DEVELOPMENT OF A RESEARCH-BASED LEARNING PROGRESSION FOR
MIDDLE SCHOOL THROUGH UNDERGRADUATE STUDENTS' CONCEPTUAL
UNDERSTANDING OF SIZE AND SCALE

by

Cesar Delgado

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Education)
in The University of Michigan
2009

Doctoral Committee:

        Professor Joseph S. Krajcik, Chair
        Professor Brian P. Coppola
        Professor Edward A. Silver
        Associate Professor Elizabeth A. Davis

DEDICATION

A mi familia

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

Size and scale are crosscutting ideas integral to scientific understanding. However, research shows that students have little understanding of the size of objects, particularly objects too small to see with the unaided eye. Using a cross-sectional study with 101 middle-school through undergraduate students, a teaching experiment with 24 middle school students, and a theoretical task analysis, I built a learning progression for one-dimensional size and scale that focused on four aspects: ordering by size, grouping by size, relative scale (how many times bigger one object is than another), and absolute size. Over 90% of students develop their conceptual understanding of size and scale by connecting the aspects of size and scale in a specific order. Learners first connect the two qualitative aspects: ordering and grouping. They next connect ordering and relative scale, and then connect ordering and absolute size. Learners connect the two quantitative aspects, relative scale and absolute size, last.

A major assumption underlying the learning progression theory is that they can be used to design learning materials that will advance student understanding. To provide empirical support for this idea, a teaching experiment was conducted to determine if students could advance in the manner suggested by the learning progression. The teaching experiment resulted in statistically and educationally significant learning gains. It showed that students increase the accuracy of their factual knowledge in tandem with their increased connectedness of knowledge, by establishing landmark objects that help them construct a mental measurement line. Measurement units including micrometers and nanometers were shown to be powerful tools for student learning about the unseen world.

The task analyses revealed a variety of strategies that learners can use in addressing size and scale tasks. Types of strategies include using recall, using a single aspect, and employing the connection across aspects. Required logical and mathematical skills that precede proportional reasoning were also identified.

This dissertation provides a model of how to develop a learning progression for a core idea, and showed that a learning progression can inform the design of curriculum materials that can move students to more advanced levels on the learning progression.

CHAPTER 1: INTRODUCTION

This dissertation characterizes a trajectory that students tend to take in gradually constructing conceptual understanding of size and scale. It shows how a carefully designed instructional unit based on a learning progression can spur this understanding and increase students' factual knowledge of the size and scale of important scientific objects. This dissertation also analyzes in detail what mathematical and logical skills and content knowledge learners need in order to construct their understanding of size and scale. Understanding the size and scale of science concepts across disciplinary boundaries may help learners develop a more connected knowledge of science. This learning may also help learners grow in their mathematical reasoning, and even in their abstract thinking.

In this chapter, I explain the need to need for a learning progression for size and scale, describe the scope of this dissertation, and describe the approach I took in iteratively developing this learning progression.

The concepts of size and scale are important both in science and for science learning. Because of the importance of size and scale to science, students need to comprehend size and scale in order to understand and learn science. However, research shows that learners have difficulty learning about the size and scale of scientifically important objects; for instance, some people in every age group from children to adults mention a small visible object when asked for the smallest object of which they can think (Waldron, Spencer, & Batt, 2006). A lack of understanding of size and scale may be behind student lack of understanding about the particulate nature of matter:

> Students ascribe macroscopic properties to particles. For example, particles may explode, burn, contract, expand and/or change shape. This primitive reasoning

prohibits understanding of the nature of a chemical reaction (Kind, 2004, p. 13).

A good understanding of size and scale could be instrumental in better understanding not only atomic-molecular explanations for chemical phenomena, but many other important scientific concepts as well. The purpose of this dissertation is to develop a learning progression to guide the development of more effective curriculum for size and scale. In order to do so, I carry out empirical research to fill significant gaps in the literature.

<div align="center">The Importance of Size and Scale in Science</div>

Size is a characteristic of every object, and is the magnitude or extent of the object. The size of an object is established by comparing it to a standard, which functions as a scale. Scales are "the spatial, temporal, quantitative, or analytical dimensions used by scientists to measure and study objects and processes" (Gibson, Ostrom, & Ahn, 2000, p. 219). This dissertation is concerned with what learners know and how they think about the linear, one-dimensional measure of length of objects ranging in size from "submacroscopic" (too small to see with the unaided eye, e.g., an atom) to large macroscopic (e.g., a planet).

According to *Science for all Americans*,

> Science is a process for producing knowledge. The process depends both on making careful observations of phenomena and on inventing theories for making sense out of those observations. (Rutherford & Ahlgren, 1990, Ch. 1).

In both theory and observation, size and scale are important. Concerning theory, size was one of the few characteristics the atom was postulated to have in classical Greek and seventeenth century European theories of matter (Berryman, 2004; Chalmers, 2005); the atomic nature of matter is arguably the most important scientific hypothesis (Feynman, 1963, I, i 1-2). Concerning observation, every physical object can be characterized partially in terms of its size. Some regular physical objects can be well specified by just three characteristics: size, material, and shape (e.g., a copper sphere with a diameter of 10 cm). In the application of scientific knowledge, scale is a paramount consideration.

Objects or organisms of different sizes behave differently, even if scaled up faithfully. Scale has been called "the quintessential aspect of every physical theory" (Bazant, 2002), a fundamental conceptual problem in ecology (Levin, 1992) and "one of the major gateways to the modern world of science" (Hawkins, 1978).

There are many historical examples showing the important role of size and scale in the development of scientific knowledge. For instance, before 1675, the existence of microorganisms had been hypothesized but not confirmed. Then van Leeuwenhoek visualized microorganisms using a microscope he had invented. This momentous event opened to scientific study the microscale world (the region of sizes spanning from around 1 micrometer [μm], or thousandth of a millimeter, to roughly 100 μm - the thickness of a hair). This event transformed biology, founding the fields of microbiology and bacteriology, and greatly benefited the field of medicine and thus, human health and wellbeing. Before 1925, scientists thought that the entire universe consisted of the Milky Way galaxy. Using a new 100-inch optical telescope, the largest in existence at that time, Edwin Hubble detected some stars that were too distant to belong to the Milky Way; his discovery changed our view of the universe. Even today, emerging science tends to be at very large or small scales. The nanotechnology revolution we are currently experiencing, and which deals with objects that are one to 100 nanometers (nm; millionths of a millimeter, or billionths of a meter) in at least one dimension, is possible in part due to atomic force microscopes developed in the 1980s. Working at novel scales has thus opened up entirely new disciplines of science. The measurement and comprehension of sizes and distances is a core component of disciplinary domains: from the vast distances of astronomy, measured in billions of light-years; through the minute cells and bacteria of biology, invisible to the naked eye; to the counterintuitive quantum phenomena at the nanoscale, where matter exhibits unique properties.

New tools can open to scientific study entirely new size regions, but size is of key importance in selecting tools and models to study all phenomena. The size of an object will determine whether we can study it with our unaided eyes, or whether it will require magnification; and whether a magnifying lens, an optical microscope, or an electron

microscope can provide the required magnification. The selection of a model is also impacted by size: the classical Newtonian mechanical model is adequate for modeling the collision of billiard balls, but for phenomena at the nanoscale, quantum effects must be considered. Recent research confirms that scientists, as well as the practitioners of many other professions and trades, consider scale to be fundamental in their work (Jones & Taylor, 2009).

<div align="center">Rationale for the Dissertation</div>

<div align="center">*The Importance of Size and Scale in Science Education*</div>

US standards documents in science identify scale as a concept that pervades science, and that can be used to unify student learning across disciplines, topics, and grades; it is a tool that helps students understand the world (American Association for the Advancement of Science [AAAS], 1993, Ch. 11; National Research Council [NRC], 1996). According to the *Benchmarks for Science Literacy* (AAAS, 1993), "Particularly important senses of scale to develop for science literacy are the immense size of the cosmos [and] the minute size of molecules…" (p. 276). The *Benchmarks* also mention size as an important characteristic of atoms: "All matter is made up of atoms, which are far too small to see directly through a microscope…" (4D/M1, p. 78).

As new fields of science and technology emerge, such as nanoscale science and technology, science instruction and curriculum materials need to change accordingly (Gilbert, De Jong, Justi, Treagust, & Van Driel, 2002, p. 395). One of the goals of the National Science Foundation-funded National Center for Learning and Teaching in Nanoscale Science and Engineering (NCLT) is to suggest ways of incorporating appropriate nanoscale science concepts into the K-12th grade curriculum. Since a good understanding of size and scale is a requirement for inquiry into nanoscale science and engineering (Waldron, Sheppard, Spencer, & Batt, 2005, p. 375), one of the areas that the NCLT has been researching is students' knowledge and learning of size and scale. Size and scale constitute one of nine big ideas for nanoscale science and engineering, for not only does size define the nanoscale, size also determines the dominant forces and physical laws that predict how objects of a certain size range will behave (Stevens,

Sutherland, & Krajcik, in press).

The Importance of Size and Scale in Mathematics Education

Measurement is closely related to size and scale. Measurement is one of five content strands contemplated in the *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics [NCTM], 1989). This document stresses the importance of students being able to understand measurable attributes of objects, along with the units, processes, techniques, and tools of measurement. Math standards documents also include the expectation that early elementary students be able to use length to order and group objects (NCTM, 1989). The importance of size and scale to mathematics education, though, may go beyond being a component of the content strand of measurement. Resnick described a sequence of four kinds of mathematical thinking (1992). In this account, based on an interpretive review of research on the development of mathematical knowledge, the earliest stage is called the "mathematics of *protoquantities*". This kind of mathematical thinking involves qualitative reasoning about amounts of physical material via "actions that can be performed directly on physical objects" (Resnick, 1992, p. 403), including combining, comparing, and ordering by size. The next kind of mathematical thinking is the mathematics of *quantities*, again involving physical material, but now with numerically quantified amounts, such as *n* inches, and operations such as dividing a measured amount into equal shares. These two kinds of mathematical reasoning involving physical material underlie the two subsequent stages: the mathematics of numbers, and the mathematics of operators, which no longer involve physical material (Resnick, 1992). In this account, then, experiences that involve thinking about and measuring one-dimensional size are fundamental in the development of mathematical thinking. In sum, theoretical accounts of the development of mathematical knowledge as well as standards documents in mathematics and science education all point to the importance of thinking about and having experiences with the size and scale of objects.

The Role of Size and Scale in the Development of Abstract Thought

Gattis (2001) summarized a body of research that suggests that spatial cognition is used

as a "platform for building new structures essential for higher cognitive processes" (p. 1). The spatial organization of elements to be remembered may help to organize memory; to communicate relations (e.g., opposing political currents as being on the left and right), and as logical structures (allowing a learner to "preserve order between elements, to identify relations between dimensions, and to identify polarity within a dimension" – Gattis & Dupeyrat, 1999, p. 165). Many of these hypothesized ways in which spatial cognition is used as a platform involve a single dimension, for instance creating a mental spatial array to mentally depict order and help us to make transitive inferences. Transitivity allows one to know the relationship between A and C if the relationship between A and B and the relationship between B and C are known, for example, if A = B and B = C, then A = C, or if A < B and B < C, then A < C. Gattis states that "selecting and adapting an appropriate spatial structure is a demanding task" (2001, p. 6), and the capacity of the reasoner may play an important role in determining how frequent and how feasible it is to use spatial schemas to aid abstract thought. The construction of a robust conceptual idea of size and scale may thus be useful in science, in the development of mathematical understanding, and even in the development of abstract thought.

Student Difficulties Learning Size and Scale

Current US K-12 science and mathematics curriculum and instruction may not be successfully addressing size and scale, as educational researchers have identified many areas of difficulty or lack of knowledge for learners that are related to size and scale. The finding that students do not have a good idea of the size of the atom (often overestimating its size) is common in the literature (e.g., Brook, Briggs, & Driver, 1984; Griffiths & Preston, 1992). At the opposite extreme of scale, students have trouble conceptualizing the size of very large objects, for instance, often overestimating the size of the Earth relative to the size of the sun (e.g., Phillips, 1991; Hapkiewicz, 1999). US students fare worse on geometry and measurement than on other mathematical topics in international comparisons (National Center for Education Statistics [NCES], 1996). For instance, fewer than half of seventh graders were able to determine the length of a line that was not aligned to the zero point of a ruler (Clements, 2003); over a quarter of 8th grade students were unable to determine the length of an object with the zero point of the ruler aligned,

when the object reached to the midpoint of 0.8 and 0.9 meters (NCES, 2003). Data from the 1996 National Assessment of Education Progress (NAEP) show that eighth- and twelfth-grade students have high rates of success on simple conversions (within the same system) between units of linear measure, but "only about half the twelfth-grade students could successfully answer an item involving more than one conversion." (Kenney & Kouba, 1997, p. 147). While students were not tested on conversions *between* systems of units, even experts may have trouble with these. In 1999, NASA's $125 million Mars climate orbiter was lost because the two different engineering teams involved use different systems of measurement, one metric and the other English (Lloyd, 1999).

The Need for a Learning Progression for Size and Scale

Recent publications have suggested that a "learning progression" can guide the principled development of effective curriculum, instruction, and assessment for science (Duschl, Schweingruber, & Shouse, 2007; Smith, Wiser, Anderson, & Krajcik, 2006; Wilson & Bertenthal, 2005). Learning progressions are

> descriptions of the successively more sophisticated ways of thinking about a topic that can follow one another as children learn about and investigate a topic over a broad span of time (e.g., 6 to 8 years) (Duschl et al., 2007, p. 214).

Learning progressions, described in detail in Chapter 2, are a promising approach to thinking about instruction because they explicitly consider longer-term aspects of learning, paying close attention to how students gradually construct more connected knowledge. They include instructional strategies that can help students in this process, and assessments to measure progress and inform instruction. A learning progression cannot be guided solely by experts' determination of a content-oriented logical sequence, for experts are likely to have forgotten the conceptual challenges and difficulties they encountered as learners (Bransford, Brown, & Cocking, 1999), and may not take into account learners' unexpected and often counter-intuitive difficulties and pathways to understanding (Hiebert & Carpenter, 1992). Learning progressions should rely on research on student learning when available (Smith et al., 2006) and include "actual observed performances by real students." (C. Anderson, 2008, p. 4). Learning

progressions have various components:

> Learning progressions consider the interaction among the strands of scientific proficiency in building understanding (know, use, and interpret scientific explanations of the natural world; generate and evaluate scientific evidence and explanations; understand the nature and development of scientific knowledge; participate productively in scientific practices and discourse.) (Duschl et al., 2007, p. 221).

Finally, learning progressions should be organized around "big ideas" – the central, core concepts in a domain (Smith et al., 2006). The US science and math curriculum has been characterized as being a mile wide but an inch thick (Schmidt, McKnight, & Raizen, 1997). What typically results from such a curriculum is not robust, connected knowledge but a "splintered vision" (Schmidt et al., 1997). Focusing on big ideas in science education, including size and scale, may help address the fragmentation of knowledge, particularly if instruction is guided by a research-based learning progression.

Scope of This Dissertation

Content

This dissertation focuses on students' ideas about one-dimensional size and scale; concepts of two-dimensional size (area), 3-dimensional size (volume), and surface area to volume ratio are also important in science education but are more complex, and are outside of the scope of this dissertation. Since I want to inform the teaching and learning of nanoscale science and engineering concepts at the pre-university level, I study what students know about objects at the micro-, nano-, and sub-nanoscales; I also include macroscopic object for comparison purposes. While there is a large body of research concerning children's construction of concepts of space, measurement, and geometry, and of learners' estimation strategies for length, until recently this research almost invariably focused on objects that are large enough to see but not too large to perceive all at once, i.e., everyday objects. However, as the 2009 version of the *Benchmarks for Science Literacy* (AAAS) note:

> Natural phenomena often involve sizes, durations, and speeds that are extremely

small or extremely large. These phenomena may be difficult to appreciate because they involve magnitudes far outside human experience. (11D/M3, Ch. 11)

At these scales of size, the *Benchmarks* found "No applicable research findings" for the common theme of scale (AAAS, 1993). Since then, there has been research into understanding of size and scale (see Chapter 2), but significant gaps remain. A large drop in the accuracy of learners' conceptions of the size and scale of objects below visibility was found (Tretter, Jones, & Minogue, 2006), lending further urgency to study the way students think and learn about the submacroscopic world.

