged that a tiny speck had weight were more likely than the remaining students to have a finely graded sense of the dimension of weight (Infinite or Lots). However, the variables were sufficiently distinct so that the fact that one believes that tiny objects have weight does not necessarily imply that one conceives of weight as a continuum with very tiny differences among possible values. In other words, we conclude that the *acknowledgement of infinitesimally small objects* (objects too small to be perceived yet understood as having weight and taking up space) is not tantamount to acknowledging *infinitesimal differences in weight*. One imagines that these ideas each need to be discussed in the curriculum

**Table 18: Association between Attribution of weight to a tiny, invisible speck and Granularity of Weight.**

Group	(Show All)			
Count of Invisible Speck	Invisible Sp			
weight	1=yes	2=no	(blank)	Grand Total
1-Infinite	21	10		31
2-Lots	26	19		45
3-Some	71	170		241
4-None	3	25		28
(blank)				
Grand Total	121	224		345

Progress on Granularity of Lengths was generally similar to that of Granularity of Weights (Figure 28). There was clear improvement from Grades 3 to 5. However, the improvements could not be attributed to the Inquiry Treatment.

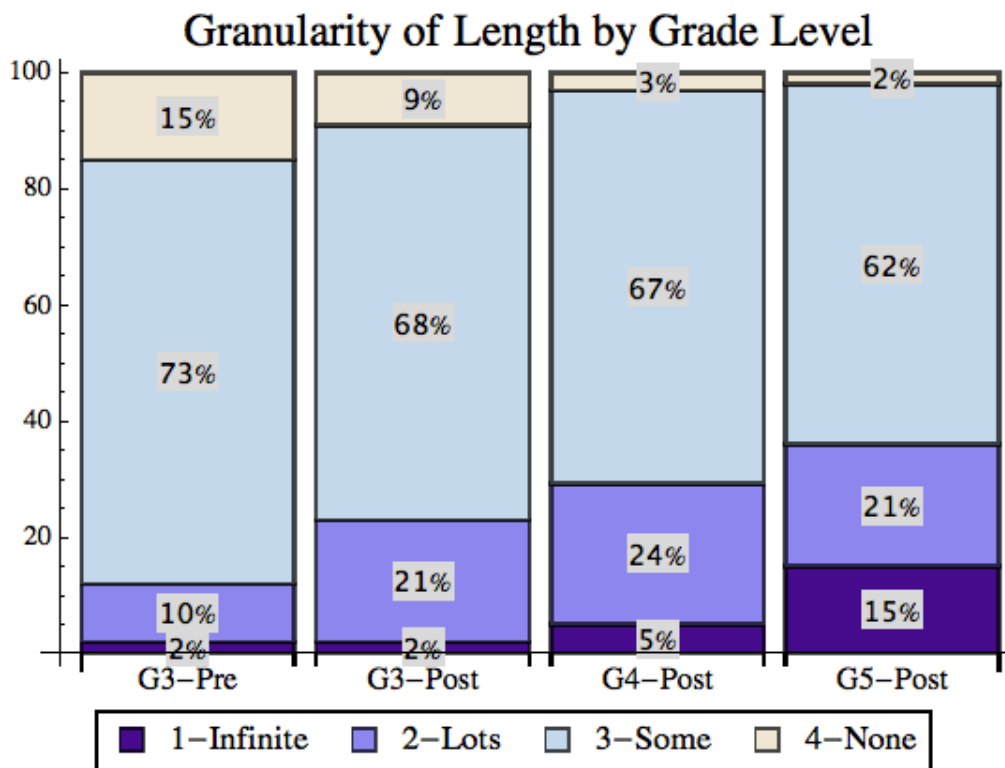
**Figure 28: Granularity of Length by Grade Level**

Table 19 gives a more precise look at the response category *Some*. In the case of number granularity, *some* corresponded to an average of just under 3 (2.8). For weight and length granularity, *some* corresponded to just over four and one-half weights and lengths, respectively. So, although the category included all responses one or greater but less than

13, they typically involved much smaller numbers. And in the case of number, the most common response (the mode) was 1. In other words, those students classified as acknowledging some numbers between 4 and 5 more often than not only recognized four and one-half as being between those values. So although the granularity of number is initially more constrained than the granularity of weight and length, by the end of grade 5 it has moved ahead of them. This corresponds to the transition from natural numbers to rational numbers that occurs in elementary and middle school (with the introduction of fractions and rational numbers on the real line). We suspect that this transition has a gradual influence on the conceptualization of physical quantities as finely graduated dimensions.

**Table 19: Granularity of number, weight, and length. Representative values of "some". (All Grades and groups)**

	Mean	Median	Mode
"Some" numbers	2.8	2	1
"Some" weights	4.6	4	3
"Some" lengths	4.6	4	3

Some additional thoughts on the relations between the number line and the conceptualization of physical quantities as dimensions are included at the end of the next section.

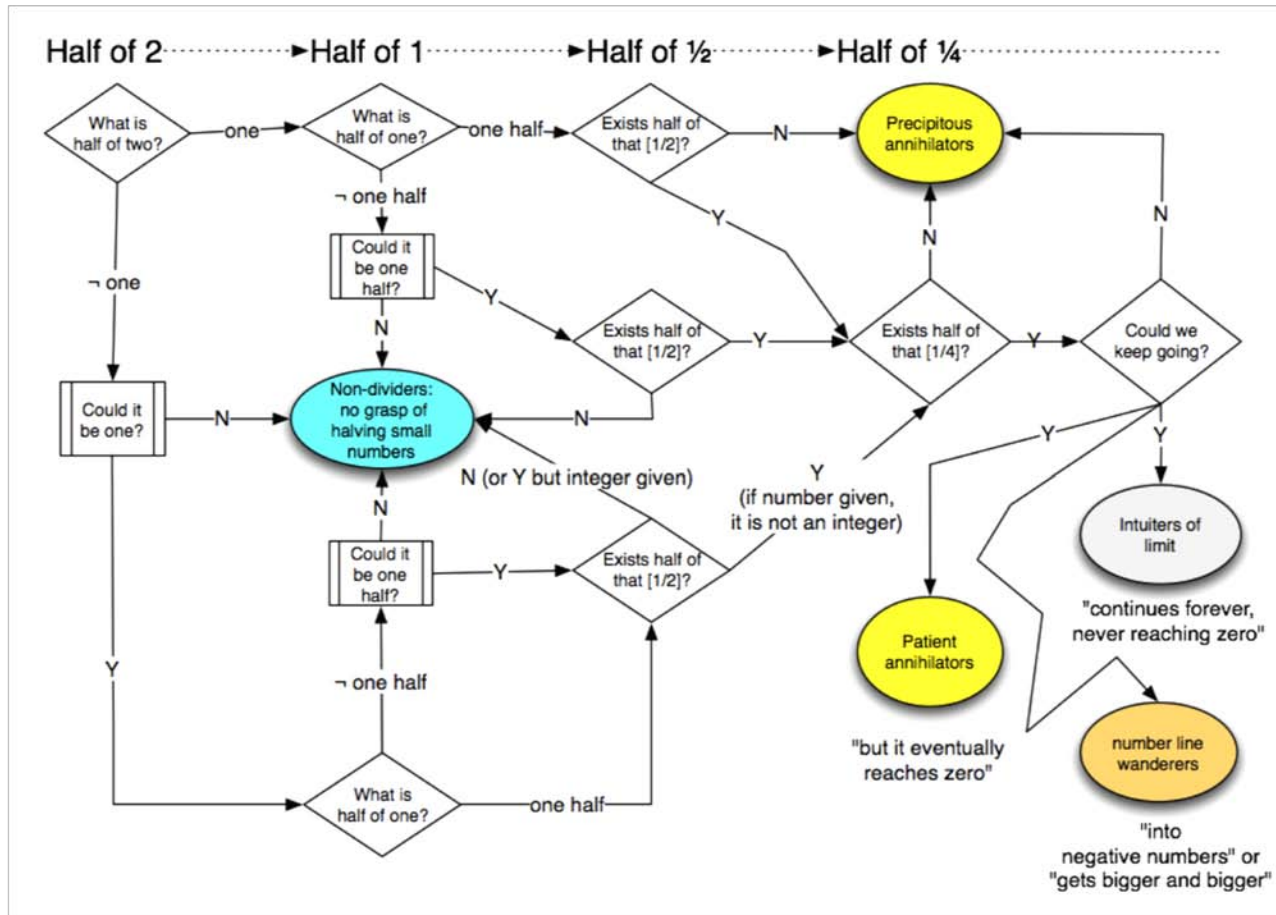
### 3. The Repeated Halving of Numbers

#### INTRODUCTION

What do students think happens when numbers are repeatedly halved, that is, divided by two? Does the process ever end? If so, when and why?

The Repeated Halving Task consists of a clinical interview (see Figure 29) around the following issues:

- What is half of two? (The student was asked to name the result correctly as one. If she hesitated, the interviewer suggested that one is half of two; if the student accepted the suggestion, the interview proceeded).
- What is half of one? (Correct answer of one half required or suggested by interviewer)
- Is there a number that is half of that? (It was not necessary that the student know that the name of the results from this point forward, although names given by students were recorded.)
- Is there a number that is half of that result (that is, half of  $\frac{1}{4}$ )?
- Could we keep going like this? Why or why not?



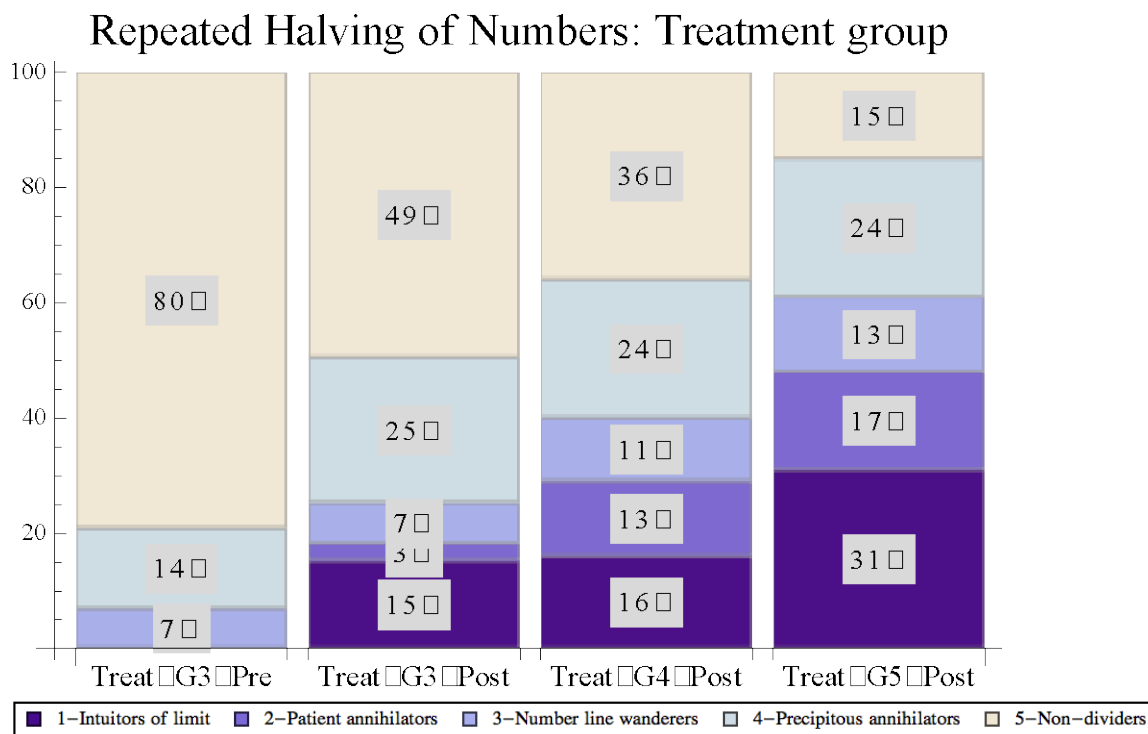
**Figure 29: Repeated Halving Questioning and Relation to Response Categories**

Two judges read the interviewer notes, including partial transcriptions of responses, and classified students into five groups:

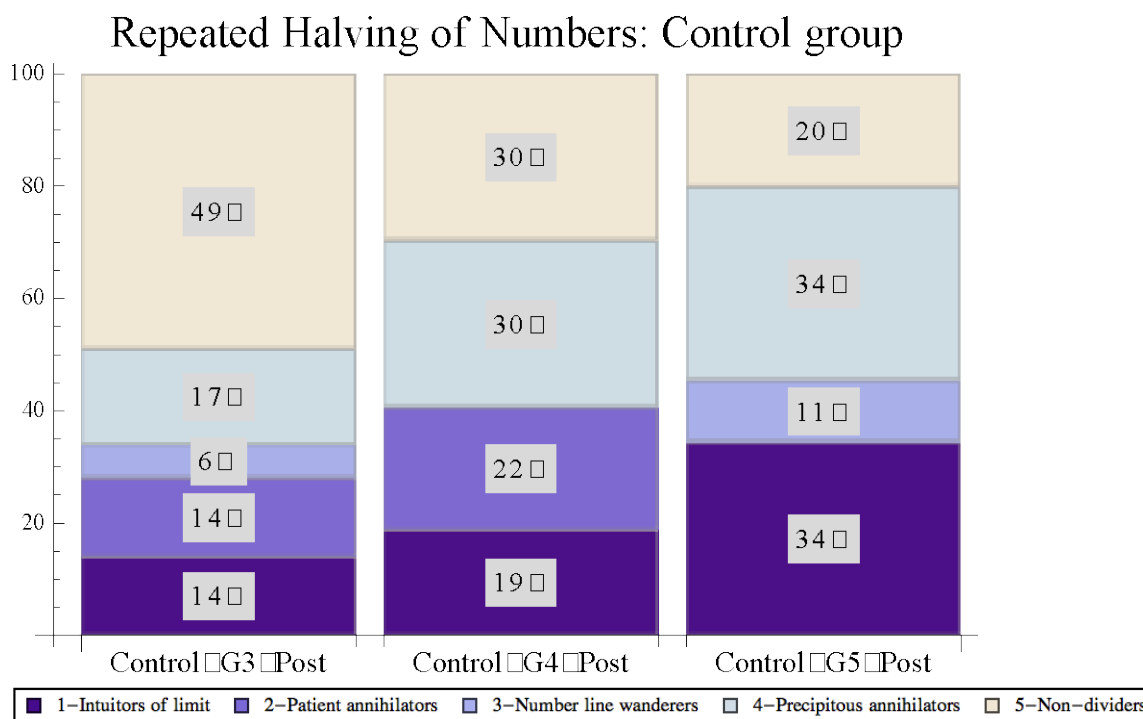
- **Non-dividers** did not display a grasp the elementary notion of taking half of a number. They failed to acknowledge (even after the possible suggestion by the interviewer) that half of two was one, that half of one was one-half, or that one-half could be halved. Those who initially gave incorrect answers but accepted the correct answers when encouraged to do so will belong to one of the remaining categories.
- **Number line wanderers** claimed the results reached zero but continued beyond into the negative numbers, never terminating. They started out correctly but failed to recognize that zero would be a limit. Some may have confused dividing by two with subtracting  $\frac{1}{2}$ .
- **Precipitous annihilators** successfully handled the cases of half of two, half of that, but thought the process ended or produced nothing at one of the next two steps.