Grades

Submacroscopic objects such as cells, viruses, and atoms are mentioned in standards documents for science education for grades six through eight (AAAS, 1993). As middle school students learn to use ratios and powers of 10, and develop greater familiarity with very large and very small numbers (AAAS, 1993), they should increasingly be able to think about the size of objects of any magnitude. Thus, I begin my investigation of students' knowledge at grade 6. Undergraduates at a selective research university are a best-case scenario of what students can achieve by the end of high school with existing curricula.

Ways of Thinking About Size and Scale

As I explain in Chapter 3, a critical first step in my design process is to clearly define the construct of interest (Pellegrino et al., 2008). I identified four different ways of thinking about size and scale from the literature. I term these different ways of thinking about size and scale "aspects". The four aspects include ordering and grouping by size, which are qualitative and consistent with Resnick's *protoquantities*; and relative scale and absolute size, which are quantitative and correspond to Resnick's *mathematics of quantities* (1992). I define the four aspects and review relevant research for each, in Chapter 2.

Types of Knowledge

In the last five years a body of research on learners' understanding of the size and scale of very large and very small objects has begun to develop. One strand of research, by Jones, Tretter, and colleagues, has characterized the mean accuracy of learners' factual

knowledge of the size of objects for fifth grade through graduate students (Tretter, Jones, Andre, Negishi, & Minogue, 2006; Tretter, Jones, & Minogue, 2006), experienced and pre-service teachers (Jones, Tretter, Taylor, & Oppewal, 2008), and visually-impaired students (Jones, Taylor, & Broadwell, in press). Other recent studies have investigated the mean accuracy of factual knowledge of submacroscopic objects in learners of all age groups (e.g., Waldron et al., 2006; Batt, Waldron, & Broadwater, 2008; Castellini et al., 2007). Some of these studies begin to go beyond factual knowledge and delve into conceptual knowledge, but are not sufficient to inform the development of a learning progression. These studies' investigations of factual knowledge have assessed student knowledge via the four aspects mentioned earlier, but examining each aspect in isolation. An important component of conceptual understanding is the *connections* between pieces of knowledge, according to both constructivist theory and empirical studies of experts and novices (e.g., Piaget, 1983; Linn, Davis, & Eylon, 2004; diSessa 1988; Chi & Ceci, 1987; Bransford et al., 1999; Snir, Smith, & Grosslight, 1993). In this study, then, I investigate students' factual knowledge and connectedness of knowledge. Since I cannot directly observe students' mental connections, I infer them using tasks designed to make their thinking visible (Pellegrino, Chudowsky, & Glaser, 2001) by observing the *consistency* of individual students in their ideas about the size and scale of objects *across* aspects.

Research Questions

The current research base, which focuses primarily on the mean factual knowledge of learners by age group or grade, is insufficient to generate a learning progression. The purpose of this dissertation is to fill this gap. I envision this dissertation as the first two cycles of a design research effort: an "iterative, situated, and theory-based [attempt] simultaneously to understand and improve educational processes…" (diSessa & Cobb, 2004, p. 80). Successively more refined iterations of the learning progression guide the development of improved instructional interventions for size and scale, which in turn inform the revision and improvement of the learning progression.

Teaching experiments or longitudinal studies provide the strongest support for a

proposed learning progression (Duschl et al., 2007). Due to the logistical complications of a longitudinal study, and the difficulty of designing curriculum for a teaching experiment when little is known about conceptual understanding, I first used a cross-sectional design to investigate student understanding of size and scale and generate a first iteration of a learning progression. I considered both factual knowledge and consistency of knowledge. For both factual knowledge and consistency of knowledge, I described average performance by grade and also examined student responses for *patterns* that may be useful in defining qualitatively different levels of thinking.

The following research questions guided Study One:

1) What do middle school through undergraduate students know about one-dimensional size and scale, given their current curricular experiences?
   A) What is students' factual knowledge of the size of important objects in science, by year in school?
   B) What patterns of factual knowledge are observed in students?
   C) What connections (as inferred from consistency) do students make across aspects of size and scale, by year in school?
   D) What patterns of consistency are observed in students?

The results from Study One, presented in Chapter 4, allowed me to build the first iteration of a learning progression (LP1), which guided the development of a focused curriculum on size and scale. Study Two investigated the impact of this curriculum on middle school students. It provided stronger evidence than the cross-sectional Study One concerning the actual paths of learning that students undergo when provided with focused opportunities to learn about size and scale, thus informing the second iteration of the learning progression (LP2). Study Two also provided additional information about the upper anchor, instructional activities, and the relationship between factual knowledge and consistency of knowledge. The following research questions guided Study Two:

2) What do middle school students know about one-dimensional size and scale, after a

focused curricular experience?

A) What is students' factual knowledge of the size of important objects in science?

B) What is students' consistency across aspects of size and scale?

C) How do the patterns of factual knowledge and consistency differ from those of students who have not experienced the focused curriculum, if at all?

The results of research question 2 are presented in Chapter 5, and provide data towards confirming the accuracy of and/or revising LP1 in order to generate LP2.

Since the learning progression includes both factual knowledge and consistency of knowledge, it is important to further examine the relationship between factual knowledge and consistency of knowledge. As Chi and Ceci (1987) note, examining how content knowledge interacts with the development of meta-knowledge is a promising strategy in studying cognitive development (p. 135). I conducted a theoretical task analysis for each aspect of size and scale, and used the data from Study One and Study Two, in order to address the following research question:

3) What is the relationship between factual knowledge and consistency of knowledge, if any?

My findings for research question 3 are presented in Chapter 6. I then use the findings from all three research questions to develop LP2, presented in Chapter 7. Chapter 8 summarizes my findings and describes how they build upon and extend prior research, along with implications for curriculum and instruction of the important constructs of size and scale.

Chapter Summary and Overview

In this chapter I reviewed the importance of size and scale to science as well as to science and mathematics learning, and presented an argument for the need for a learning progression to guide efforts to improve student learning in this area. Size and scale are important in science because the extent of an object or phenomenon is a measurable

property that can impinge upon model and tool selection. Emerging science tends to be at extremes of size, even today with the advent of nanoscale science and engineering. Most importantly for science learning, size and scale constitute a crosscutting theme that can help students develop a more connected understanding of science and the world around them. Experiences interacting with and measuring amounts of physical substance are thought to underlie mathematical learning about numbers and operations, and spatial reasoning is thought to be a platform for abstract thought. Since prior research shows that students do not have extensive knowledge or conceptual understanding of size and scale, better curriculum materials and instruction need to be developed. A learning progression can guide this development process, but existing research is insufficient for the generation of a learning progression. Given the paucity of research into this area, I identified the need to characterize conceptual understanding of size and scale and examine its relationship to factual knowledge.

In this chapter I also presented the cyclical nature of my investigation. The existing literature and a cross-sectional study informed the first iteration of the learning progression, which guided the development of curriculum and instruction implemented in a teaching experiment; these, along with task analyses, allowed the revision and refinement of the learning progression. I focus on one-dimensional aspects of size and scale, including ordering by size, grouping by size, relative scale, and absolute size, for middle and high school.

In Chapter 2, I review the relevant findings from previous research that inform my dissertation and present a theoretical framework. In Chapter 3, I describe my methods, including the participants and methods of analysis for each study. In Chapter 4, I present my findings about the existing state of knowledge in middle school to undergraduate students, given the various curricula they have experienced. I also present LP1, built from these cross-sectional data. In Chapter 5, I describe how LP1 guided the development of a 12-hour curriculum unit for size and scale. I then present my findings about the state of knowledge in the middle school students after they experienced the curriculum unit. In Chapter 6, I examine the relationship between consistency and factual knowledge through

theoretical as well as empirical analyses using the data from Study One and Study Two. I use the findings from both studies and the task analyses to generate LP2, presented in Chapter 7, and summarize my findings and relate these to the literature in Chapter 8.

CHAPTER 2: THEORETICAL FRAMEWORK

In the previous chapter, I discussed the importance of size and scale and the need for a
learning progression to guide the development of improved curriculum and instruction
for this important topic. In this chapter I discuss the literature on learning progressions
and situate these in terms of learning theory. I describe the components of a learning
progression, including lower and upper anchors and intermediate levels of understanding.
I also review the literature relevant to size and scale in order to define what is known
about learners' understanding of size and scale, particularly at submacroscopic sizes, and
what research is still lacking to inform the development of a learning progression.

Theories of Learning and Learning Progressions

A very basic (and outmoded) conception of learning, behaviorism envisions the learner as
an "empty vessel" to be filled with knowledge, without taking into account the learner's
prior knowledge (e.g., Skinner, 1971). This model of learning considers understanding to
be composed of small, discrete pieces of factual or procedural knowledge. Behaviorist
conceptions of learning were challenged by Piaget's theories of learning (e.g., Piaget
1983), which envision the learner as actively building knowledge upon his or her prior
knowledge and ideas (see also Ausubel, 1968). Piagetian views of learning privilege
higher-order thinking processes and the connections between pieces of knowledge.
However, one aspect of Piaget's theories has not held up well against empirical evidence:
the postulated content-independence of developmental stages of reasoning based on
logical structures (Chi & Ceci, 1987). Research studies examining the differences
between novices and experts over the last three decades have found an important role for
factual or content knowledge in expertise, including the domain specificity of expertise
(e.g., Chi, Feltovich, & Glaser, 1981; Ericsson, Krampe, Tesch-Roemer, 1993; Chi &
Ceci, 1987). Specifically in the case of size and scale, domain specificity was found by
Tretter, Jones, and Minogue (2006):

> Objects at large scale were hard to conceptualize for experts whose specialty was at small scale because they don't have any reference points for mentally jumping to that new world… The opposite tended to be true for those experts whose specialty was at the large scale. An astrophysicist described her uncertainty about small-scale objects by saying, "I'm used to thinking in galaxy sizes" (p. 1078)

However, the renewed emphasis on factual knowledge in expert-novice research does not imply a return to behaviorism, for it has also found that the *organization* of factual knowledge is key; experts in various domains have been found to have well-connected knowledge (see Chi & Ceci, 1987 for a review). It is now considered that higher-level, conceptual knowledge has to do with the relations among facts (Snir et al., 1993). As Chi and Ceci state,

> children's knowledge increases as they acquire more links among the concepts and attributes that they already have… That is, with development, knowledge becomes more accessible because there are more links interconnecting the different components of knowledge (1987, p. 116)

Conceptual knowledge is coherent and organized, and can help students to learn new ideas by connecting these to their existing knowledge (Kilpatrick, Swafford, & Findell, 2001; Linn et al., 2004). The NRC book *How People Learn* (Bransford et al., 1999) summarizes the current constructivist view that experts have extensive factual knowledge organized in ways that show deep understanding.

Current constructivist conceptions of learning have also been influenced by Vygotsky's (1978, 1985) theories about the social nature of learning. Vygotsky (1978, 1985) proposed that learners have a "zone of proximal development " that contains what they can most readily learn. He defined this zone as the difference between what the learner can accomplish alone and what he can accomplish with help from a more able peer. If a learner's current level of knowledge is known, and if a likely trajectory of learning has been identified by previous research and organized into a learning progression, then we can tailor instruction to that learner's zone of proximal development.

Learning progressions are compatible with and embody constructivist views of learning. Learning progressions value prior knowledge by incorporating a lower anchor (Duschl et al., 2007) comprised of a description of the knowledge and reasoning that students have been found to have at the beginning of the period of interest. The upper anchor of a learning progression is composed of the knowledge that society expects of students at the end of a specific point in schooling (as embodied in standards or benchmarks, for instance) and that educational research shows is feasible (Smith et al., 2006; Duschl et al., 2007). Learning progressions express this upper anchor not in terms of facts to be learned, nor as decontextualized and abstract skills, but as learning performances that involve both factual knowledge and higher-order and/or scientific inquiry skills (Smith et al., 2006; Krajcik, McNeill, & Reiser, 2008; C. Anderson, 2008). Learning progressions also

> propose the *intermediate* understandings between these anchor points that are reasonably coherent networks of ideas and practices and that contribute to building a more mature understanding (Duschl et al, 2007, p. 220; italicized in the original).

These intermediate understandings are often described as successive *levels* (e.g., Alonzo & Steedle, 2008); the levels track students as they gradually learn more factual knowledge and structure this knowledge. Learning progressions can guide teachers and curriculum developers by providing information about a learner's likely zone of proximal development (Vygotsky, 1978, 1985), provided the learner's current level is known. Assessment is thus an important component of learning progressions, as it can determine a student's level and thus what instructional activities are likely to help the learner. Assessment also helps measure the effectiveness of instructional activities and a learner's advances along the learning progression.

Learning progressions also need to consider the role of instruction in scaffolding this progress. A learning progression needs to incorporate instructional experiences that have proven effective, or in their absence, suggested learning activities that we believe will help students construct their understanding and move along the learning progression. In

keeping with constructivist learning theory, the learning progression is understood not to be a single, unique pathway to learning for all students, since each learner will experience different learning activities, starting from different prior knowledge (Duschl et al., 2007). However, learning progressions will allow us to predict what students will be able to achieve given appropriate instructional opportunities. Finally, learning progressions are research based, since

> The strongest evidence for a suggested advance comes in the form of teaching experiments that demonstrate how students can move from one set of understandings to the next or longitudinal studies showing systematic progressions in students' understanding (Duschl et al., 2007, p. 222).

In this dissertation, my theoretical framework is a learning progression approach, as described above, which incorporates elements of constructivist learning theory and values the connections among pieces of knowledge that learners construct.

Ways of Thinking About Size and Scale.

I conducted a search of the literature to identify distinct ways in which learners can think about, work with, and process the one-dimensional size and scale of objects. The words we use to talk about size, such as "big" or "little", always imply comparison (Nelson & Benedict, 1974). The size of an object is established by comparing it to a standard, which functions as a scale, allowing us to measure a dimension (Gibson et al., 2000). The standard may be normative (an implicit, stored mental standard); perceptual (a physically present object); or functional (related to the function it is to serve - Gelman & Ebeling, 1989). Size can also be defined by comparison to conventionally defined units of measurement (e.g., Lehrer, 2003). In their seminal paper on learning progressions, Smith and colleagues (2006) mention four ways of thinking about variables in general that are relevant to one-dimensional size: ordering, classifying, and quantitative measurement using standard or non-standard units. I term these ways of thinking about size and scale "aspects", and refer to them as ordering, grouping, relative scale, and absolute size, respectively. I use "grouping" instead of "classifying" because of the latter term's connotation of using pre-existing categories; I am interested in how students define the

categories themselves, and the term grouping captures this broader meaning. I use the term "relative scale" to refer to thinking about the size of one object in terms of another object, as is done when measuring with non-standard units. Relative scale also includes thinking about scaled-up or scaled-down objects; it follows the terminology in recent science education papers (Tretter, Jones, & Minogue, 2006; Batt et al., 2008). I use the term "absolute size" to refer to thinking about the size of an object in terms of standard units, such as the millimeter or the mile, instead of the term measurement using standard units, because absolute size can be determined in other ways in addition to measurement, and in order to be consistent with the terminology of recent science education papers. Recent research findings about students' knowledge of size and scale fit well into these four aspects of size and scale.

Early Development of the Four Aspects

Resnick proposed a sequential development of four qualitatively different kinds of mathematical thinking (Resnick, 1992). In this account, the first kind is the mathematics of *protoquantities.* This kind of thinking involves reasoning about amounts of physical material via "actions that can be performed directly on physical objects" (p. 403), including combining and ordering. These are qualitative and correspond to ordering and grouping. Their next kind of mathematical thinking is the mathematics of *quantities*, involving measured physical material. These are quantitative and can be expressed in linguistic terms like *n* inches (i.e., absolute size) and operations such as dividing a measured amount into equal shares (corresponding to relative scale). Two further types of mathematical thinking no longer relate to physical materials and are thus not directly relevant to size and scale: the mathematics of numbers and the mathematics of operators (Resnick, 1992).

Piaget and Inhelder (1971) proposed a similar developmental pathway in constructing the idea of space, focusing on part-whole relations but applicable in my opinion to any set of objects (two objects of different lengths can be conceptualized as a whole length and a partial length, or both can be visualized as a part of the length of a third, longer object). This pathway starts with children who do not understand the relationship between parts

and whole. Then they acquire qualitative "sub-logical operations" that only show that the whole is greater than the parts. The next stage of understanding involves non-metric "extensive" quantity, allowing for comparison between parts; this enables ordering and grouping. The pathway culminates in the understanding of "metric" quantities, which involve the idea of how many times larger or smaller one part is than another part or a whole, and with sizes expressed in terms of units, i.e., relative scale and absolute size.