- **Patient annihilators** acknowledged that halving continued for more steps but stopped or reached zero at some point.
- **Intuitors of limit** expressed the view that halving went on and on forever, never reaching zero.

At the beginning of the study (Grade 3) the majority of students had difficulty in understanding the division of numbers for the simplest three cases. Some even asserted that one did not have a half. Interestingly, there was evidence that Treatment students were making more progress than Control students from Grade 3 to the end of Grade 5. More specifically, among the Treatment students, the number of students in the two most sophisticated categories (Intuitors of Limit or Patient Annihilator) more than doubled from the end of grade 3 (18%) to the end of grade 5 (48%), a highly significant improvement ( $\chi^2(1, N=121) = 11.33, p < .001$ ). In contrast, there was no significant change among the control. They started off more sophisticated with 28% in the Intuitors of Limit or Patient Annihilator category; by the end of grade 5 only 34% were in one of these categories ( $\chi^2(1, N=73) = .11, p = .74$ ). There was also evidence that understanding repeated division of number was strongly related to granularity of number for both treatment and control students.



**Figure 30:** Repeated Halving of Number (Treatment Group)



**Figure 31:** Repeated Halving Of Number (Control Group)

#### 4- Summary of results on granularity and repeated halving of numbers

The results regarding repeated halving of numbers and granularity surprised us considerably. When we began the study, we thought, perhaps naively so, that students would regard measurable properties such as length and weight as continua and, if not exactly continua, tightly packed dimensions. We knew from earlier research by Smith and her colleagues that students would hesitate in attributing weight and volume to objects that were so very tiny. This also seemed consistent with the view that their judgments about physical magnitudes were heavily reliant on their sensations: if an object felt as if it had no weight, there was no apparent reason to infer it had weight. (As we saw, this belief is rather easily changed through instruction.)

However, the notion that there are very few distinct weights between any given examples was very surprising, especially given the fact that students had a great amount of practice, in the treatment group, associating weights and volumes with positions on the number line (or quantity lines).

For some reason this experience with number lines does not necessarily translate into a reconceptualization of quantities as finely graduated. In fact, length itself is not treated as having many distinct instances. Does this suggest that students are simply casual about what they consider distinct values? Do they only acknowledge differences that are sufficiently large as to be obvious?

Why would this occur for numbers? If a student has any inkling that a great many distinct fractions exist, even if they are very close in value, we would imagine that they would have mentioned them.

There seems to be much to be learned about the relationship between number line understanding and the conceptualization of quantities as dimensions. We do not have time and space to go into this here, but a few remarks would seem to be merited.

A number line is a visual depiction of how numbers are ordered; there is a one to one correspondence assumed between each real number and its respective location on the idealized line.<sup>8</sup> The model of the real numbers is only partially revealed through number line diagrams. In actuality, the real number line model includes *operations* on real numbers, most prominently, addition, subtraction, multiplication, and division. Arithmetical operations and fundamental properties of real numbers, known as the Field Axioms, can be represented dynamically in the model as actions on line segments corresponding to intervals.<sup>9</sup>

When a teacher first introduces a number line in an elementary mathematics classroom, she and her students will be having a discussion about rather different things. Her young students will almost invariably believe that the number line contains only the numbers they know: the counting numbers<sup>10</sup>, 1, 2, 3.... They believe that no numbers exist between consecutive counting numbers. From our point of view, they believe the number line to be “sparsely populated.” Furthermore, young students will fail to understand how subtraction and division relate to operations on the number line. And, until and unless they receive instruction on the matter, they will not understand how a fractional operator (e.g.  $\frac{2}{3}$ ) relates to multiplication and division by integers. Their expanded knowledge of number systems will go hand in hand with adjustments in their understanding of the number line. The number line also provides a foundation for two-dimensional and higher-order coordinate spaces of use for visually representing functional dependency.

A physical quantity can be thought of as a *dimension* along which the amount or intensity of an attribute (weight, volume, brightness, distance...) can be spatially ordered. The dimension has a metric if there is an agreed upon way to assign values to locations and distances (intervals) on the continuum.

Because young students readily make comparative judgments of quantities and numbers (one object is heavier, wider, warmer...than the other; 6 is bigger than 2), we know they

---

<sup>8</sup> In a drawn number line, points necessarily have size; in the idealized model of the number line, mathematical points do not have size. There are other differences between a drawn number line—which is a signifier, a symbolic token—and the ideal number line model—the signified or referent. It is even prudent to think of the number line as a symbolic representational system rather than an isolated symbol.

<sup>9</sup> The additive axioms (associativity, commutativity, identity, and inverse) are straightforward; the multiplicative ones are not. E.g. a multiplicative inverse requires that there be a metric.

<sup>10</sup> The counting numbers are also known as the natural numbers or whole numbers (although some authors consider zero to be a whole number).

treat them as (ordered) magnitudes. But how fine are the smallest distinctions they acknowledge? How many numbers are there in a given interval? How finely grained are their notions of length and weight? At issue are students' presumptions about the granularity of quantities and numbers. The present results suggest that children have much more limited granularity than we do; even by Grade 5, only about half acknowledge there are many numbers (infinite, lots) between 4 and 5 or that one can keep dividing numbers in half, and fewer than half acknowledge there are many weights or lengths between two noticeably different weights or lengths.

Such questions happen to bear directly on the two issues mentioned earlier; namely, the shift away from an over-reliance on explanations based on perceptual judgment and the view of measurable quantities as dimensions having properties akin to those of real numbers. In grade 4, Inquiry students observed the effect of adding a small object to a glass of water: the water level rose appreciably. However, when asked in class discussion to consider whether a rock cast into a lake would cause the water level to rise, they were greatly divided. Only some of the students would acknowledge *imperceptible changes in magnitude*. We suspect their willingness to acknowledge these changes relates to their growing understanding of the granularity of number.

---

#### **E. Relating scientific concepts and quantification of components of a mixture (sugar and water)**

The Inquiry Curriculum focused on developing children's quantitative understanding of specific quantities--weight and volume--and their coordination in an idea of heavy for size as part of developing a deeper understanding of matter and materials. This is a rich context for developing an understanding of measure and proportional reasoning and the previous section demonstrated some of the progress Treatment children made in that context. But how would children think and reason about other (related) contexts that call for proportional reasoning not considered explicitly in the Inquiry Curriculum, such as reasoning about concentrations of different mixtures? Here our findings indicate that such generalization is not automatic, raising important questions about the complex interactions between the understanding of specific content and mathematical structure.

On tasks involving concentrations of liquid mixtures, children are considered to reason proportionally insofar as their judgments of an intensive quantity (e.g. sweetness) are in accordance with the given ratios of amounts. When the *ratios* of solute to solvent are equal (e.g. one mixture of 2 cubes sugar in 4 units water is compared to another of 3 cubes sugar in 6 units of water) the child predicts the mixtures will be equally sweet; when the ratio for one mixture is smaller (even though its *raw amounts* may be greater), that mixture is inferred to be less sweet.

Researchers (e.g. Piaget, Karplus, Noelting) typically interpret the task in terms of whether the child recognizes the mathematical structure inherent in the problem (e.g.  $a:b < c:d$ , where  $a, b$  might correspond to the weights of two cylinders and  $c, d$  to their volumes) or incorrectly attends to other mathematical properties or relations (e.g.  $b-a >$

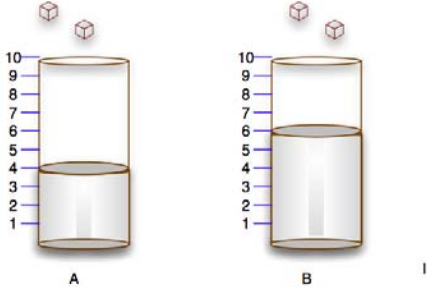
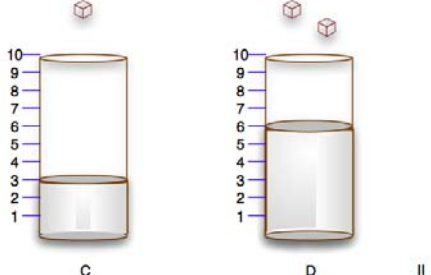
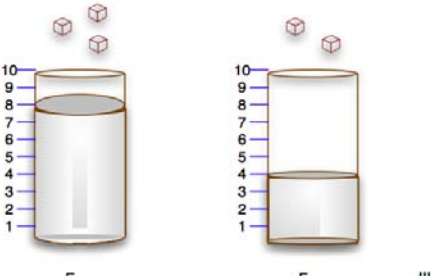
d-c; or  $b > d$ ). The choice among competing mathematical models is to be resolved by a student's care in identifying the structure of the problem.

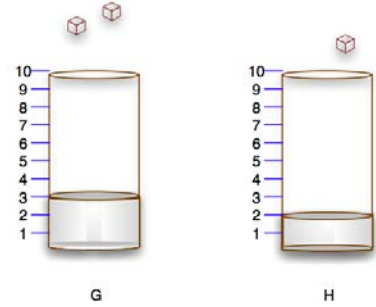
As an empirical problem, the task is first and foremost a matter of figuring out how ingredients contribute to the taste of the final mixture. It only becomes a matter of comparing ratios if and when the scientific reasoning evokes a model of ratio and proportion. So one could argue that poor performance is due to a lack of understanding about how mixtures work (the concentration of solute determines the intensity of the taste).

Here we examine both the emergence of proportional reasoning and children's ideas about how sugar and water combine to produce solutions of differing sweetness. We wonder whether it is possible to distinguish between those cases where unsuccessful reasoning is due to the lack of proportional reasoning or to misunderstandings about what determines the sweetness of a mixture.

Participants were shown the four diagrams in Table 20 and asked if, after mixing, the drink in one pitcher would taste the same as in the other or whether one would be sweeter.

**Table 20:** Mixture Tasks, showing questions and also diagrams shown to students.

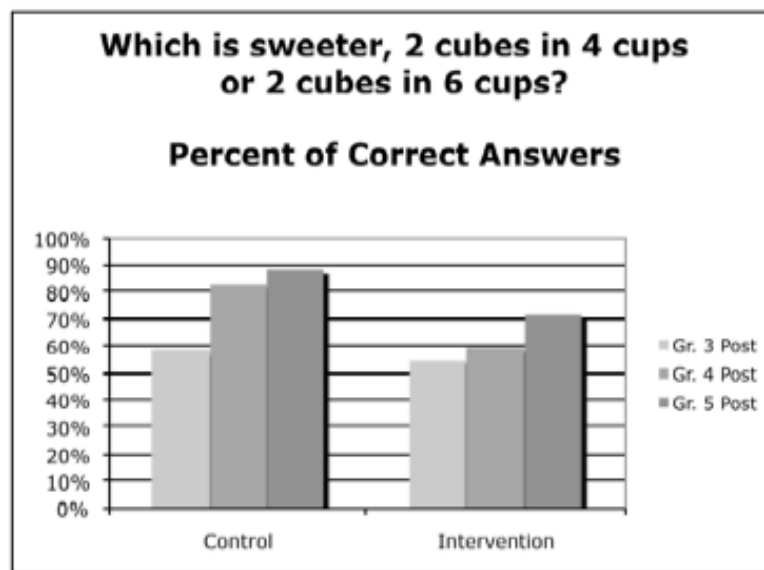
The Questions	Cards shown to students
<p>Task 1:</p> <p>Which mixture will taste sweeter?</p> <p>2 sugar cubes in 4 cups of water or</p> <p>2 sugar cubes in 6 cups of water?</p>	 <p style="text-align: center;">A                      B                      I</p>
<p>Task 2:</p> <p>Which mixture will taste sweeter?</p> <p>1 sugar cube in 3 cups of water or</p> <p>2 sugar cubes in 6 cups of water?</p>	 <p style="text-align: center;">C                      D                      II</p>
<p>Task 3:</p> <p>Which mixture will taste sweeter?</p> <p>3 sugar cubes in 8 cups of water or</p> <p>2 sugar cubes in 4 cups of water?</p> <p>How could you make the two mixtures equally sweet?</p>	 <p style="text-align: center;">E                      F                      III</p>

<p><b>Task 3:</b></p> <p>Which mixture will taste sweeter?          2 sugar cubes in 3 cups of water or          1 sugar cube in 2 cups of water?</p>	
---	--

### 1-Same Amount of Sugar, Different Amounts of Water

Task 1 was the easiest of the four since the answer depended solely on the amount of water, given that the amounts of sugar were equal. Children's difficulty in this task may not stem from lack of proportional reasoning, since Task 1 can be solved by focusing on a single variable, without taking into account the specific number of units: the beakers have the same amount of sugar but B has more water, so it must contain a less sweet solution.