Even though the qualitative aspects developer earlier than the quantitative ones, I expect all of these to be in place by middle school, to some degree. For instance, even though middle school students may have trouble measuring an object (e.g., Lindquist and Kouba, 1989; Lubienski, 2003), they "have a general understanding of the attributes of length and appropriate units for measuring length" (Strutchens, Martin, & Kenney, 2003); and the *Benchmarks* (AAAS, 1993, 2009) include the expectation that fifth-graders understand the fundamentals of measurement with standard and non-standard units.

Factual Knowledge of Size and Scale

Next, I further define the four aspects and review what is known about students' factual knowledge in each aspect, focusing on submacroscopic objects.

Ordering

An example of ordering is ant < beetle < mouse < cat < cow, where " < " indicates "smaller than". Ordering or seriation "is the product of a set of asymmetrical transitive relations connected in series" (Inhelder & Piaget, 1969, pp. 5-6), and is qualitative. This means that ordering is composed of successive comparisons of pairs of objects, establishing for instance that A is smaller than B, and (separately), that B is smaller than C. By transitivity, A is also smaller than C: If A < B and B < C, then A < C. Through this serial process, the order A < B < C is established. As I will show in Chapter 6, however, the comparison of two objects can be done in various ways, including direct comparison, comparing the absolute sizes of both, comparing the relative scale of each in relation to a third object, and so on. Science and math standards documents include the expectation that early elementary students be able to order everyday objects by length, or numbers

and everyday fractions by magnitude (AAAS, 1993; NCTM, 1989).

Studies focusing on students' understanding of diverse objects in science often mention as one of their findings lack of accurate understanding of the size of the objects. These studies often refer to qualitative mistakes in student beliefs about the size of objects, for instance mentioning that students believe that the size of cells is similar to the size of atoms and molecules (Flores, 2003) or that students confuse cells and atoms (e.g., Harrison and Treagust, 1996). Only 15% of 11-13 year olds were able to correctly order germ, molecule, and atom by size (Waldron et al., 2006); this result informs the lower anchor. However, no data are provided for high school students that might inform the upper anchor. In another study, only 7% of 495 respondents aged seven-90 years old were able to correctly order atom, water molecule, bacterium, and cell; 45% of the respondents ordered the atom smallest (Castellini et al., 2007). Data are not reported by age group, so this study too only informs the lower anchor. The 5th, 7th, and 9th grade students in Tretter's study (Tretter, Jones, Andre, et al., 2006) collectively generated a mean ranking for hair width that was smaller than all or all but one of five submacroscopic objects: atom, atomic nucleus, cell, bacterium, and virus. This may mean that establishing the hair as a landmark at the boundary of the macroscopic and submacroscopic worlds may be important in the developments of students' understanding of size and scale. That study also found that students in the elementary and middle school grades perceive vertical distances as larger than horizontal distances of the same magnitude; these students on average ordered vertical distances (e.g., the height of the tallest building) as being much larger than they in fact were. In sum, we cannot expect middle school students to order submacroscopic objects correctly, and can expect some problems even with familiar macroscopic objects.

Determining the smallest object or measurement unit of which one knows involves ordering as well (A < all other known objects). Castellini and colleagues (2007) found that around 15% of students in grades six through eight mentioned a small visible object when asked for the smallest thing of which they could think, with 57% atom and around 24% microscopic objects. Waldron and colleagues (2006) found over 40% of 11-13 year

21

old students responded with a macroscopic object. This shows that lower anchor cannot assume knowledge of submacroscopic objects. Over 70% of high school-age students in both studies responded with submacroscopic objects, including over 40% with atom or other nanoscale objects (Castellini et al., 2007; Waldron et al., 2006); the upper anchor can thus include the expectation that all students at the end of high school know of objects as small as atoms, with improved curriculum. Waldron and colleagues report that "most" 11-13 year old students were unable to accurately order millimeter, micrometer, and nanometer by size (2006, p. 573); 57% of students of all ages could not accurately order the units (Batt et al., 2008). Another study showed that only 18% of 7-9th grade girls were able to define the millimeter as 1/1000 of a meter, and few had knowledge of units smaller than millimeter (Jones et al., 2007). This shows that the lower anchor cannot assume knowledge of submacroscopic units, or even of the millimeter.

Grouping

Grouping involves placing objects of similar size into a group, and objects of different size into different groups (Inhelder & Piaget, 1969). An example of grouping by size is {carbon atom, water molecule} < {skin cell, red blood cell} < {flea, ant, grain of salt} < {cat, dog}; like ordering, grouping is qualitative. Transitivity plays a role in grouping: if object B is larger than A but smaller than C, then B must be placed in the same group as A and C (because if A < B and B < C, then A < B < C). Science and math standards also include the expectation for young students to be able to group by attributes (AAAS, 1993; NCTM, 1989). Grouping objects by a continuous characteristic such as size is more complex than grouping by a distinction of kind such as animal species. This is because any two unequal numbers resemble or are dissimilar to each other only in relation to other numbers, for example, 5 is similar to 6 when compared to 100, but dissimilar to 6 when compared to 5.01 (see Sera & Smith, 1987, for a similar discussion on the terms big and little). Furthermore, there may be several satisfactory ways to group objects by size, using different numbers of groups. As I will show in Chapter 6, there are many possible strategies for grouping, some of which involve absolute size or relative scale and some of which are more qualitative.

Expert scientists conceptually group objects into "worlds" characterized by measurement units and the instruments used to study them (Tretter, Jones, & Minogue, 2006). It is likely that each world is characterized by objects that are exemplars of a size range – these have variously been called landmarks (Tretter, Jones, Andre, et al., 2006), reference points (Joram, Subrahmanyam, & Gelman, 1998), benchmarks (Joram, 2003) or anchors (Petrov & Anderson, 2005). Landmark submacroscopic objects for experts and gifted seniors include the atom and "microscopic objects", respectively (Tretter, Jones, Andre, et al., 2006). That study also reports conceptual boundaries for size categories. All age groups (fifth grade through graduate students) aggregately conceptualized macroscopic objects ranging in size from human to astronomical distances as falling into three categories: room size, field size, and larger. Middle school students collectively had a single category for objects smaller than the human, while gifted high school seniors collectively differentiated these into small macroscopic, very small macroscopic, and microscopic. The authors' conclude that experts appear to make more groups than novices but note that the aggregate data may obscure individual differences. The literature thus provides little guidance regarding what to include for grouping in the lower and upper anchors.

Inhelder and Piaget found that "the development of seriation [ordering] is almost exactly parallel to that of classification [grouping], and tends to precede it step by step." (1969, p. 4). The fact that ordering precedes grouping developmentally does not mean that ordering is easier than grouping. Correctly grouping by size involves looser constraints than correctly ordering by size, as grouping allows for ties in size – there is no distinction in size between the objects in a given group. Grouping accurately is easier than ordering accurately, for a given set of objects. Consider the three smallest objects in a series of four or more. If those three objects are all of a different size, then there are six ways to order them, only one of which is right. However, assuming all three are grouped together, then there is only one way to group, and it is automatically correct.

Relative Scale

Relative scale is a third way of thinking about size without reference to conventionally

23

defined units. Relative scale expresses the size of one object in terms of another, for instance, the thickness of DNA is around 20 times wider than a carbon atom. Relative scale is quantitative, unlike ordering and grouping. Relative scale can be obtained through measurement with non-standard units (Smith et al., 2006) using "iteration" – placing a unit end to end with no gaps or overlaps to cover a length (Wiedtke, 1990; Lehrer, 2003; Clements & Stephan, 2004). As I show in Chapter 6, relative scale can be obtained in other manners; for instance, it can be calculated if the absolute sizes of both objects are known. Scales such as are used in a map, architectural blueprints or mock-ups, and toy cars also involve relative scale. Vergnaud (1988) reported that not all students at the end of elementary school understand expressions like "three times more", or "three times less", often not realizing that these are of a multiplicative nature (p. 156). Thus, we cannot take for granted that students will understand the very idea of relative scale.

Relative scale should be more difficult than ordering because relative scale involves ordering (e.g., object A is bigger than object B) but also a quantification of the relative difference in size (e.g., object A is 12 times bigger than object B).

Relative scale is employed by experts, who "unitize" (Lamon, 1994), expressing the size of one object in terms of a landmark object (Tretter, Jones, Andre, et al., 2006). For instance, scientists have defined units such as the astronomical unit and light year to facilitate dealing with enormous distances of the solar system and beyond. Studies focusing on students' understanding of diverse objects in science may include mention of what are essentially relative scale misunderstandings. The finding that students do not have a good idea of the size of the atom (often overestimating its size) is common in the literature (e.g., Brook et al., 1984; Griffiths & Preston, 1992). Previous research shows that students often overestimate the size of the Earth relative to the size of the sun (e.g., Phillips, 1991; Hapkiewicz, 1999).

Tretter, Jones, and Minogue (2006) investigated the accuracy of learners' estimates of the size of multiple objects, in terms of body lengths. Respondents were asked to provide an

object of given size ranges. Even the gifted high school seniors had only a 20% accuracy rate for objects at the micro- and nanoscale (acceptable accuracy meaning within a factor of 10). They tended to provide objects that were too large. Accuracy at the millimeter range was much higher, at 80% for gifted seniors and around 45% for middle school students. Batt and colleagues (2008) asked learners to estimate the size of small, macroscopic objects if they were scaled up 100 million times, after providing the learners with an example. Respondents were told the size of a golf ball scaled up 100 million times and asked to estimate the size of a pinhead at that scale, or vice versa. Accuracy was below 50% for most groups. Thus, both papers found that the relative scale tasks were very difficult for learners.

At the lower anchor, then, students cannot expected to be able to estimate the relative scale of objects at the millimeter range or below. At the upper anchor, these findings show that high school seniors can estimate or calculate relative scale for millimeter-sized objects, under their existing curricula. In Chapter 5 I report on middle school students' accuracy of relative scale estimates after a focused curriculum on size and scale.

Absolute Size

In the sense that size is determined by comparison to a scale, size is always relative; however, magnitudes that are established in relation to conventionally defined units, for instance 5 mm, have come to be called "absolute" (e.g., Graham, Ernhart, Craft, & Berman, 1964). The updated *Benchmarks for Science Literacy* (AAAS, 2009) include the expectation that fifth-graders should know that a quantity consists of both a number and a unit (9A/E3, Ch. 9).

In almost all the countries in the world, the SI system – based on the earlier metric system – is in customary use; this system is used almost universally in science. The SI unit for length is the meter, and standard prefixes (such as milli- for one thousandth) can be used in combination with the base unit. In the United States, however, the Imperial system – based on the earlier English system – is in customary use, with units for length such as the inch, the foot, and the mile. (Following customary usage, I henceforth refer to these

two systems as metric and English.)

Both absolute size and relative scale can be determined through measurement. Lehrer (2003) proposed that conceptual understanding about measurement is constituted by a "network or web of ideas related to unit" (p. 180) involving eight foundational ideas:

(1) Unit-attribute relations: the appropriate units must be used to measure, e.g., units of length to measure length.
(2) Iteration: repeatedly placing the unit end to end from starting point to end point.
(3) Tiling: leaving no gaps between units when iterating
(4) Identical units: the same unit is used for a measure, or mixtures are explicitly labeled (e.g., 5 feet and 10 inches)
(5) Standardization: the use of conventional units (e.g., metric or English units) facilitates communication.
(6) Proportionality: the number of units to represent a measure is inversely proportional to the size of the units.
(7) Additivity: the total distance is equivalent to the sum of the parts.
(8) Origin: any location can serve as an origin.

In Lehrer's (2003) account, students gradually coordinate these eight foundational ideas in constructing a theory of measurement.

Wiedtke's (1990) foundational ideas for measurement additionally include the ideas that counting the units iterated defines the length of a segment, that a segment can be assigned a length of one, and transitivity. Clements and Stephan (Clements & Stephan, 2004; Stephan & Clements, 2003) additionally propose *partitioning* – the idea that an object can be divided into units, and *conservation* – the idea that displacing an object does not change its length (conservation of length was described and studied earlier by Piaget and colleagues: Piaget & Inhelder, 1971; Piaget, Inhelder, & Szeminska, 1960).

Absolute size can be obtained through means other than measurement. If the absolute size

of one object is known, and the relative scale of that object to a second object is also known, then the absolute size can be calculated. Joram and colleagues (1998) review various strategies for the estimation of absolute size, which rely either on relative scale or mentally visualizing a measurement tool next to the object (see Chapter 6).

Previous research has shown that learners of all ages find it difficult to come up with objects of given absolute sizes, with lower accuracy at the extremes of size and particularly at the micrometer range or below (Tretter, Jones, & Minogue, 2006). Jones and colleagues (2008) found a similar pattern among novice and experienced teachers. Jones and colleagues (in press) also found a similar pattern but higher accuracy among visually impaired students. Gifted high school seniors had accuracy (within 10X) near 100% for objects between 1 mm and 100 m, but only 40% at the micrometer range, and 20% at the nanometer range; middle school students had 100% accuracy at 1 m but only 80% accuracy at the millimeter range. The lower anchor can thus include the expectation that students can only estimate the size of objects that are near human size, with millimeter-sized objects also at the upper anchor. However, it remains to be seen what accuracy of absolute size estimation can be achieved with a focused curriculum (see Chapter 5).

Absolute size appears to be more difficult than relative scale in perceptually based tasks, but not for tasks that are not based on perception. Vasilyeva and Huttenlocher (2004) examined 4- and 5-year old children's ability to interpret two-dimensional maps in order to determine their ability to scale (however, the map was a long, thin rectangle, thus involving essentially one dimension). The findings show that small children who are not familiar with units and cannot yet use formal proportional reasoning still have early, non-quantitative, perceptually based scaling abilities, whereby they appear to "mentally transform…a layout in a way that preserves metric relations." (p. 688). Graham and colleagues (1964) also found that "it is easier for children between 2 and 4 ½ years old to learn to choose a stimulus of the same relative size than to choose one of the same absolute size." (p. 32). With tasks that do not depend on perception, however, it is not clear whether relative scale or absolute size is easier for learners. Tretter, Jones, and

Minogue (2006) found that the accuracy of respondents coming up with objects of a certain size range expressed in metric units (absolute size) or body lengths (relative scale) did not vary predictably.

Absolute size is more difficult than ordering. Research studying teachers' conceptions of geological time found that "relative" judgments (corresponding to what I term ordering) were easier and more accurate than absolute judgments (Trend, 2001; Dahl, Anderson, & Libarkin, 2005). Tretter, Jones, Andre and colleagues (2006) report a similar finding for the size of objects.

Summary

In sum, I study four aspects of size and scale. Ordering and grouping involve the qualitative comparison of one object to other objects. Relative scale involves a quantitative comparison of one object to another object. These three aspects of size and scale depend on comparison to objects, and thus are relative in nature. Absolute size, in contrast, depends on comparison to conventionally defined units of measurement, and thus is not relative. Both absolute size and relative scale are quantitative, and can be obtained through measurement with standard and non-standard units, respectively. Absolute size depends on both a number and a unit for its characterization. The four aspects are summarized in Figure 1.

The recent body of research into learners' understanding of the submacroscopic world helps characterize the lower anchor, considering middle school as a starting point. Students may not know of any objects that are too small to see (Castellini et al., 2007; Waldron et al., 2006), and cannot accurately order submacroscopic objects by size (Castellini et al., 2007; Waldron et al., 2006). They do not know of units smaller than a millimeter and may not have a good grasp of the millimeter (Waldron et al., 2006; Batt et al., 2008; Jones et al., 2007). They can estimate the relative scale of objects (or provide

	Qualitative	Quantitative	
Harder	Ordering by Size^	Relative Scale^	Similar difficulty*
Easier	Grouping by Size	Absolute Size^	

^Ordering is easier than absolute size or relative scale
*Relative scale is easier with perceptible objects

Legend: ▨ = Relative   ☐ = Absolute

Figure 1: A taxonomy for aspects of size and scale

objects of a given relative scale) near human size, but not at the millimeter size range or below (Tretter, Jones, & Minogue, 2006). They can estimate the absolute size of objects near human size, but not at the millimeter range or below (Tretter, Jones, & Minogue, 2006; Jones et al., 2008; Jones et al., in press).

The research also helps characterize the upper anchor, considering the end of high school as an endpoint. Students are able to think of nanometer or sub nanometer-scale objects like molecules, atoms, or subatomic particles as their smallest known objects (Castellini et al., 2007; Waldron et al., 2006). Students can estimate the relative scale and absolute size of objects from human size down to millimeter-sized objects (Tretter, Jones, & Minogue, 2006).