Here, more than half of the third Graders correctly stated that the mixture with two sugar cubes in four cups of water would taste sweeter than the mixture with two sugar cubes in six cups of water. In most cases they justified their correct answers by stating that there was more water in the second container. Performance was consistently higher for the Control Group and improved from Grades three to five for both groups (see Figure 32).

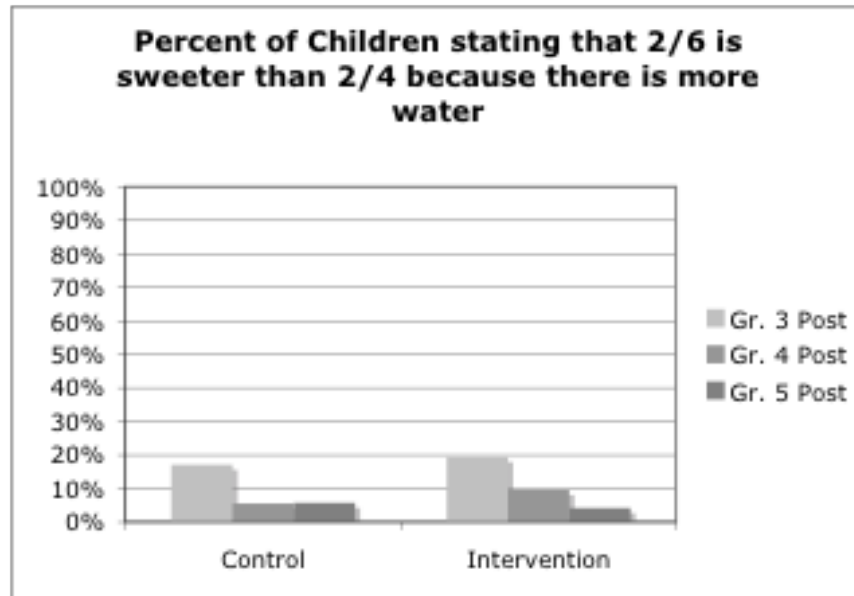


**Figure 32:** (2 cubes in 4 cups) sweeter than (2 cubes in 6 cups)

Wrong answers that the solutions would be equally sweet were justified by stating that the beakers had the same amount of sugar. This occurred for less than 10% of all Control students and for about 25% of all Intervention students. Examples of such justifications are: "Water has nothing in it, you're just putting sugar in it so it has the same taste" or "It doesn't matter how much water." These students seemed to see water as merely the means for delivering the total amount of sugar. They may have failed to realize that

sweetness is a property of an arbitrary sampling (e.g.: a gulp) rather than of the total amount.

Surprisingly, other students, mainly third Graders, believed that solution B was sweeter because there was more water to dilute the sugar (see Figure 33), as in the following examples: “The more water it has, the better the sugar cube will dissolve” or “The sugar eats the water, makes it turn into sugar.”

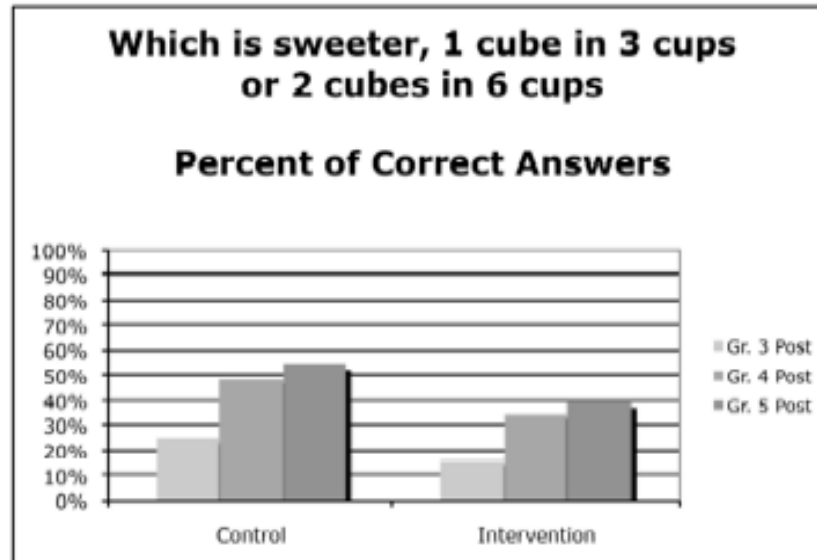


**Figure 33:** (2 cubes in 6 cups) claimed sweeter than (2 cubes in 4 cups) because “there is more water”

These two kinds of justifications support the idea that, at least in younger ages, students’ performance in mixture tasks may be more related to their understanding of how sugar and water interact to produce sweetness when they are mixed than to awkwardness with the mathematical structure.

### 2-Amount of Sugar Proportional to Amount of Water

In task 2, the ratios of sugar to water were equal. Children’s performance improved from Grades three to five (See Figure 34) and was again better for the Control students.



**Figure 34:** (1 cube in 3 cups) just as sweet as (2 cubes in 6 cups)

Slightly less than half of the correct answers were justified by noting that 1 in 3 was the same as 2 in 6. The following are examples of this kind of justification: “It’s like saying every three cups you have, you put a sugar cube in,” “One jar had double (or half) the sugar but also double (or half) the water,” “This (C) has less water and 1 sugar cube but this (D) has more water and two sugar cubes,” “D has double everything that C has so if you put the cubes in and took a drink it would taste same.” Other children who gave correct answers used a distribution approach: “Because if we took this one away and then took away 3 more cups of that they’d be the same but if we add 3 cups and then a whole sugar cube it’d be the same.” Some just reasoned qualitatively: “C has a little bit of water, so it would be just enough for it, D has a lot of water,” “Because C has less water and less sugar and D has more water and more sugar cubes.”

Wrong answers stated in 19% of the cases that C was sweeter, and in 64% of the cases that D was sweeter. Justifications for incorrect answers typically mentioned only one of the ingredients (sugar or water).

### 3-Different Ratios

In task 3 children had to compare two different ratios, with different values for each variable. Again, the Control Group performed better than the Intervention group (Figure 35). Here, approximately one third of the children justified correct answers (a) by comparing the ratio between the number of sugar cubes with the ratio between the number of cups of water, as in the following examples: “In E there is double the water but not double the sugar”; “4 is half of 8, but 2 is more than half of 3, so it will have extra flavor” or (b) by comparing the ratios of sugar to water, as follows: “F has 1 sugar cube for 2 cups water, E has less than 1 sugar cube for 2 cups water.”

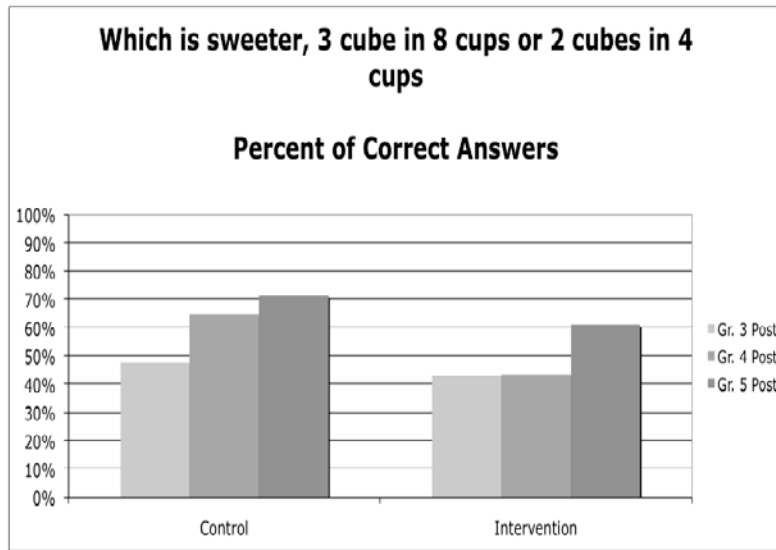
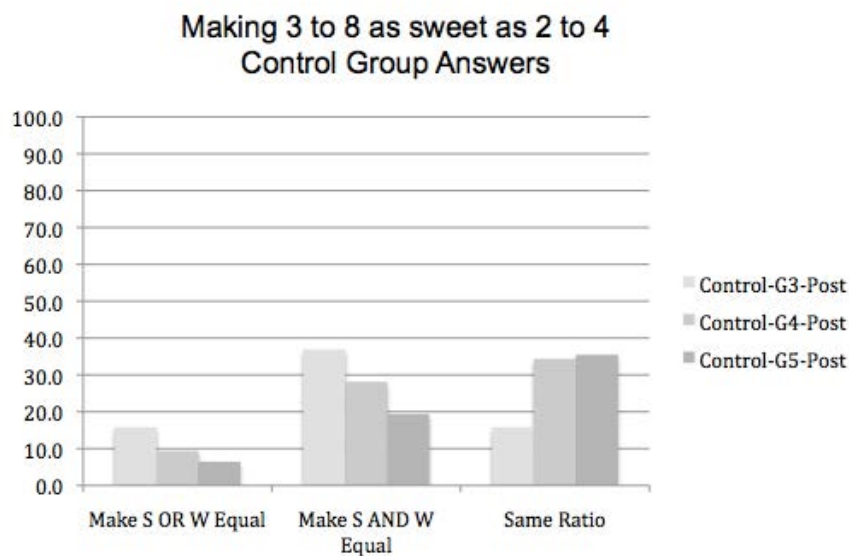


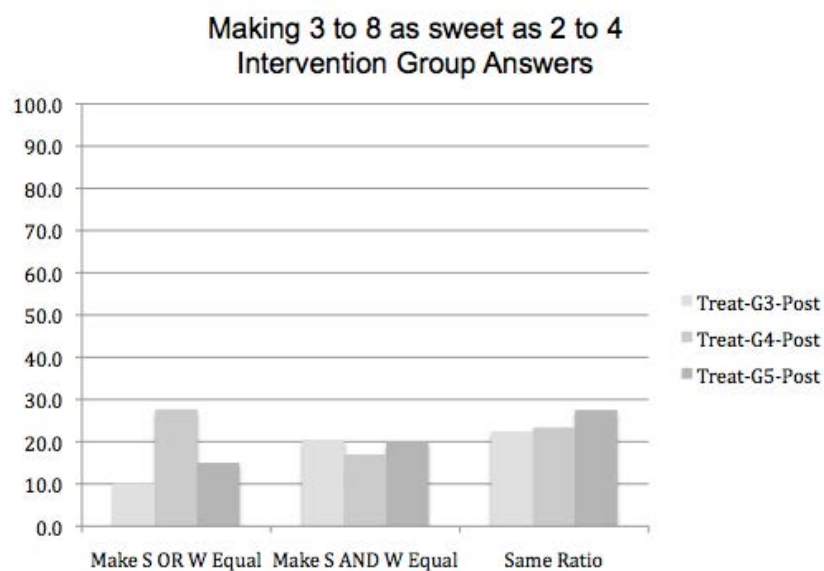
Figure 35: (3 cubes in 8 cups) less sweet than (2 cubes in 4 cups)

Justifications for incorrect answers typically mentioned only one of the ingredients (sugar) or did not mention precise quantification. As much as 44% of the children answered that E was sweeter and 8% judged that E and F tasted the same. The children who stated that E was sweeter justified their answers by the fact that there was more sugar in E, further explaining, for example, that “The water doesn’t matter, only the sugar cubes.” The children who said that E and F would taste the same explained their answers with statements that E had more water and more sugar as in the following example: “Because with only a little water and two cubes, F could be sweet. E has a lot of water and three sugar cubes, when they dissolve, they might taste the same.”

In task 3, children also answered the question “Is there any way we could make them taste the same even though they have different amounts of sugar?” For the Control Group, children’s performance systematically improved from Grades three to five (see Figure 36), with more children in higher Grades proposing solutions that would make the two mixtures have the same ratio of sugar to water, even though the amounts were different. For the Treatment group (Figure 37), the trend is less clear, but still more children proposed equal ratios in higher Grades.



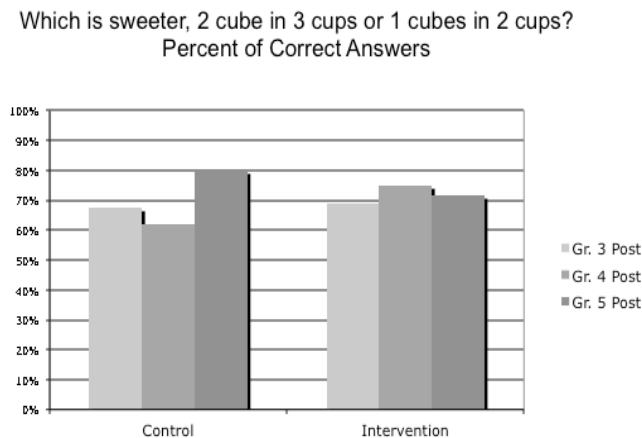
**Figure 36:** Making two mixtures equally sweet (Control)



**Figure 37:** Making two mixtures equally sweet (Treatment)

#### 4-Different Simplified Ratios

More than 60% of the children in both groups gave a correct answer to this task (Figure 38), with the fifth Graders in the Control Group showing the best performance.



**Figure 38: (2 cubes in 3 cups) sweeter than (1 cube in 2 cups)**

About one quarter of the fifth Graders in the Control Group justified their answers with statements that involved proportionality. For the other Grades, in both groups, less than 10% of the children offered proportional justifications for their answers. Instead, a little less than one third of the answers in all groups and Grades mentioned only either the difference in amounts of sugar or the difference in the amounts of water.

#### 5-Summary of Results

Our results indicate that children's early difficulties may rest on conceptions about how sugar and water contribute to sweetness. For most children only the amount of sugar matters; for a few, more water makes the mixture sweeter. The results also show no evidence of influence of the curriculum (which incidentally did not directly focus on sugar and water mixtures or proportionality).