Additional Research Required for Factual Knowledge

Individual learners' grouping strategies and levels of accuracy have yet to be characterized adequately. For ordering (including the smallest object and unit known), as well as relative scale and absolute size, it remains to be seen what students can achieve with improved instructional opportunities. It is also necessary to investigate how students' factual knowledge tends to grow, and what instructional activities are effective. There is a need for improved assessment that goes beyond calculating mean performance

levels at different age or grade levels and allows examination of individual thinking. The reliance on paper and pencil surveys and instruments in most prior research may not be allowing unexpected findings to emerge (Ambert, Adler, Adler, & Detzner, 1995), and may be ineffective in gauging what students can achieve with clarification or additional prompting.

Conceptual Understanding of Size and Scale

From a learning progression stance, the understanding of size and scale includes but necessarily goes beyond knowing the size of objects as isolated facts. Previous research on size and scale and related concepts has taken some steps towards characterizing the organization of students' knowledge of size and scale, in two distinct ways. One approach examines how students use the known sizes of landmark objects in order to get a sense of size, to estimate the size of other objects, and to develop conceptual size categories. Another approach examines how students' knowledge of one type of size and scale information is related to another type of size and scale information. In this dissertation, I build upon and integrate both approaches in constructing a learning progression for size and scale. I discuss these two approaches next.

The Use of Landmark Objects and Conceptual Boundaries

Tretter and colleagues investigated conceptual size categories, which they propose are a learner's way of organizing knowledge about those objects (Tretter, Jones, Andre, et al., 2006; Tretter, Jones, & Minogue, 2006). Each category is characterized by exemplars or landmark objects, and experts' categories also may involve an associated unit of measurement and tool. By chunking knowledge in their area of expertise, experts develop scale "worlds" that they can access. Experts may also unitize (Lamon, 1994): in other words, they may define an object as a unit that is useful in a particular world. In this manner, a learner can develop a sense of size. This means that experts can use and relate grouping, the absolute size of landmark objects, and relative scale (by unitizing). However, this body of research has not directly investigated individual learners' ability to connect grouping, absolute size, and relative scale.

30

A very recent paper by this research group (Jones & Taylor, in press) sketches a trajectory for the learning of scale. In this trajectory, novices develop number sense and measurement estimation skills, conceptualize relative sizes, and learn to use measurement tools. Developing learners convert measurements and scales, consider surface area to volume relationships, become aware of changing scales, use body rulers for measurement and estimation, visualize scales and understand different types of scales, and develop proportional reasoning and visual spatial skills. Experienced learners display automaticity and accuracy, create reliable scales, relate one scale to another, develop accuracy in using scale, and apply conceptual anchors when estimating scale. This trajectory provides a large-grained description of how the experts arrived at their state of knowledge concerning a broad set of skills and concepts; this dissertation takes a finer-grained approach to examining the development of understanding of one-dimensional size and scale. The trajectory proposed by Jones and Taylor (in press) relies in part on retrospective self-reports from the experts, but experts may have forgotten what they struggled with as learners (Bransford et al., 1999 – see also Hook & Rosenshine, 1979, for additional cautions concerning retrospective self-reports). I provide a specific illustration of how a learner did not recall exactly how she arrived at her advanced understanding in Chapter 4. Thus, the trajectory proposed by Jones and Taylor (in press) is useful in coordinating the learning progression generated here to other, related areas of size and scale (e.g., surface area to volume relationships). It is also useful in pointing out the importance of diverse estimation techniques and the creation of scales.

In this dissertation, I closely examine how learners actually learn and apply these skills, for one-dimensional size and scale. I also examine how students develop landmark objects to help define and create size categories and to provide a "stable frame of reference on which to base [a] relational web of scale sizes" (Tretter, Jones, Andre, et al., 2006, p. 307).

Landmark objects are also important in the model proposed by Joram and colleagues (1998) for thinking about length, based on observing measurement estimation (estimation of absolute sizes). This model is thus most pertinent for objects with which the learner

has directly interacted but may still be informative for submacroscopic objects, particularly as some of the estimation tasks deal with objects that are not physically present. In this model, a learner builds a set of reference points or benchmarks, such as the height of a human being 6 feet or the thumb being 2 inches long:

> Benchmarks typically consist of non-standard units whose lengths are used to represent the lengths of standard units (e.g., the length of the top half of my thumb equals about 1 inch in length) or multiples of units (e.g., the height of an average man is about 6 feet)…[benchmarks] typically evolve into mentally represented objects… measurement benchmarks allow one to generate estimates about unfamiliar quantities. (Joram, 2003, p. 58)

These reference points are more useful than using a standard unit of measurement because they may be more familiar to the learner or may require fewer iterations (e.g., estimating the height of a 12-ft wall requires only two iterations of a 6-ft human, but 12 iterations of a foot). In this model, learner constructs a *mental measurement line* with well-known reference points at the corresponding positions. The learner can estimate intermediate sizes by mentally iterating reference points, which Joram and colleagues call "knowledge of how a scale is constructed." (1998, p. 427). Joram and colleagues caution that children may not go beyond knowing isolated reference points to build a measurement line if not pushed to do so by their teacher. This instruction would have to focus on the "principles and constraints that underlie the construction and use of measurement scales for estimation" (p. 431). In my terminology, this involves knowing how to use relative scale.

Summary of Prior Research

By using landmark objects of known absolute size and knowledge of scale creation, students develop a mental measurement line (Joram et al., 1998). These landmarks are exemplars of size ranges that may also include an associated unit and tool, and can be chunked into "worlds"; these size ranges are conceptual categories to help students develop a sense of size (Tretter, Jones, & Minogue, 2006). At the lower anchor, students may possess no submacroscopic landmarks; at the upper anchor, students may have a rather vague landmark of "microscopic objects" (Tretter, Jones, Andre, et al., 2006) in the

absence of a focused curriculum on size and scale. In the lower anchor, students may not be able to construct a scale to understand the size of objects between landmarks (Tretter, Jones, Andre, et al., 2006; Joram et al., 1998).

Additional Research Required for Conceptual Understanding

In building upon this work it is important to determine more specific promising landmark objects for submacroscopic regions, beyond the ones proposed (atom for experts, and microscopic objects for gifted seniors -Tretter, Jones, Andre, et al., 2006), and to investigate whether and how students appropriate and learn to use these landmark objects. Scale creation, mentioned by both research groups, also needs to be more clearly defined. Given the foundational ideas in measurement described earlier (Wiedtke, 1990; Lehrer, 2003; Clements & Stephan, 2004), I believe that understanding relative scale is at the heart of scale creation, since relative scale is the essence of iteration. By coordinating the absolute size of landmark objects and the relative scale involved in the iteration of these landmark objects, a mental measurement line (Joram, 1998) or relational web (Tretter, Jones, Andre, et al., 2006) can be created. Thus, characterizing the way in which students gradually learn to coordinate absolute size and relative scale is of fundamental importance to a learning progression for size and scale.

Relationship Across Types of Knowledge of Size and Scale

The section above focusing on landmark objects implied but did not explicitly examine the importance of relating or linking across aspects or types of knowledge. A recent paper (Batt et al., 2008) defines a "think score" for size and scale based on the smallest object of which a respondent could think, and relates the think score to performance on the relative scale task mentioned earlier (scaling up a pinhead or golf ball 100 million times, given an example). By relating knowledge of smallest object (ordering) to performance on a scaling task (relative scale), this study attempts to characterize networks of ideas. However, a trend showing increased accuracy on the pinhead estimation for higher think scores was found, but not for the golf ball, so the usefulness of the findings to a learning progression are limited. Respondents were not probed about the size of the smallest object (C. Batt, personal communication, April 13, 2009). I have observed that respondents' answers depend on the wording of the probe: many students reply with a

small macroscopic object (e.g., grain of salt) when asked for the smallest thing of which they can think, but recall cells, atoms or electrons when probed for an object "too small to see" (see Chapter 4). Thus, this think score is based on what we have found to be a measure that is highly sensitive to context. It may be that levels of thinking about size and scale could be more comprehensive in nature, taking into account performance across various aspects of size and scale, so that they are truly coherent networks of ideas. As I show in Chapter 4, the measure I develop to characterize students' level of thinking about size and scale is more highly correlated to relative scale, absolute size, and ordering tasks than the smallest object known (even though respondents were also probed for an object too small to see in my interviews).

Summary of Prior Research

Batt and colleagues (2008) developed a measure to reflect the relationship between ordering and relative scale. This is a step towards characterizing conceptual understanding. Additional research might more fully illuminate how to characterize conceptual understanding and what should be included in the lower and upper anchors of the learning progression.

Additional Research Required for Conceptual Understanding

Given the importance of connections among pieces of knowledge in current learning theory and the need to identify coherent networks of ideas in order to define levels for the learning progression, it is essential to develop measures of connectedness among all four aspects of size and scale. In the following section, I briefly explain my approach to measuring connectedness across aspects of size and scale and review pertinent literature.

Connections Across Aspects of Size and Scale

The four aspects of size and scale are logically connected. Consider the diameters of the balls used in several sports: squash (4 cm), tennis (6.5 cm), baseball (7.5 cm), volleyball (21 cm), and soccer (22 cm). Knowing the absolute size of the sports balls informs their ordering by size: squash < tennis < baseball < volleyball < soccer; it also allows one to calculate the size of one ball relative to another (e.g., the soccer ball is 22 cm / 7.5 cm = 2.9 times bigger in diameter than the baseball). Conversely, knowing that the diameter of

the soccer ball is 5.5 times bigger than that of the squash ball, along with the diameter of one ball, allows one to calculate the absolute size of the other. The balls can also be organized into groups, such as {squash} < {tennis, baseball} < {volleyball, soccer}, based on their relative or absolute sizes. I consider that the connections among the different aspects of size and scale are not only a logical necessity, but also an important component of conceptual knowledge of these ideas.

Since I cannot directly observe students' mental connections, I infer them by observing the *consistency* of individual students in their answers to tasks designed to make their thinking visible (Pellegrino et al., 2001); my tasks involve the same objects across different aspects of size and scale. Thus, I can measure consistency independently of accuracy of factual knowledge (Vosniadou, 2003). For instance, some students believe that relative scale and absolute size are not related, and do not realize that the relative scale of two objects (how many times bigger one object is than another) and the absolute size of one object uniquely determine the size of the second object. Such students tend to have inconsistent answers for the same object on relative scale and absolute size tasks.

Consistency Across Aspects of Size and Scale

While there is little research on the connections among the aspects of size and scale (see above), research from various fields suggest that these connections may be non-trivial for students.

Consistency Across Ordering and Grouping

 A learner who orders five objects A < B < C < D < E, but groups them {A, C} < {B, D, E}, is not being consistent, as B < C in the order task, but C < B in the grouping. This can be interpreted as showing that she has not connected ordering and grouping. The literature on decision making has found that even adults occasionally display "intransitivity" of preferences (e.g., Tversky, 1969) when a respondent prefers A to B, and prefers B to C, but then prefers C to A (the transitive choice is to prefer A to C). Intransitivity may also occur in linked ordering and grouping tasks.

Consistency Across Ordering and Relative Scale

For three objects ordered A < B < C, if B is 10 times smaller than C, then A must be smaller than C by *more than 10* times. Consistency across ordering and relative scale involves the inverse relationship between unit size and unit number, or what Lehrer (2003) calls the idea of proportionality: the smaller the unit, the more units are required to cover a given distance. More A's than B's are required to compose C, since object A is smaller than object B. Hiebert found that unlike most measurement-related tasks,

> Tasks that involve the inverse relationship between unit number and unit size… apparently depend on more general logical reasoning abilities [such as] conservation and transitivity… (Hiebert, 1981, p. 207).

Hiebert found that while most basic concepts and skills of linear measurement could be taught to students who did not display conservation of length or transitivity, the inverse relationship between unit number and unit size could not. In his words, the "constraints imposed by the absence of these reasoning abilities are not removed by specific instruction" (1981, p. 207). Thus, students who lack conservation or transitivity may have trouble being consistent across ordering and relative scale. A specific instance of lack of consistency across ordering and relative scale was noted in word problems about the amount of food needed by three fish of different sizes. Some learners initially assigned the same amount of food to each of the two larger fish, instead of more food to the largest fish (Clark & Kamii, 1996). This was characterized as an "absence of serial correspondence" (recall that seriation is equivalent to ordering). Thus, there is reason to suspect that learners may have trouble making this connection.

Consistency Across Ordering and Absolute Size

For two objects ordered A < B, the absolute size estimated for A must be smaller than the absolute size estimated for B. Learners may not know convenient measurement units for submacroscopic objects, and this may force them to rely on fractions or decimals of familiar units (e.g., 0.01 mm, 1/200 in). However, these rational numbers are difficult for students (e.g., Post, Cramer, Behr, Lesh, & Harel, 1993; Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1980; Hiebert & Wearne, 1986). Psychological research has shown

that even college students may incorrectly interpret decimals such as 0.02 to mean "1 in 200" instead of 2 in 100, possibly as an overgeneralization from 0.01 meaning 1 in 100 (Cohen, Ferrell, & Johnson, 2002). This misinterpretation of the meaning of decimals may lead some students to assign larger sizes to smaller objects, resulting in a lack of consistency between ordering and absolute size. Even with familiar objects and units, students may be inconsistent, for instance estimating the width of a car as larger than the width of a road (Markovits, Hershkowitz, & Bruckheimer, 1989). It also highlights the role of mathematical skills that underlie students' knowledge of size and scale and helps them to be consistent (this is examined in Chapter 6).

Consistency Across Absolute Size and Relative Scale

If object A is estimated to be 100 times smaller than object B (relative scale), and the absolute size of object B is given (say, 1 mm in length), then an estimate for the absolute size of A should also be 100 times smaller (e.g., 1/100 mm or 0.01 mm). This connection requires proportional thinking – a pivotal concept that "is the capstone of children's elementary school arithmetic [and] the cornerstone of all that is to follow." (Lesh, Post, & Behr, 1988, pp. 93-94). However, a large body of research has shown proportional thinking to be difficult for students (e.g., Lesh et al., 1988).

Consistency Across Grouping and Absolute Size or Relative Scale

The size differences between objects can be used to group objects, often in more than one way. For instance, objects that measure 1cm, 2 cm, 5 cm, 7 cm, 30 cm, and 40 cm in length, could be sensibly grouped as {1 cm, 2 cm}, {5 cm, 7 cm}, {30 cm, 40 cm}, following a rule that places objects that differ by a (multiplicative) factor of more than 2 in different groups. Alternate grouping strategies can be made that are *not* consistent with absolute size, e.g., {1 cm, 5 cm}, {2 cm, 7 cm}, {30 cm, 40 cm}. The same holds true for relative scale. Inhelder and Piaget noticed that many young children spontaneously switch criteria when asked to classify (1969, p. 285). Some students may thus have trouble being consistent across ordering and relative scale, or ordering and absolute size.

Most of the types of consistency mentioned above can be measured independently of the accuracy of the factual knowledge about the size of objects. For instance, if a student

mistakenly believes that a virus is a small multicellular organism (rather than being smaller than a cell or bacterium, as is the case), then she might order objects as follows: atom < carbon dioxide molecule < Staph A bacterium < red blood cell < virus < dust mite. Even though the order is wrong, the student's grouping can still be evaluated for consistency with ordering (see Vosniadou, 2003 for a similar discussion). If the student groups the objects as {atom, carbon dioxide molecule}, {Staph A, blood cell}, {virus, dust mite}, the ordering is consistent with the grouping, even though the grouping is also wrong. Students often do not make connections that appear obvious to an adult; they may treat information as separate pieces (Hiebert & Carpenter, 1992; diSessa, 1988). We cannot take for granted that learners will have well-connected knowledge of size and scale at the outset. In constructing a learning progression for size and scale, I thus focus both on the accuracy of individual learners' knowledge about the size of objects and their consistency across aspects of size and scale, and organize my findings not only by aggregate mean performance by year in school, but by patterns of individual thinking.

In sum, there is little science education research about the connectedness of students' knowledge across aspects of size and scale. However, research from other areas including decision-making, mathematics education, and psychology suggests that students may find it difficult to make accurate and productive connections.

A systematic exploration of students' consistency across aspects of size and scale is still needed. As I suggested above and describe in Chapter 3, I undertake this exploration asking students to carry out tasks for different aspects of size and scale using the same objects.

Chapter Summary

In this chapter I described learning progressions and showed how they are compatible with contemporary, constructivist views of learning. Learning progressions acknowledge the importance of prior knowledge by incorporating a lower anchor describing what students actually know about a big idea in science at the beginning of the period of

interest. Learning progressions set learning goals in the form of an upper anchor informed by standards and learning research. Learning progressions also describe intermediate levels of understanding. Learning progressions include instructional activities and assessments. I explained why both conceptual and factual knowledge are important, and outlined how connections between pieces of factual knowledge are a component of conceptual understanding.