Our data suggest that to prepare students to develop an understanding of proportionality as it relates to mixtures, explicit and systematic experiences with the resulting mixtures, together with a stronger focus on quantification, are needed.

---

### **F. Conclusions**

Our results support the following main conclusions:

First, at the beginning of Grade 3, students' concepts of weight, volume, material, and matter are different from, and often at odds with, those of scientists. Such concepts are typically overlooked in traditional science curricula.

More specifically, at the start of third Grade, children treated material, weight, and size as closely tied to their perceptual judgments of objects and materials rather than as objective properties that are tightly interconnected in an explicit model of matter. Further, they did not yet conceptualize physical quantities as measures that could be operated on additively and multiplicatively. They tended to view fractions as a fairly small set of numbers falling between integers (often only 1 or 2 fractions between consecutive integers). Likewise, they thought there were relatively few possible examples of distinct weights and lengths that would be greater than one example and smaller than another (even when the differences between the examples was striking). They did not clearly distinguish volume

from length, area, shape, and weight; they did not clearly distinguish density from weight and coordinate it with size (volume). Nor did they recognize that materials differed in density. By and large, these youngest students in our study also lacked a concept of matter as something that includes all solids and liquids, as something that takes up space and has weight, and that also includes gases and gas-like mixtures. Thus the conceptual relations in these students' knowledge network were very different from the relations observed among experts.

Second, Treatment students (those who received the Inquiry Project curriculum) made clear progress from Grade 3 to 5 in understanding weight, volume, density, material, and matter. There was evidence that they were engaged in a productive, but drawn-out process of reconceptualizing the concepts in their matter network. Although they still showed difficulty on various items and were not at ceiling on all many items even at the end of grade 5, the majority had made substantial progress in making changes that allowed them to meaningfully view matter as something that takes up space and has weight and that also includes gases as well as solids and liquids.

More specifically, by the end of Grade 3, over 90% of the Treatment children understood that the identity of a material remained unchanged with decomposition into little pieces; 88% understood that reshaping did not change the amount of material, and almost 70% linked weight to amount of material and consistently judged that reshaping did not change the object's weight or how it would affect a balance scale (up from 63%, 64%, and 41% at pretest, respectively). Grade 3 Treatment students made especially dramatic progress in their understanding that tiny things both have weight and take up space (up from 7% in the pretest to 67% in the Grade 3 posttest). They also made significant progress in understanding that reshaping did not change an object's volume (from 33% to 65% in the Blocks Rearrangement task, and from 25% to 45% in the Clay Transformation task). They made only limited progress, however, with regard to other aspects of volume (e.g., volume measurement, water displacement), "heaviness of materials," and matter by the end of Grade 3. The fact that children's understanding of materials and weight leads their understanding of volume, density, and matter is consistent with the Inquiry curriculum's assumption that weight and material can constitute productive entry points for learning about matter.

In Grade 4, although children did not make further progress with most of the weight or materials tasks, or the two tasks that assessed volume invariance across reshaping, they continued to make steady progress on all three of the density tasks and significant progress on volume measurement and water displacement, a focus of the 4<sup>th</sup> grade curriculum. Thus, by the end of Grade 4:

- Over half of the children (53%) were systematically distinguishing the heaviness of objects from heaviness of materials (up from 17% at pretest and 33% at the end of Grade 3);
- Over half (53%) were correctly inferring what material two "mystery" objects were made of using information about heaviness for size (not simply weight) (up from 12% at pretest and 19% at the end of Grade 3).
- Almost half (44%) were able to provide explanations for why different size objects weighed the same that involved consideration that the objects were made

of different materials and/or the smaller was made of a heavier kind of material (up from 10% at the pretest, and 25% at the end of Grade 3).

- Over half (56%) understood the amount of water displaced by a submerged object depended on its volume rather than its weight (up from 10% at pretest, and 21% at the end of Grade 3).
- Finally, more than a third (36%) correctly used the small cubes to judge the size of two rectilinear objects according to volume (rather than length or area). Although the majority still had difficulty with this task, this represents a significant improvement from pretest when only 5% used the cubes in this way, and the end of Grade 3, when 21% did.

In Grade 5, children were far more likely to have an explicit concept of matter and realize that even very tiny things have weight. By then 76% had come to understand that tiny pieces of clay take up space and have weight, and 86% understood that when sugar dissolves in water, the amount of sugar and its weight remain invariant (up from 44% at pretest). They tended to view matter as including solids, liquids, and gases; and understood that matter takes up space and has weight. [Three-quarters (76%) classified all solids and liquids as matter, (up from 37% at pretest), and almost two-thirds (63%) were also including some gases in these groupings (up from 25% at pretest). Further, 60% now explicitly stated that all matter takes up space or has weight (up from 10% at the end of grade 3), and 63% were systematically distinguishing the weight of objects from the heaviness of materials.]

However, the Treatment Students did not make further progress with volume and indeed showed some (slight) regressions on all the volume tasks. We suspect this may reflect the shift in curricular focus. Disappointingly, fewer than a third of the treatment students judged size on the basis of volume (as opposed to area) at the end of Grade 5. In addition, although the vast majority of students understood that even tiny things have weight, fewer than a third treated weight as a continuous dimension (with lots or an infinite number of weights between the weight of two balls). These findings point to the need to keep volume front and center even after grade 4, and to have explicit discussions about the granularity of weight.

Third, progress among the Control students (those who had the standard science curriculum) was much spottier. On some key tasks their progress was delayed relative to the Treatment students, although they eventually caught up by the end of Grade 5 (e.g., on the tasks probing their understanding the invariance of materials across grinding, understanding that shape change does not affect object weight, balance, or volume when objects were composed of unit cubes, and understanding volume measurement.) On a number of other tasks, however, they either made no progress at all across grades 3 to 5 or very limited progress, so that by the end of Grade 5 there remained significant differences between the Treatment and Control students. More specifically, the vast majority of Grade 5 Control students still did not think tiny pieces of clay both took up space and had weight (31% Control compared to 76% Treatment), that the amount and weight of sugar was conserved on dissolving (46% Control compared to 85% Treatment), that the volume of clay was conserved on reshaping (26% Control compared to 42% Treatment), or that

volume, not weight was relevant on water displacement. (14% Control compared to 47% Treatment). The majority did not systematically distinguish the heaviness of materials from the heaviness of objects (34% Control compared to 63% Treatment; they did not form systematic matter groupings that included at least some gases as well as all solids and liquids (29% Control compared to 63% Treatment); nor did they articulate the idea that all matter has weight or takes up space (9% Control compared to 60% Treatment). On these tasks, their developmental trajectories were relatively flat and their level of performance significantly lagged behind that of the Treatment students. It appears that the Inquiry curriculum was more likely than the standard curriculum to help students develop a sound model of matter.

Fourth, there were encouraging indications that young students can make use of quantity lines (number line where the units are in units of measure such as grams or cubic centimeters) to systematically order sets measurements and to think of physical quantities as dimensions. However, they tended to regard the distinct values on quantity lines and number lines as being “few and far between.” In principle, treating quantities and numbers as dimensions would seem to be consistent with the emergence of proportional reasoning. There was evidence in the interviews (Granularity of Weight, Length, Number; Sweetness) that both Treatment and Control students made some progress from Grades 3 to 5 in these regards, although in both cases the progress was limited and there were no differences between Treatment and Control students. One suspects that proportional reasoning about physical quantities and understanding quantities as continua may need more direct nurturing.

Finally, our data showed there were many similar relationships across tasks for Treatment and Control students, despite their different levels of performance and curricular histories. For example, understanding that tiny pieces of clay have weight and take up space was strongly related to developing precursor understandings of density and matter; understanding volume measurement was strongly related to understanding the measurement of area and focusing on volume as the invariant in the Blocks Rerangement task; understanding the granularity of weight was related to both understanding of the granularity of number and understanding that tiny pieces have weight. These relations may reflect conceptual constraints that need to be exploited in any Learning Progressions curriculum.

---

### III. References & Reports

- Carraher, D.W., Sue Doubler, Jodi Asbell-Clarke, Roger Tobin, Analucia D. Schliemann, Carol L. Smith, Marianne Wiser, Paul Wagoner, Chunhua Liu, Sally Crissman, Nick Haddad, Sara Lacy, Lynn McCormack (2009). "Seeing weight, grasping density.", Poster presented at annual meeting of the National Science Foundation, DRK-12 Principal Investigators' Meeting, November, Washington, D.C. [Carraher-2009-NSFPoster.pdf]
- Carraher, D. W. (1996). Learning about fractions. In L. P. Steffe, P. Nesher, G. Goldin, P. Cobb & B. Greer (Eds.), *Theories of Mathematical Learning*. Hillsdale, NJ: Erlbaum.
- Carraher, D.W. & Cayton-Hodges, G. (2011). Distinguishing Volume from other Magnitudes. Symposium paper presented at the 2011 Annual Meeting of the American Educational Research Association, April, New Orleans, LA.
- Carraher, D.W. & Schliemann, A.D. (1991). Children's understanding of fractions as expressions of relative magnitude. *Proceedings of the XV International Conference Psychology of Mathematics Education*. Assisi, Italy.
- Carraher, D.W., Smith, C.L., & Kavanagh, C. (2011). The Granularity of Number, Length and Weight. Symposium paper presented at the 2011 Annual Meeting of the American Educational Research Association, April, New Orleans, LA.
- Carraher, D.W., Smith, C.L., Wiser, M., Schliemann, A.D. & Cayton-Hodges, G. (2009) Assessing students' evolving understanding about matter. Paper presented at the "Learning Progressions in Science" Conference, Iowa City, IA, June 24-26, 29pp. [Carraher-2009-LeaPS.pdf & Carraher-2009-LeaPS-ppt.pdf (PowerPoint)]
- Doubler, S., Asbell-Clarke, J., Carraher, D.W. & Tobin, R. 2007. The Inquiry Project: An IMD Learning Progression, Paper presented at the NSF DRK-12 Annual Meeting for PI's. November, Washington, D.C. [Doubler-2007-DRK-12.pdf]
- Doubler, S., Carraher, D.W., Asbell-Clarke, J., Tobin, R. (2007). The Inquiry Project (#0628245): Year 1, Annual report written for the National Science Foundation, Aug. 14, 2007, 70 pp. [Doubler-2007-NSFrpt.pdf]
- Doubler, S., Asbell-Clarke, J., Carraher, D.W., Tobin, R. (2008). The Inquiry Project (#0628245): Year 2, Annual report written for the National Science Foundation, Aug. 15, 2008, 15 pp. [Doubler-2008-NSFrpt.pdf]
- Doubler, S., Carraher, D.W., Asbell-Clarke, J., Tobin, R. (2009). The Inquiry Project (#0628245): Year 3, Annual report written for the National Science Foundation, Jul. 29, 2009, 44 pp. [Doubler-2009-NSFrpt.pdf]
- Massachusetts Department of Education (2007). 2007 Massachusetts Comprehensive Assessment System Report, 231 pp., available at <http://www.mcasservicecenter.com/documents/MA/Technical%20Report/2007/04-09-08%202007%20MCAS%20Tech%20Rpt%20Final%20PDF.pdf>.

- National Research Council (2007). *Taking science to school: Learning and teaching science in Grades K-8*. Washington, D.C.: National Academies Press.
- Schliemann, A.D. & Carraher, D.W. (1992). Proportional reasoning in and out of school. In P. Light & G. Butterworth (Eds.) *Context and Cognition*. Hemel Hempstead, Harvester-Wheatsheaf, 47-73.
- Schliemann, A. D., Liu, C., Wagoner, P., & Carraher, D. (2011). Understanding Materials, Weight, Size, Volume, and Density. Symposium paper presented at the 2011 Annual Meeting of the American Educational Research Association, April, New Orleans, LA.
- Schliemann, A. D., & Nunes, T. (1990). A situated schema of proportionality. *British Journal of Developmental Psychology* (8), 259-269.
- Schliemann, A. D., Wagoner, P., Liu, C., & Carraher, D. (2011). Understanding Sugar and Water Mixtures. Symposium paper presented at the 2011 Annual Meeting of the American Educational Research Association, April, New Orleans, LA.
- Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and atomic-molecular theory. *Measurement: Interdisciplinary Research and Perspectives*, 14(1-2), 1-98.
- Smith, C.L., Wiser, M., & Doubler, S. (2011). Abstracting a general concept of matter among Grade 3-5 students: Lessons from the Inquiry Project. Symposium paper presented at the 2011 Annual Meeting of the Jean Piaget Society, June, Berkeley, CA.
- Smith, C. L., Wiser, M., Carraher, D. (2010). Using a comparative longitudinal study with upper elementary school students to test some assumptions of a Learning progression for Matter. Paper presented at NARST, Philadelphia, PA, March 24, 2010.
- Wiser, M., & Smith, C. L. (2008). Teaching about matter in Grades K-8: When should the atomic-molecular theory be introduced? In S. Vosniadou (Ed.), *International handbook of research on conceptual change*. Hillsdale, NJ: Erlbaum.
- Wiser, M., Smith, C.L., Asbell-Clarke, J., & Doubler, S. (2009). Developing and Refining a Learning Progression for Matter: The Inquiry Project: Grades 3-5. Paper presented at AERA Learning Progressions for Matter Symposium, San Diego, CA, April 14, 2009.
- Wiser, M., Smith, C.L., Doubler, S. & J. Asbell-Clarke. Learning Progressions as tool for curriculum development: Lessons from the Inquiry Project. (2009) Paper presented at the "Learning Progressions in Science" Conference, Iowa City, IA, June 24-26, 2009, 26pp. [Wiser-2009-LeaPS.pdf]
- Wiser, M., Smith, C., & Doubler, S. (in press). Learning Progressions as Tool for Curriculum Development: Lessons from the Inquiry Project. To appear in A. Alonzo & A. Gotwals (Eds.), *Learning Progressions in Sciences*. Sense Publishing