I next synthesized the research literature on size and scale, highlighting four aspects: ordering, grouping, relative scale, and absolute size; and tracing a pathway showing how these develop in early childhood according to Resnick (1992) and Piaget and Inhelder (1971). I presented definitions of the four aspects and discussed their relative difficulty, based on prior research or theoretical considerations. I reviewed a body of recent research that characterizes students' mean factual knowledge about the size of submacroscopic objects through ordering, grouping, relative scale, and absolute size by grade or age bands, usually showing that learners do not have a good command of size and scale. Recent research has proposed that learners gradually construct a mental measurement line or relational web of sizes. The relationship between the smallest object a learner knows and other size and scale tasks has been investigated, and a descriptive trajectory from novice through developing to experienced learner proposed.

I then summarized prior research into factual knowledge of the four aspects, which informed mainly the lower anchor, but also the upper anchor. The lower anchor at middle school can only assume that students can estimate the absolute size and relative scale of objects near human size; they may not have a good grasp of the millimeter and may be unaware that some objects are too small to see with the unaided eye. Conservatively (due to the absence of a focused curriculum on size and scale), the upper anchor can include knowing about objects as small as the atom or subatomic particles and being conversant with relative scale and absolute size estimations for objects as small as a millimeter. However, this is subject to modification based on the results given improved curriculum.

Not much is known about students' conceptual understanding of size and scale,

particularly whether and how they make connections across aspects of size and scale. Additional research is required to look into grouping, and into the landmarks students can construct for the submacroscopic region. Research is needed to determine the upper anchors for ordering, relative scale, and absolute size, and how students can connect these aspects to enable scale creation and other forms of conceptual understanding. It is also critical to see how learning of these connections actually occurs in individuals, how it is related to factual knowledge, and what learning activities are effective. These investigations will require improved assessment that allows examination of individual thinking and how individual students actually develop their understanding.

Next, I describe my methods for the two studies I conducted, in Chapter 3.

CHAPTER 3: METHODS

In the previous chapters I showed that additional research is required to establish the upper anchor for factual knowledge; to characterize conceptual knowledge by assessing consistency across aspects; to investigate the relationship between factual knowledge and consistency of knowledge; and to find levels and study how students progress from level to level. I described my iterative research design, involving a cross-sectional study (Study One) followed by a teaching experiment (Study Two), along with a theoretical task analysis. In this chapter, I describe my methods for these studies. I describe how I assessed consistency across aspects as well as factual knowledge, and my strategy for characterizing levels of understanding. I also describe how I investigated the relationship between factual knowledge and consistency of knowledge.

Participants

Study One

My collaborators and I interviewed 101 middle school through undergraduate students, using the interview protocol described below. The students came from a low-mid SES (50% free or reduced lunch), ethnically/racially diverse public school district in a small city (N = 65); a mid-high SES, mainly non-Hispanic White private school in a college town (N = 31); and a selective research university (N = 5); all in the Midwestern US. See Table 1.

Table 1: *Participants in Study One*

Grade	Public Middle & High School N = 65		Private Middle & High School N = 31		Research University N = 5		Total N = 101	
	Male	Female	Male	Female	Male	Female	Male	Female
6th	6	8	0	0			6	8
7th	6	9	4	5			10	14
8th	0	3	0	0			0	3
9th	7	4	5	5			12	9
10th	2	4	6	5			8	9
11th	7	9	1	0			8	9
Undergrad					2	3	2	3
Total	28	37	16	15	2	3	46	55

Around a quarter of the students (N = 24) were selected by virtue of their participation in the summer camp described in Study Two (only their pre-camp interviews are considered in Study One). For the remaining pre-college students, we used stratified purposeful sampling (Patton, 2002) in order to obtain results that are typical of their school and may shed light on differences by gender and ethnicity/race. The undergraduate participants were volunteers from among work-study or undergraduate research students in the school of education, that is, a convenience sample (Patton, 2002). I purposely included few undergraduates in order not to create trends that might not hold for the middle and high school students. The sample includes few eighth graders and no 12th graders due to logistical constraints. Some participants received a token of appreciation for their participation (refrigerator magnet, cookie, etc.) but none received payment or class credit.

Study Two

A collaborator and I interviewed 24 public middle school students before and after they experienced a focused 12-hour curriculum for size and scale in our summer nanoscience camp, using the interview protocol described below. (The pre-camp interviews of these students are included in Study One.) See Table 2. The students came from the low-mid SES, diverse public school mentioned above. These self-selected students may have had greater than average interest in science, but did not necessarily differ in achievement. We initially screened applicants via an essay, but ultimately accepted all wait-listed

applicants, after some students were unable to attend. Two of the campers declined to participate in the study. In this study I excluded five students who had participated in the camp one year earlier (where a shorter curriculum on size and scale was enacted), and students who did not attend all sessions of the size and scale curriculum, because the impact of the curriculum on these students would presumably be different than on students experiencing the full curriculum for the first time.

Table 2: *Participants in Study Two*

Grade	Public Middle School		
	Male	Female	Total
6th	6	8	14
7th	3	4	7
8th	0	3	3
Total	9	15	24

Data Collection

Development of the Interview Protocol

Selection of Content

In order to explore students' knowledge of the size and scale of important scientific objects, I created an interview protocol that asks open-ended questions with precise wording, following Patton's (2002) standardized, open-ended format (see Appendix A). This format is meant to predictably elicit responses about the same topics, while allowing unexpected findings to emerge (Ambert et al., 1995). This interview assessed each student's ability to order by size, group by size, estimate relative scale (size relative to a small macroscopic object), and estimate absolute size, using several scientifically important objects that are potential landmark objects (Tretter, Jones, Andre, et al., 2006). The interview thus tested student factual knowledge of size and scale via the four aspects described earlier. The same questions also allowed me to determine whether students were consistent across these aspects, because they used the same objects. Consistency can be measured independently of factual accuracy (Vosniadou, 2003).

I selected 10 objects to use in the tasks, five submacroscopic and five macroscopic. Four submacroscopic objects - atom, molecule, virus, and red blood cell - are included in K-12 science education standards documents (AAAS, 1993, and NRC, 1996). The fifth submacroscopic object - the mitochondrion – is not mentioned in the standards documents but is included in seven of eight high school biology textbooks examined in a recent study (Beyer, Delgado, Davis, & Krajcik, in press). I included the mitochondrion in order to represent the size range between virus and cell. Castellini and colleagues report that over half of the second through fourth graders mentioned ants, bugs, and germs as the smallest objects of which they knew (2007); however, I used the red blood cell rather then a germ because some dictionaries define germ as a microorganism *or virus*. I also included familiar macroscopic objects at the millimeter, centimeter, meter, kilometer, and thousands of kilometers sizes: pinhead, ant, human, mountain, and earth, respectively. I included these macroscopic objects in order to contextualize students' accuracy for submacroscopic objects; previous research has found that students' accuracy is lower for submacroscopic objects (Tretter, Jones, & Minogue, 2006). These ten objects span the range from sub-nanometer or atomic to planetary. I list their sizes in Table 3.

Table 3: *Absolute size of objects used in interview*

Object	Size	Object	Size
Atom (carbon)	~0.15 nm	Pinhead	~1 mm
Molecule (water)	~0.3 nm	Ant	~2-20 mm
Virus (adenovirus)	~70-130 nm	Human (average adult male)	~1.8 m
Mitochondrion	~1-10 μm	Mountain	~1-9 km
Red blood cell	~5-10 μm	Earth	~12,700 km

Design Process

I developed the interview protocol following an iterative, construct-centered design (CCD) process (Pellegrino et al., 2008). This approach builds on learning-goal-driven design (Krajcik et al., 2008) and evidence-centered design (Mislevy, Steinberg, Almond, Haertel, & Penuel, 2003), and is consistent with current thinking in instructional and assessment design (e.g., Wilson, 2005; Pellegrino et al., 2001). The steps in CCD are to

clearly define and "unpack" the construct; specify claims that describe what we wish students to be able to do with their knowledge of the construct; define what we will take as evidence that the student has met the claim (mastered the knowledge); and develop tasks that will produce the evidence (Pellegrino et al., 2008). I presented part of the unpacking for one-dimensional size in Chapter 2; and include another part in Chapter 6. In Table 4 I present an example of claim, evidence, and task; this task is one of the interview questions. See Appendix B for the claim, evidence, and task for the remaining interview tasks.

Table 4: *Claim, evidence, and task for relative scale*

Claim	The student is able to estimate the size of a range of objects in terms of a convenient and familiar reference object.
Evidence	Given a series of objects and a reference of known size, the student's work includes estimates of sizes that are accurate to within one order of magnitude.
Task	How many times bigger or smaller than the head of a pin (1 mm in diameter) do you think the following objects are: diameter of an atom, diameter of a red blood cell, height of an average adult human, diameter of the earth?

Due to the small research base on size and scale, and particularly due to the lack of studies on the connectedness of students' knowledge, I used an iterative process that involved gathering empirical data to further inform the unpacking and development of the interview protocol. Collaborators and I pilot-tested the interview on around two dozen middle and high school students. I found evidence of lack of consistency of knowledge in my pilot interviews. For example, one sixth grader estimated an absolute size for the diameter of a red blood cell that was *larger* than the given size of a pinhead, despite ranking the cell as smaller, displaying a lack of consistency across absolute size and ordering. Another student estimated the diameter of a red blood cell at one-tenth of a millimeter, despite previously having estimated that it was one million times smaller than the 1 mm head of a pin (relative scale-absolute size inconsistency). This finding reinforced my theoretically-driven decision to systematically assess consistency across aspects of size and scale. After pilot testing, I revised the interview protocol for clarity and completeness. For instance, many students did not recognize the type of pin

portrayed on the card, so we incorporated the demonstration of an actual straight pin.

Interview Questions

We first asked respondents for the smallest object of which they could think (related to ordering), and the units with which to express its size (related to absolute size). We asked these questions first in order to avoid influencing their answers, since the following tasks provide objects and units. If the student responded with a macroscopic object, we probed for an object "too small to see". In later interviews, we asked students who did not know a unit for this object for the smallest unit they did know.

We then asked students to order cards of the 10 objects mentioned above, each with an image and the name, by the actual size of the objects. This direct ordering task included various submacroscopic and microscopic objects and is different from previously reported ordering tasks. These included separate tasks for macroscopic and submacroscopic objects (Waldron et al., 2006), various submacroscopic objects and a single macroscopic object (Castellini et al., 2007), or asked respondents to place objects into provided size ranges, which could allow for ties in ranking (Tretter, Jones, Andre, et al., 2006). We showed respondents a straight pin (of the sort used in packing men's dress shirts), since pilot testing revealed that many students did not recognize the type of pin portrayed on the card. We also told students that the pinhead was "around 1 mm, or a little less than 1/16 of an inch". While the difference between two- and three-dimensionality between the cards and the pin may affect students' thinking (e.g., Barufaldi & Dietz, 1974), showing the pin resolved a more serious problem, that of students not knowing the size of the reference object for the tasks. When we saw students ordering in an unorthodox fashion, or trying to judge the size of the *pictures* (e.g., measuring images with their fingers), we reminded them to order by the size of the actual objects depicted; we would point out, for instance, that humans are not actually 3 inches tall, as shown on the card. We expected middle and high school students to be able to understand the difference between the images as objects and the images and symbols, as prior research has shown that very young children can do so (DeLoache, Peralta de Mendoza, & Anderson, 1999). We also asked students to explain how they had ordered

the submacroscopic cards.

Next, we asked students to group the cards they had just ordered by size, making as many groups as they thought made sense, and placing as many cards as they wished into a given group. We asked students to explain their grouping strategy, by asking what the cards in each group had in common and/or what the student would title each group. To my knowledge, prior research has not reported results for a direct grouping task for submacroscopic and macroscopic objects; however, Tretter and colleagues (Tretter, Jones, Andre, et al., 2006) found that experts aggregately make more groups than novices, based on statistical processing of a task in which learners placed objects into pre-existing size ranges.

Next, students did the relative scale task shown in Table 4. This relative scale task employs the 1-mm head of a pin as reference object, rather than the human body as in previous research (Tretter, Jones, & Minogue, 2006). I selected this reference object because the millimeter may be known to students from their work with rulers (Wiedtke, 1990) and results in smaller relative scale numbers for submacroscopic objects than if using the human body as reference.

Next, students estimated and recorded their estimates for the absolute size of atom, red blood cell, human, and Earth. If they used English units, we asked them to also provide an answer in metric units. The relative scale and absolute size tasks in this study ask for the size of specified objects, whereas previous studies asked students to provide an object for specified sizes (Tretter, Jones, & Minogue, 2006). Each approach has its advantages. Providing size ranges to the learner minimizes the problem of not knowing how to represent small sizes, while providing specified objects reduces the problem of students not recalling objects that they do have a sense of size for.

Finally, we asked students how they thought relative scale and absolute size were related, if at all. To my knowledge, this question has not been posed to learners previously. However, it is an important question, because it may be at the heart of scale creation,

identified by researchers as a crucial skill in developing conceptual understanding of size and scale (Joram et al., 1998; Jones & Taylor, 2009). For the relative scale and absolute size tasks, if the students did not rank atom < cell < pinhead, but did have one of these and another submacroscopic object ranked smaller than pinhead, we substituted objects (e.g., for a student who ordered molecule < virus < cell < atom < mitochondrion < pinhead, we would use atom and mitochondrion to compare to pinhead. We avoided using the molecule as a substitute due to their enormous range of sizes). However, with students who ranked one or no submacroscopic object as smaller than the pinhead, we could not carry out some of the relative and absolute tasks; I explain how we coded these responses below. The use of the same objects across the four tasks allows us to determine whether a student's answers were consistent across aspects, independently of the factual accuracy of their answers (Vosniadou, 2003). The interview protocol is included as Appendix A.

Interviewing Procedure.

I conducted 54% of the interviews; three doctoral students in science education and two PhDs involved in the NCLT conducted the remaining interviews. Each one-on-one interview was audio recorded. We recorded student responses for ordering, grouping, and the relative scale estimates on a pre-formatted answer sheet (see Appendix C); respondents recorded their own absolute size estimates on the same answer sheet. We probed student responses following the interview protocol, and asked for clarifications when needed. We asked students to explain their thinking in an effort to ensure that students were responding to the prompts as intended. Scratch paper, pencil, and calculators were made available to the students. We provided information about the objects when these were unfamiliar to students, for instance, saying that the mitochondrion is a part of the cell that provides the energy. Most interviews took place in the library or a classroom of the schools; a small number were carried out in a meeting room of a public library.

Validity and Reliability

Validity

According to the Standards for Educational and Psychological Testing,

> Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests… The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations. (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 9.)

Evidence for the validity of test interpretation comes from five sources, as described below for this interview protocol.

Evidence based on test content involves ensuring that the test adequately represents the content domain (Goodwin & Leech, 2003). I examined both factual knowledge and consistency for four important aspects of size and scale that are present in the literature. These four aspects are compatible with Resnick's (1992) framework of four types of mathematical thinking and are thus important for the development of mathematical reasoning. The four aspects are also important in scientific thinking, as discussed above. I assess student knowledge using submacroscopic objects that have been identified as possible landmark objects by prior research (Tretter, Jones, Andre, et al., 2006). Three PhDs in science or science education examined the interview protocol for completeness. I also used the interview protocol to interview two experts, full professors in science or science education at a major research university and whose PhDs were in Chemistry and Theoretical Physics, respectively. Finally, the construct-centered design process ensures a tight linkage between test items and construct. Thus, content validity is strong.

Evidence based on response processes can also bolster the validity of a study. The interview format allowed for clarification and rephrasing of the questions, helping to ensure that responses matched the intended construct. Pilot testing allowed us to improve clarity, and asking students to explain their thinking allowed us to evaluate if they were

responding to questions as intended. In order to avoid errors due to procedural math mistakes, the interviewer provided calculators and scratch paper. Our use of scoring rubrics with examples and practice rounds of coding helped ensure that raters were coding as intended.

Evidence for validity based on internal structure is provided by the value for Cronbach's alpha for the set of interview questions, which was high, at 0.80. This shows that the interview protocol had high internal consistency. In other words, different subsets of questions from the interview had sub scores that are highly correlated, showing that they appear to measure the same underlying variable (knowledge of size and scale).

Evidence based on the relation to other variables can also bolster the validity of a study. Since there are no established tests to measure conceptual understanding of size and scale, I was not able to correlate the data obtained through this interview with any external measure. However, the scores on consistency of knowledge and on factual knowledge were strongly and significantly correlated. As will be discussed in Chapter 5, the mean score on most items increased after a focused intervention to build understanding and knowledge of size and scale.

The final source of evidence for validity comes from the consequences of testing. The curriculum implemented in Study Two was guided by the results of our interviews with middle school students at the same school district our summer campers attended. The fact that this curricular intervention managed to increase overall levels of knowledge provides evidence based on the consequences of testing for the validity of the construct. (Clearly, this evidence and the evidence based on relations to other variables are not independent in this case. Consequential validity is a relatively new and controversial idea, and it is not very clear how to study it – Goodwin & Leech, 2003).

Reliability

In order to assess and improve the reliability of this interview protocol, I conducted an item-response theory (IRT) analysis using ConstructMap (Kennedy, Wilson, Draney,

Tutunciyan, & Vorp, n.d.). This analysis can help detect items that are not working as intended, by examining the "infit" values. The infit or weighted mean squares value compares the expected slope of the item response function (an S-shaped curve showing the probability of answering an item correctly as a function of student ability) to the observed slope (Wilson, 2005). The infit values for all but two items were within the acceptable range of 0.75-1.33 (Wilson, 2005). One was grouping – which I did not consider to be hierarchical and so did not expect to be able to discriminate between respondents by ability. The second item with a poor infit was the final interview question, asking about the relationship between relative scale and absolute size, in general. I suspect that this poor infit is due to the wording of my prompt, which some respondents interpreted as referring to their specific estimates for relative scale and absolute size, instead of the relationship in general. The item generated useful and interesting information, in the students who interpreted it as intended, and I use it to illustrate students' thinking in Chapters 6 and 7. I did not use this item in defining levels - I substituted a measure with similar difficulty and better infit value: partial relative scale-absolute consistency. I removed the two items with poor infit values from the final IRT
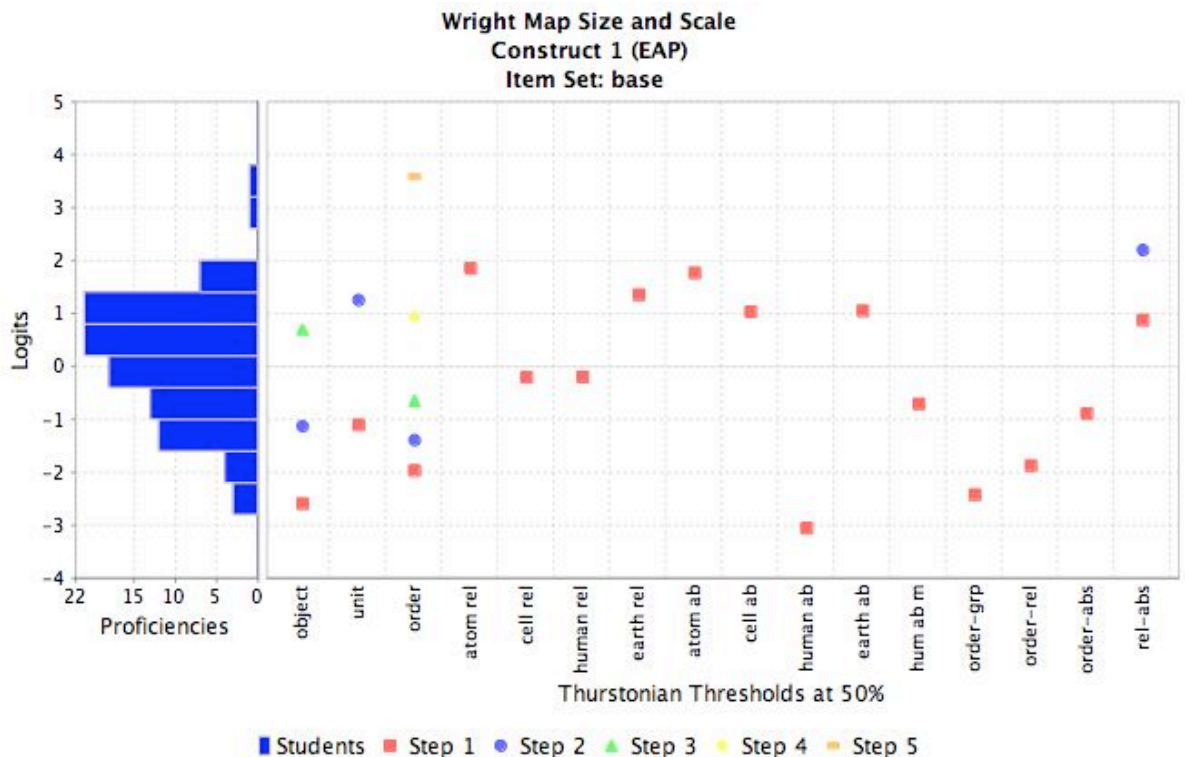
Figure 2: Wright map of IRT analysis

51

analysis diagnosing the difficulty level and infits for the items, and that generated the value of Cronbach's alpha.

Figure 2 shows a one-parameter Wright map displaying the estimated ability of respondents and the relative difficulty of the items on a single graph. The levels of ability of students (shown in a sideways histogram) and difficulty of items can both be conceptualized as measuring the amount of the construct present in a respondent or an item (Wilson, 2005). These levels are expressed in "logits", a probabilistic measure of the log-odds of an event occurring. A respondent at a given location has a 50% probability of correctly answering an item at that same location, a 27% probability for an item one logit higher, 12% for an item two logits higher, 5% for three logits higher, and just 2% for four logits higher. Conversely (and symmetrically), that respondent would have a 73% chance of correctly answering an item one logit lower than his or her location, and so on. Figure 2 shows the estimated difficulty of the items in "logits". There is a large gap in difficulty between order-group, order-relative, and order-absolute consistency, on the one hand (all at -1 logits or lower), and full relative-absolute consistency on the other hand (shown as the blue dot at +2 logits). The conceptual question about relative-absolute consistency and partial relative-absolute consistency were of very similar difficulty on the initial IRT analysis, and filled the gap (at around +1 logits). The IRT analysis also showed that the interview protocol covered a range of difficulties that spanned the abilities of the middle school through undergraduate students.

Cronbach's alpha for the set of interview questions was high, at 0.80, showing that the interview protocol had high internal consistency. Inter-rater reliability was 94% for factual coding and 95% for consistency (the coding procedure and calculation of inter-rater reliability is described below).

Data Coding

Coding for Factual Knowledge

I created a coding rubric that focuses on the accuracy of students' factual knowledge about the size of objects (included as Appendix D).

Smallest Object

For the smallest object, the rubric initially contemplated five exhaustive, non-overlapping size regimes of hundred-fold size differences, and one for non-matter responses (e.g., computer virus). After coding revealed that two categories were empty, I simplified to four: macroscopic object (coded 0), cell or microorganism (1), atom or molecule (2), and subatomic particle (3). These categories are similar to those in previous studies (Castellini et al., 2007; Batt et al., 2008).

Smallest Unit

For the smallest unit items, I initially generated five categories, but simplified to three after the IRT analysis showed that the difficulty was similar for two sets of two categories. The simplified coding categories were "do not know" or non-length unit (e.g., Newton-meter) (coded 0), macroscopic (e.g., inch, millimeter – coded 1), and submacroscopic (e.g., 1000th of an inch, micrometer, nanometer – coded 2). I used the smallest unit that the student came up with in responding to two questions: the unit for the smallest object known, and the smallest unit known. The second question (smallest unit known) was not asked in some of the interviews.

Ordering

The scoring categories for the 10-card ordering task focused on successively finer-grained distinctions that students might draw upon in ordering. The lowest category, coded 0, corresponds to errors in ordering the macroscopic objects (pinhead, ant, human, mountain, earth), or interspersing macroscopic and submacroscopic objects. This category can capture the students who believe that "The smallest thing that they can 'see' is the same as or similar to the smallest thing that they can 'think of'" (Waldron et al., 2006, p. 571). The remaining categories require correct ordering of macroscopic objects. Responses that ranked cell smaller than atom were coded 1. The cell (or single-celled microbes) may be the first submacroscopic object students encounter, in elementary school (AAAS, 1993). When they then encounter the atom in middle school (AAAS, 1993), they may not initially realize that the atom is smaller than the cell, and code 1 captures these students. Code 2 corresponds to ordering the atom as smaller than the cell, but not knowing how to order the atom relative to virus, mitochondria, or molecule. Code

3 was used for responses ranking atom smallest but cell not the largest of the submacroscopic objects. This code is designed to identify students who have established the atom as a landmark (Tretter, Jones, Andre, et al., 2006), but do not yet know where the cell belongs by size. Code 4 included responses with atom smallest and cell largest of the submacroscopic objects, but errors among molecule, virus, and mitochondria. This code is for students who appear to have established both atom and cell as landmarks. Code 5 was used for respondents who had all objects correctly ordered *and* could justify their ordering. For instance, a student response coded 5 might rely on recalling that atoms make up molecules (and thus atoms are smaller), knowing that mitochondria are parts of cells (and thus smaller), and knowing that cells and organelles can be seen with an optical microscope but atoms and molecules cannot (thus establishing that atoms and molecules are smaller than cells and organelles). If a student further recalled seeing an illustration of a virus injecting its genetic material into a cell that showed that the virus is smaller than a mitochondrion, then this information would suffice to establish the order atom < molecule < virus < mitochondrion < red blood cell. Students who ordered all objects correctly but who could not justify their ordering or admitted to guessing were coded 4.

Grouping

The coding scheme for grouping, unlike the others, is not hierarchical, as there are many correct ways to group and we did not ask or probe for a specific type of grouping. For instance, a macroscopic vs. submacroscopic distinction would result in two groups, while defining groups by the instrument required to visualize the object (eye, optical microscope, scanning probe microscope, etc.) might result in three or more, and defining groups by units might result in even more groups. This rubric assigns a code of zero if any groups are incorrect (e.g., grouping atom and virus without including molecule, which is intermediate in size). A code of 1 was used for responses with correct groups that however mix macroscopic and submacroscopic objects in the same group. The codes 2-5 were used for two correct groups, three correct groups, four correct groups, or five through nine correct groups (placing every object in its own group is coded as zero, as it does not involve making any groups; similarly for placing all 10 objects in a single group).

Relative Scale and Absolute Size

Some researchers have proposed "criteria of reasonableness" for estimates of quantity that are more exacting for smaller numbers; for instance, estimating that 100 objects are 50 would be considered less accurate than estimating that 10,000 objects are 5,000– despite the fact that the ratio of the actual number of objects and the estimate are identical in both cases (see Sowder, 1992). However, a large body of research has confirmed the empirical validity of Weber's Law, which states that only the ratio of two magnitudes affects their discriminability (e.g., McCrink, Dehaene, & Dehaene-Lambertz, 2007). Thus, I used the same criterion regardless of the size object being estimated. Student responses were coded as 1 for accurate if they were within one order of magnitude (ten times) of the accepted value, 0 otherwise (a criterion similar to that employed by Tretter, Jones, & Minogue, 2006). Students who ranked atom and red blood cell larger than the head of a pin were of course not asked to estimate how many times smaller these were than the head of a pin. If they did rank other submacroscopic objects as smaller than the head of a pin, these were substituted for the atom and/or cell, with corresponding criteria for accuracy. Else, the responses were coded as zero.

Coding for Consistency

The rubric distinguishes between two codes for consistency between ordering and grouping: 1 for evidence of consistency, and 0 otherwise (including evidence of inconsistency and lack of evidence due to the inability to order or group). The rubric is similar for order-relative, and order-absolute consistency. I coded relative-absolute consistency with an additional *partial* code: 2 for fully consistent (consistent for all four objects), 1 for partially consistent (two to three objects), and 0 for other (consistent in only one object, or none). The consistency rubric is presented as Appendix E. I further explain my coding rubric with an example using a student's responses, shown in Figure 3. This figure shows a transcription of the student's ordering and grouping at the top. The student ordered atom as smallest, virus as next smallest, and so on up through the Earth. The boxes enclose the student's groups. For instance, she grouped atom with virus, and cell with pin. The middle section of the figure show the student's estimates for the relative scale tasks, recorded by the interviewer (e.g., she estimated that the atom was

3000 times smaller than the head of a pin). The bottom section of the figure shows the student's estimates for absolute size, recorded by her; for instance, she estimated the human to be 5'8" and the Earth's diameter to be 700 cm. These responses were coded as shown below.
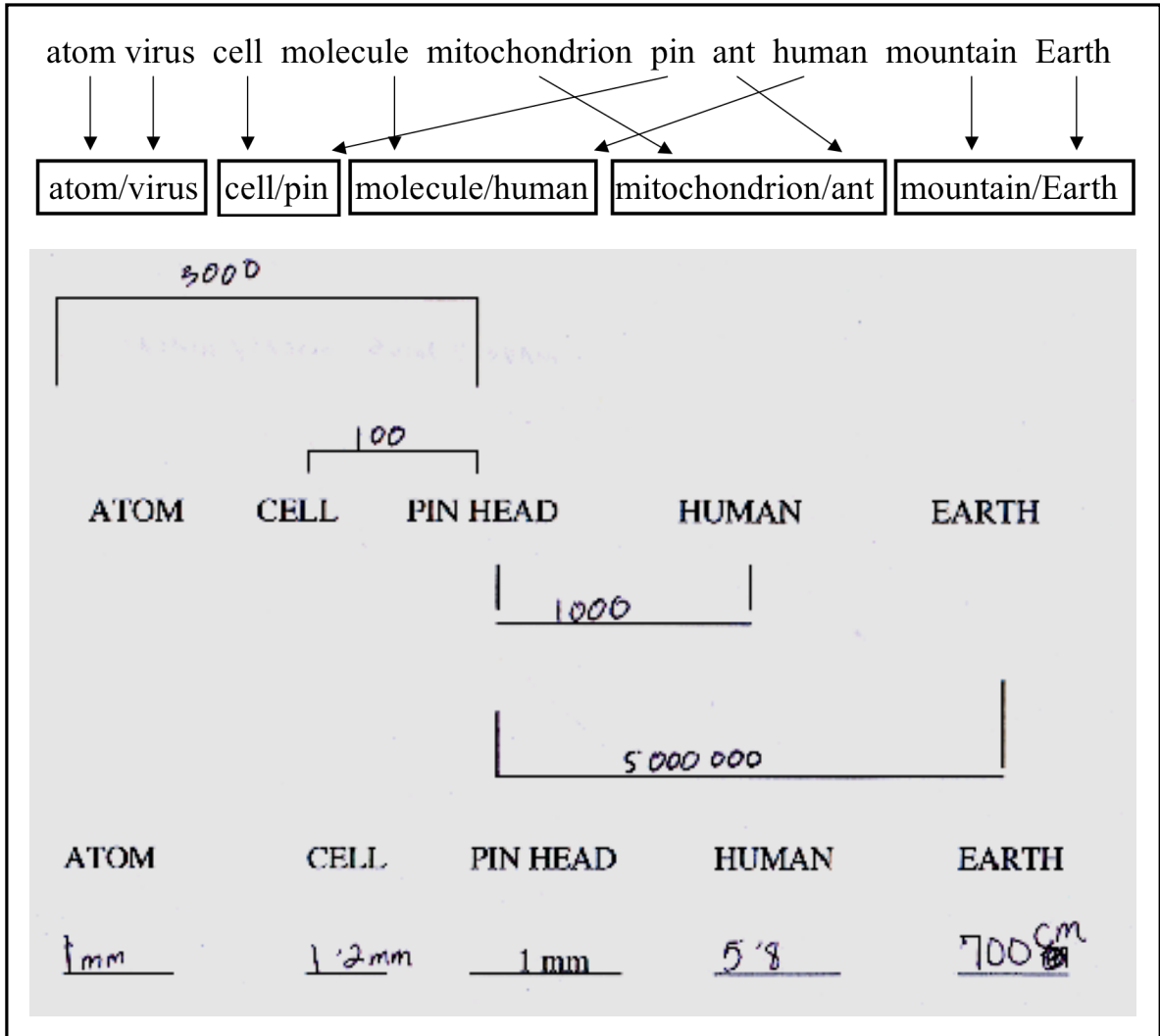
Figure 3: Sample student response (#1001)

Order-Group Consistency

Some of the groups are inconsistent with the ordering (as revealed graphically by the crossing arrows). For instance, the group containing mitochondrion and ant should also contain pin, as the pin was ranked intermediate in size in the ordering. This response was

thus coded zero. I probed the student about her grouping molecule with human:

> Interviewer: You grouped human and molecule together. Tell me about them…
> (pause 20 s).
> Respondent: I think that it's possible that they can be not the exact same size but
> (pause 8 s)… kind of similar…
> Interviewer: And they can still go together?
> Respondent: Mm hmm [Yes].

This unusual response, which appears to be due to a change in criterion by which the student is grouping, is discussed further in Chapter 6.

Order-Relative Scale Consistency

The student estimated a *smaller* number of times bigger or smaller for objects closer in size to the head of a pin, displaying consistency across order and relative scale. This response was thus coded one. An example of inconsistent order and relative scale is the student who estimated the cell was five times smaller and the atom three times smaller than the head of the pin, despite having ranked the atom smaller than the cell (#8020).

Order-Absolute Size Consistency

The student in Figure 3 estimated an absolute size for the cell that is larger than the head of a pin, despite ranking the cell smaller than the head of the pen. This is coded as inconsistent (zero).

Relative Scale-Absolute Size Consistency

In order to be consistent across relative scale and absolute size, the student would have had to estimate that the cell was 1/100 of a millimeter and the atom 1/3000 of a millimeter, given her relative scale estimates. Her absolute size estimates for human and Earth are likewise inconsistent with her relative scale estimates. Thus, the student's response was coded zero. Following the interview protocol, I probed her ideas about the relationship between absolute size and relative scale:

> Interviewer: Did you use the numbers up here [relative scale] to think about the
> numbers down here [absolute size]?
> Respondent: Uh uhh [no].

Interviewer: Like for instance, you said that a cell was 100 times smaller than a pinhead. How would that number…affect what you write down here…?
Respondent: [It's] not very related.

Scoring Procedure

The raters coded the items from the answer sheet or directly from the recording. I prepared a summary for each interview, which includes the responses not recorded on the answer sheet, and transcriptions of relevant or interesting passages (see Appendix F for a sample summary). I coded all the interviews, while a second rater coded a representative sample (10%) to ensure inter-rater reliability. The second rater did not use my summary in coding. After one round of coding and editing the rubric for clarity, we achieved inter-rater reliability of 95% for consistency and (with a different rater) 94% for factual knowledge. The inter-rater reliability is calculated as the proportion of codes that were in agreement relative to the total codes. All differences were resolved by discussion.

Data Analysis

Research Question 1A

What is students' factual knowledge of the size of important objects in science, by year in school, with existing curricula?

In order to characterize the factual knowledge of students, I generated bar graphs of mean codes for each task by grade, and clustered bar graphs showing percentages of students at each grade band (middle school, high school, undergraduate) that responded at each level of the coding rubric. This approach is similar to that employed by previous studies (Batt et al., 2008; Waldron et al., 2006; Castellini et al., 2007). I also created an overall score for factual knowledge including smallest object, unit, ordering, sum of relative scale estimates, and sum of absolute size estimates. I weighted the top code for each to three for smallest object and unit (for instance, the codes for smallest unit range from zero to two, and were multiplied by 1.5), and to five for ordering, relative estimates, and absolute estimates. I weighted smallest object and unit less heavily because they deal with a single object or unit, whereas the others deal with four or more. (Weighting all tasks to five results in very similar results.) This resulted in a variable with possible integer values from 0-21, which is a good approximation of a continuous variable. Scores were

normally distributed, permitting the use of a t-test in analyses. I generated a scatter plot of the overall factual knowledge score by grade in order to look for general trends. Using SPSS for Mac (v. 11.0.4) I used a t-test for independent samples to compare the overall factual knowledge of two groups of students from among those who did not mention the millimeter or smaller unit: those who subsequently did use the millimeter, and those who continued to use a centimeter or inch. Since learning progressions focus on the process of student learning and not just outcomes, it is important to characterize students' incorrect answers in order to get insight into their thinking and to tailor instruction accordingly. I examined student estimates for relative scale and absolute size to detect the direction in which students tend to err in their estimates for each object. Student estimates ranged over several orders of magnitude, so I plotted the base-10 logarithm of the estimates on a histogram.

Research Question 1B

What patterns of factual knowledge are observed in students with existing curricula?
In seeking to define *levels* of understanding for a learning progression for size and scale, I looked for patterns in the factual knowledge of students both *across* and *within* aspects. I also compared the degrees of difficulty of different aspects, where the coding schemes are similar enough to support these comparisons.

Across Aspects

A learning progression includes intermediate understandings that are "reasonably coherent networks of ideas and practices and that contribute to building a more mature understanding" (Duschl et al, 2007, p. 220). Two hypothetical levels of understanding that rely on networks of factual knowledge *across tasks and/or aspects* in the context of size and scale might be as follows:

A) A state of knowledge where a student does not know of units smaller than a millimeter. This hampers his ability to estimate the absolute size of a cell or an atom, and this in turn hinders him from building an accurate idea of their relative scale. Thus, the codes for these three tasks are 1 (macroscopic unit), 0 (inaccurate relative scale), 0 (inaccurate absolute size).

B) A state of knowledge where a student has recently learned of units like the micrometer and the nanometer. This helps her accurately estimate the absolute size of cell and atom, which helps her also develop an accurate idea of the relative scale. The codes for these three tasks are now 2 (submacroscopic unit), 1 (accurate relative scale), 1 (accurate absolute size).

If most students tend to follow this hypothetical pathway, then there should be a strong correlation between the score on smallest unit known, and the codes for absolute size and relative scale of atom and cell, as they would all increase in tandem (except for those few cases where students have very recently learned about units and still have not restructured their knowledge in terms of relative scale and absolute size). Therefore, I calculated and examined the correlations between codes for factual knowledge in order to detect this type of patterns across aspects. I used Kendall's tau-b correlations (a non-parametric measure of association) between tasks with four or more hierarchical levels, and the chi-square based measure of association phi for tasks with binary coding.

Within Aspects

The second type of pattern I looked for that can help define levels is *within* a single aspect. The coding rubrics for smallest object, unit for smallest object, and ordering by size already impose hierarchical levels based on theory (e.g., for the smallest object question, a macroscopic object is not as good an answer as a cell, atom is a better answer than cell, and subatomic particle is better yet), and the coding for grouping is non-hierarchical. However, the relative scale and absolute size estimates for the four different objects have independent codes for accuracy. It might be that students tend to learn how to estimate the relative scale and absolute size of the human, first; only after they can do this, will they be able to estimate the size of the other objects correctly. Thus, there might be a pattern in the order in which accuracy of estimation for the objects develops. I looked for a pattern within aspects using the K-means classification-only routine in SPSS 11 for Mac, which automatically gathers like student answers into a cluster. By assigning the maximum number of theoretically possible clusters (24 for four objects, the number

of possible permutations), I can see how many of the possible patterns actually exist in the data. Then I can see if these are consistent with a developmental pathway.

Having examined the central tendency of student estimates for research question 1A, I examined the spread of estimates for the relative scale and absolute size estimates for all four objects, by generating graphs of the logarithm of the estimates and summarizing these results with box and whisker plots.

Difficulty of Aspects

I compared the difficulty of ordering and grouping by comparing percentages of fully correct answers, using a McNemar test, which is a non-parametric test for two related dichotomous variables. I also compared the percentages of students who know of submacroscopic objects with the percentage of students who know of submacroscopic units, using a McNemar test. The quantitative tasks of relative scale and absolute size involved five objects, but the qualitative tasks (ordering and grouping) involved 10 objects, thus I could not compare difficulty across the qualitative-quantitative divide. I did however compare the difficulty of absolute size and relative scale estimation by comparing overall success rates at the same criterion for accuracy, using both a t-test for paired samples and the non-parametric Wilcoxon signed ranks test. Neither test is ideal because the t-test requires continuous variables, and these are five-level ordinal variables that only approximate a continuous variable. The Wilcoxon test is also not ideal because of the presence of multiple tied cases (42 out of 94). Nevertheless, both tests provided an indication of the significance of the difference in difficulty between relative scale and absolute size estimation, if any. I also compared the difficulty of absolute size and relative scale estimation for each individual object using a McNemar test. I generated a box and whisker plot to visually compare the distribution or spread of student estimates of relative scale and absolute size for all four objects.

Research Question 1C

What connections (as inferred by consistency) do students make across aspects of size and scale, by year in school, with existing curricula?
In order to characterize the consistency of knowledge of students, I generated a bar graph

61

of mean sum of consistency codes by grade, and a clustered bar graph showing percentages of students at each grade band (middle school, high school, undergraduate) that were at each sum of consistency codes. I also calculated percentage of students who were consistent for each type of consistency.

Research Question 1D

What patterns of consistency are observed in students, with existing curricula?
I looked for a pattern using the K-means classification-only routine in SPSS 11 for Mac, as described above (within aspects).

Research Question 2A

What is students' factual knowledge of the size of important objects in science, after a focused curriculum?
I calculated the statistical significance of students' gains for each item using a paired-samples test. For items with dichotomous coding (e.g., accuracy of estimate of absolute size for cell), I used the McNemar test. For items with four or more levels, I used the McNemar-Bowker test, which is an extension of the McNemar test for square matrices. When the McNemar-Bowker test was not feasible due to empty cells creating a non-square matrix, or in the case of variables with many possible levels approximating a continuous variable (i.e., overall factual knowledge score) I used a paired-samples t-test. This test is for continuous variables, which the items with 6 or more levels approximate. With the t-test, I also calculated effect sizes by dividing the gain by the pooled standard deviation (the square root of the average of squares of pre- and post-camp standard deviations).

Research Question 2B

What is students' consistency of knowledge across aspects, after a focused curriculum?
I followed the same procedure as for research question 2A.

Research Question 2C

How do the patterns of factual knowledge and consistency differ from those of students who have not experience the focused curriculum, if at all?

I followed the same procedures as described above for research questions 1B and 1D. I also calculated the Pearson correlation for initial level of consistency vs. consistency gains, to test for patterns of learning gains.

Research Question 3

What is the relationship between factual knowledge and consistency of knowledge, if any?

In order to address this question, I conducted a theoretical task analysis for each task, and then an empirical task analysis where I examined students' actual strategies. I carried out the theoretical task analysis by generating all the strategies of which I could think or which I found in the literature to carry out *generic* factual knowledge tasks for each aspect of size and scale. I used data from Study One to calculate the correlation between factual knowledge and consistency of knowledge. I used Pearson's correlation since the factual knowledge score ranges from 0-21 and is close to normally distributed, while the consistency score has six ordinal levels, and thus approximates a continuous, normally distributed variable as well. In order to contextualize the strength of this correlation, I compared it to the correlation between factual knowledge and grade using Cohen and Cohen's t-test (1983, p. 57). I expected the level of factual knowledge to be correlated positively to grade level, as formal schooling is likely to be the main source of information about the size of objects like the atom, the cell, and the earth.

Using the data from Study Two I conducted a multiple regression using the final overall score of factual knowledge as outcome, and the initial levels of factual knowledge and consistency as predictors, in order to evaluate the relative importance of each type of knowledge in the subsequent accumulation of additional factual knowledge. I conducted another multiple regression using the final level of consistency as outcome, with the initial levels of factual knowledge and consistency as protectors; this enables me to evaluate the relative importance of each type of knowledge to constructing an increased connectedness across aspects of size and scale

General Limitations

In this section I enumerate various limitations that are general to the dissertation. In addition, I describe limitations specific to each study in the corresponding chapters.

One limitation of this study is that I did not examine all possible ways of thinking about size and scale. For instance, I did not examine analogies of the sort "imagine that the Earth were the size of the head of a pin. Then the Sun would be a diameter of 10 cm and at a distance of 10 m from the Earth." (Tretter, Jones, Andre, et al., 2006, p. 286). However, size analogies can be seen as the coordination of two relative scale relationships or four absolute sizes, so the components of analogies are included. Furthermore, such analogies would not help students understand relative scale or absolute size, only the "relative size" of one object to another (Tretter, 2006) in qualitative terms. I also did not consider thinking about distances in terms of time, e.g., "the distance between Chicago and Detroit is four hours" (driving time, on the freeway); however, in students' everyday life, it would not be applicable in thinking of the size of a cell or an atom.

Another limitation is that size and scale, by virtue of its being a common theme (AAAS, 1993) that cuts across content areas, is not a typical science topic. In terms of the four strands of scientific proficiency (Duschl et al., 2007), learning about size and scale aids students in participating productively in scientific practices and discourse. However, learning about size and scale per se does not involve knowing, using, and interpreting scientific explanations of the natural world; generating and evaluating scientific evidence and explanations; or understanding the nature and development of scientific knowledge. Instead, it is a component in these strands. Thus, the learning progression for one-dimensional size and scale developed here needs to be interwoven with other learning progressions for core concepts of science (Duschl et al., 2007) in order to fully comply with the postulated characteristics of a learning progression. However, as size and scale is known to be a common theme that relates to student learning across topics, disciplines, and grades (AAAS, 1993), there should be ample opportunities for learners to interweave their learning progression of size and scale with others for more typical scientific content.

In fact, this learning progression could conceivably be used as a framework to organize science and mathematics learning, as discussed below.

An additional limitation of this study, and threat to validity, is the question of whether students were familiar with the objects for which they were ordering, grouping, and estimating sizes. Chi and Ceci (1987) reviewed various studies showing that familiarity or salience of objects to be classified had a strong impact on performance. We addressed this during the interview by asking whether the respondent was familiar with all of the objects. Students who did not recognize the mitochondrion, for instance, were told that it was a part of a cell responsible for energy. Whether students effectively "knew" about the objects they claim to be familiar with was not tested; and this is a weakness of the present study. However, this is a feature that the study shares with previous studies on this topic, and a feature that is to a large degree inevitable in science learning. Students are building their knowledge about scientific objects in part by learning about their size, but also building knowledge about size by learning about scientific objects. Lack of knowledge about the specific objects is not a threat to the validity of the assessment of consistency, as consistency does not depend on the accuracy of students' factual knowledge of the size of objects (Vosniadou, 2003). Thus, my strategy of examining students' consistency across aspects reduces the impact of students having inaccurate content knowledge.

As there is no prior research on students' consistency of knowledge across aspects of size and scale, the coding rubric was developed without the support of the literature. However, a research group that includes PhDs and doctoral students in science education provided helpful feedback in the development of this rubric.

The question about the smallest unit known was not asked in the first round of interviews, making the results less representative than would otherwise have been the case.

We were unable to assess the accuracy of relative scale and absolute size estimation for submacroscopic objects in the case of students who ordered the atom and the cell larger than the pinhead. These students could conceivably know of other submacroscopic

objects and even be capable of estimating accurate relative scale or absolute size estimates for these. However, the likelihood of this possibility is low, since the atom and cell are the fundamental units of chemistry and biology, and are likely to be among the best-known submacroscopic objects.

Chapter Summary

In the previous chapter, I pointed out that factual knowledge and consistency of knowledge are both important in studying student learning of size and scale. In this chapter I outlined how I study both types of knowledge simultaneously, through tasks using the same objects across different aspects. By asking students to order, group, estimate relative scale, and estimate absolute size for the *same* objects, consistency can be assessed independently of the accuracy of their factual knowledge. This procedure also yields data concerning students' actual knowledge.

In this chapter I also set forth my methodology to develop a first iteration of a learning progression, LP1, using cross-sectional data and correlational analysis. I search for levels by looking for *patterns* within and across aspects. Patterns *within* aspects can be detected using a data classification routine, and can reveal the order in which factual knowledge develops for various objects. The data classification routine can also reveal the order in which consistency develops, as the learner establishes successive connections between pairs of aspects. Patterns in factual knowledge *across* aspects can be detected with correlational analysis, and can show networks of understanding at given levels. However, a cross-sectional study provides only a rough approximation of student learning. A true longitudinal study or teaching experiment can show how students actually progress.

I also described my methodology to develop a second iteration of a learning progression, using data from a teaching experiment and theoretical and empirical task analyses. These data sources allow me to verify whether students actually progress along the levels generated from the cross-sectional study, further characterize the upper anchor, and explore how students advance from level to level.

I described my strategy of using a theoretical task analysis for each aspect to reveal if factual knowledge and consistency of knowledge *could* be related. An empirical task analysis using the data from the cross-sectional study and the teaching experiment then shows if students actually employ the strategies that depend on relationships between factual knowledge and consistency. I then use a correlational analysis to reveal the empirical importance of this relationship, given students' current curriculum. Next, I use regression analyses to further characterize the relationship between factual knowledge and consistency of knowledge, in the context of a teaching experiment.

I next report my findings from Study One, a cross-sectional study that helped me generate a first iteration of a learning progression for size and scale.

CHAPTER 4: RESULTS AND DISCUSSION FOR STUDY ONE

In this chapter I present and discuss my findings from Study One, a cross-sectional study of 101 students from middle school through undergraduates, who had experienced their normal curriculum. This study addresses research question 1:

1) What do middle school through undergraduate students know about one-dimensional size and scale, given their current curricular experiences?
   A) What is students' factual knowledge of the size of important objects in science, by year in school?
   B) What patterns of factual knowledge are observed in students?
   C) What connections (as inferred from consistency) do students make across aspects of size and scale, by year in school?
   D) What patterns of consistency are observed in students?

For each research question, I report results by interview task or analysis, then discuss how the findings build upon and extend the literature, and state the implications for the learning progression.

Students' Factual Knowledge by Year in School

In this section I address research question 1A: What is students' factual knowledge of the size of important objects in science, by year in school? These data can inform the lower and upper anchors of the learning progression. As I mentioned in Chapter 2, there are previous research studies that shed light on students' factual knowledge, but they leave important gaps.

The mean scores of students' responses for the smallest object of which they could think are displayed in Figure 4 (macroscopic responses are coded 0, cell/microorganism 1, atom/molecule 2, and subatomic particle 3). There is an overall trend towards more sophisticated answers for older students.
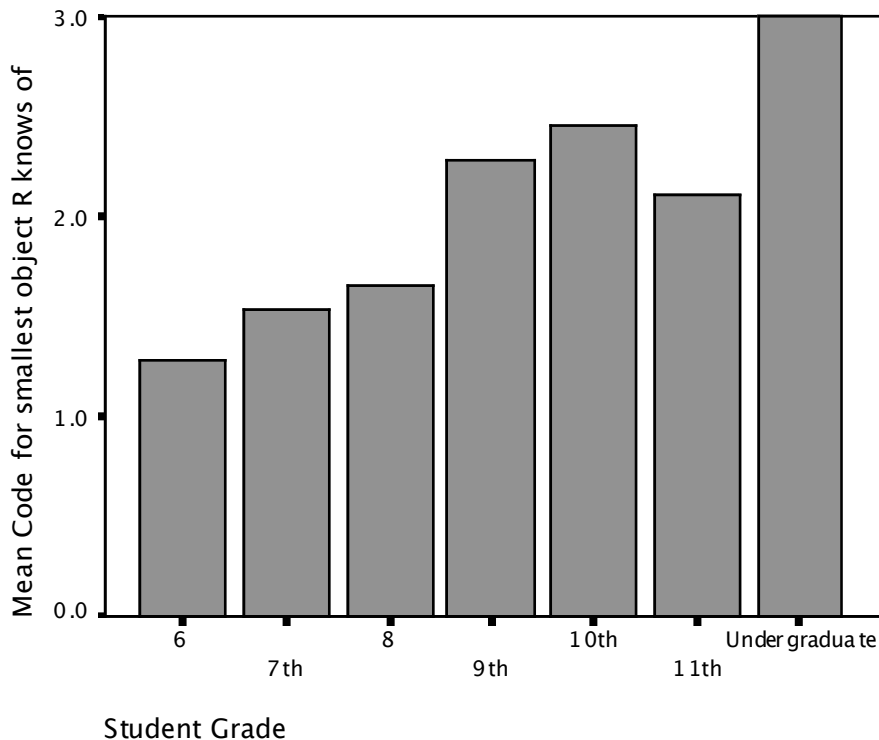
Figure 4: Mean score for smallest object of which respondent can think, by grade

In this graph, 11th graders can be seen to have performed more poorly than 10th-graders – and the same trend is visible for several other interview questions, as shown below. This trend is due to a handful of low-performing students, not a general trend for 11th-graders to perform at lower levels than 10th-graders. These low-performing 11th-graders are discussed further in Appendix G.

The clustered bar graph in Figure 5 shows the percentage of students at each grade band (middle school, high school, undergraduate) that answered within each category of

response. For instance, around 7% of the middle school students responded with a subatomic particle - as shown by the black section at the bottom of the middle school bar. The dark gray section above the black section, extending to around 53%, represents the 46% of middle school students who responded with the atom or molecule (53% - 7% = 46%); followed by 32% of responses of cell or microorganism in light gray and around 15% macroscopic objects in white. Middle school students are most apt to think of atom/molecule or cell/microorganism; high school students tend to respond with atom or subatomic particle; and the five undergraduates responded with a subatomic particle. Many students initially responded with a macroscopic object, but came up with a submacroscopic object after being prompted for "an object too small to see with the naked eye". However, six of 41 middle school students and three of 55 high school students continued to respond with a macroscopic object, despite being prompted. Macroscopic responses included grain of sugar, skin flakes, baby ant, and gnat. Over half of the high school students who had taken or were taking Chemistry responded with a subatomic particle.
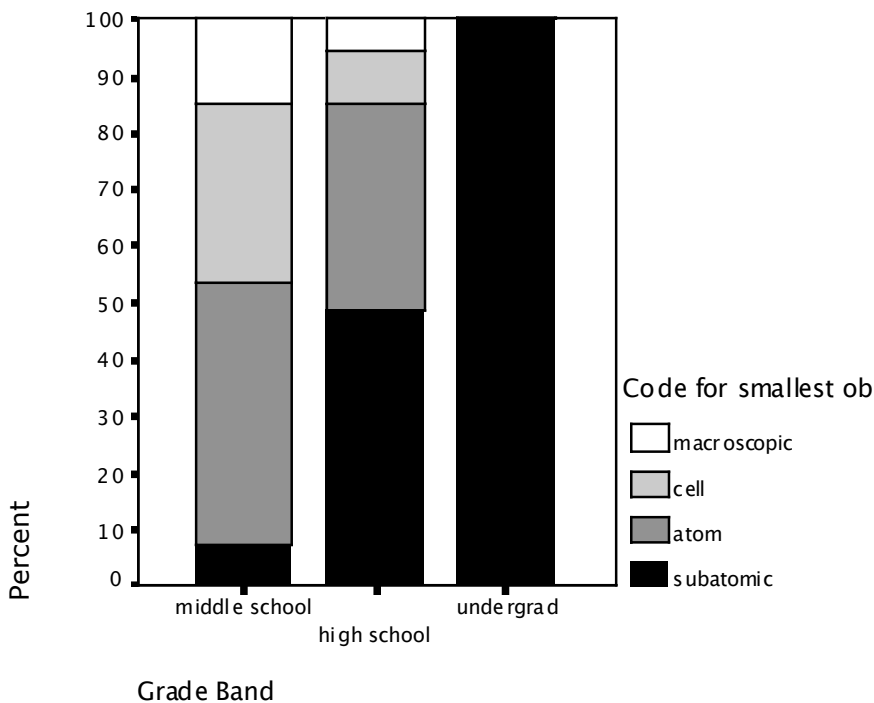
Figure 5: Percentage of students who answered at each category for smallest object question, by grade band

Discussion

These results are broadly similar to those of previous studies with equivalent questions, with older respondents usually responding with smaller objects and a gradually increasing proportion of sub-atomic responses. Waldron, Batt, and colleagues (Waldron et al., 2006) reported a much higher percentage of macroscopic responses among respondents of all age groups (children through adults). Their survey-based study asked about the smallest object the respondent could *see* immediately before asking about the smallest object the respondent could think of, possibly predisposing some respondents to continue thinking of macroscopic objects (the survey methodology did not include prompting or probing - C. Batt, personal communication, April 13, 2009). As noted above, many students in my study first responded with a macroscopic object but then provided a submacroscopic answer after being probed about objects too small to see. For example, student #0144 said: "I have to think molecular, don't I? First I thought a flea…" before responding with an atom and then nucleus of an atom. Student #1004 answered "ant" initially, then mentioned the cell after the prompt.

A survey-based study by Castellini and colleagues (2007) found a smaller proportion of 9-10th grade students responding with the atom than middle school or 11-12th grade students, but I did not observe this drop. The respondents in my study also performed at a higher level than those in this study, with an absence of "nonsense" answers (up to 14% in some age ranges, in the Castellini study). This is probably due to the probing and clarification opportunities inherent in the interview format. This comparison across studies appear to indicate that the smallest object of which students can think is sensitive to both context (previous question and follow-up prompt) and format (an interview format allows for probing), and that an interview format is likely to produce more sophisticated answers. However, this conclusion is tentative because the studies involved different students in different states.

Atoms and cells/microorganisms are potential landmark objects for students, since they brought these up before we had introduced them in the interview (67% of students mentioned an atom or part of an atom; 15% a cell or a one-celled organism). This finding

is consistent with and elaborates upon Tretter and colleagues' (Tretter, Jones, Andre, et al., 2006) finding that the atom and "microscopic objects" are landmark objects for experts.

Implications for the Learning Progression

Most but not all middle school students in this study and others thought of an atom or subatomic particle, or a cell or microorganism, as the smallest object of which they knew. Many others in this study initially responded with a macroscopic object, and only thought of the atom, subatomic particles, cells, or microorganisms after prompting for an object too small to see. This suggests that the lower anchor should consider small macroscopic objects as the prior knowledge middle school students are apt to have. The learning progression should include instructional activities to help build the atom and the cell as landmark objects for the submacroscopic world. Subatomic particles can feasibly be included in the upper anchor, as improved curriculum with a greater focus on the common theme of size and scale should increase the percentage of pre-university students that respond at this level.

Unit for Smallest Object

The mean scores by grade of students' responses for the unit of measurement for their smallest object and/or smallest unit known are shown in Figure 6 ("do not know" or non-length units were coded 0, macroscopic unit 1, and subatomic unit 2). Pre-college students had little knowledge of measurement units appropriate for submacroscopic objects, but all five undergraduates mentioned the nanometer (1 billionth of a meter) or picometer (1trillionth of a meter). No student mentioned the Angstrom, a non-SI unit that equals 0.1 nm and which is often used in chemistry textbooks to express the radius of atoms and ions. The sixth graders' good performance in this and some of the other tasks (see below) may be explained by the fact that they all were summer campers. While this interview took place before camp, these students may have greater than typical interest in science; additionally, they may have remembered hearing about nanometers in the camp orientation meeting that took place several days before the pre-camp interviews. A few younger students replied using a fraction of a macroscopic unit. The nine students who responded with macroscopic objects are not included in Figure 6.
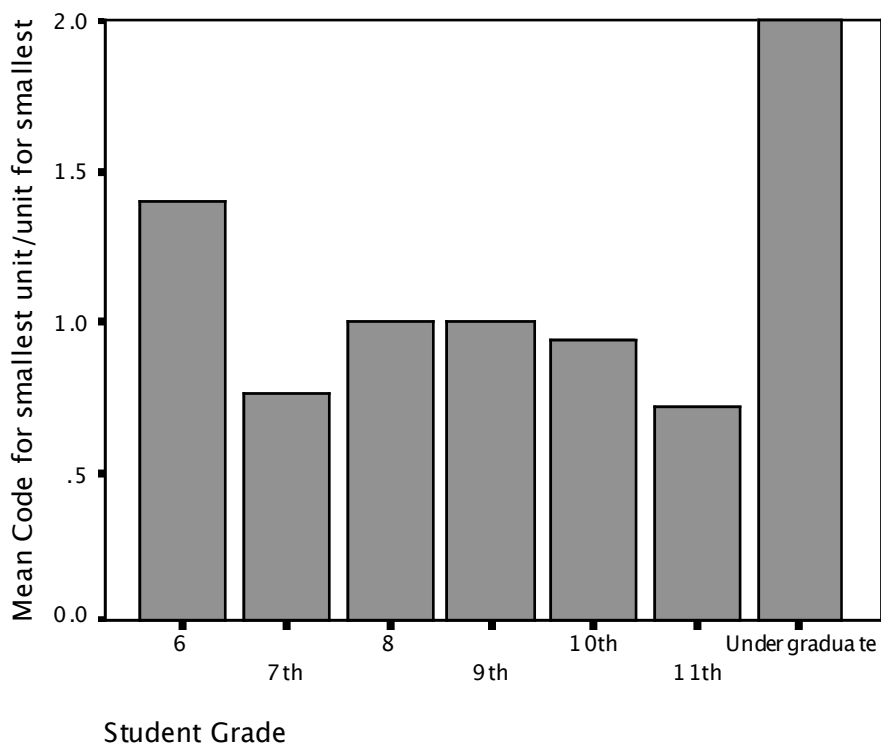
Figure 6: Mean score for unit for smallest object, by grade

The clustered bar graph in Figure 7 shows the percentage of students at each grade band that answered within each category of response. Around 85% of middle school students and 70% of high school students were unable to provide a unit of length for submacroscopic objects. Any length can be expressed in any unit (e.g., a carbon atom is around 0.1 nm in diameter, but this can be expressed as 10^{-10} m or 10^{-13} km), but only a few students thought of this strategy (e.g., #9037, 9026, 0087, 0109). Out of 66 students who did not know of a unit for their smallest object, we asked 38 for the smallest unit of which they did know (we did not ask this follow-up question in early rounds of interviews). Only five students then responded with a submacroscopic unit. Thus, the percentages of students who know of submacroscopic units *in general* (as compared to in relation to the smallest object of which they know) may be slightly larger than presented in Figure 7, due to the 28 who were not asked the follow-up question.
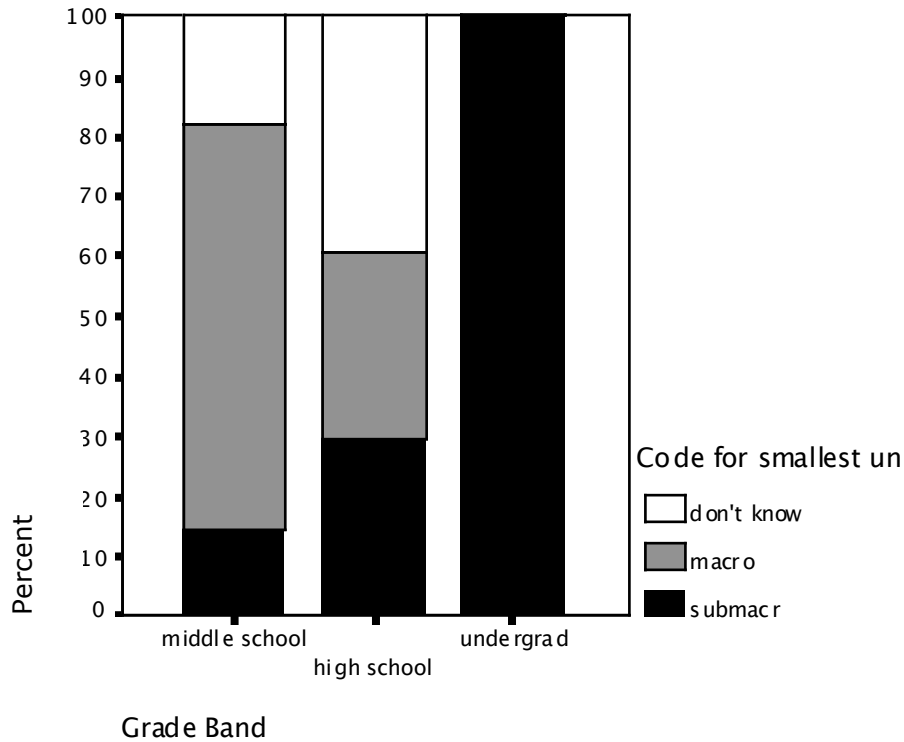
Figure 7: Percentage of students who answered at each category for smallest unit questions, by grade band

Discussion

A recent study by Jones and colleagues (2007) asked 7-9th grade girls what a millimeter is; only 18% correctly defined the millimeter as 1/1000 of a meter, and "knowledge of specific measures of metric scale beyond the meter was uncommon" (p. 198). This finding is consistent with those of larger scale assessments (NCES, 1996). In my study, around 41% of students mentioned units larger than a millimeter (e.g., centimeter or inch) in response to the interview question(s) concerning units. Half of these students did not use the millimeter even after this unit of measurement had been provided to them in the subsequent interview tasks. These students who continued to use the centimeter or inch scored significantly lower on factual knowledge and consistency of knowledge than those who did not initially respond with the millimeter but later used it ($p < 0.01$ for t-test, and effect sizes above 1, in both cases). This study thus shows that some students may know of millimeters but not initially recall them, while others simply appear not to know the unit. This is very similar to our finding with the smallest known objects, where some

74

students knew of submacroscopic objects but did not initially recall them while others seemed not to know of submacroscopic objects at all.

Implications for the Learning Progression

Among middle and high school students in this and other studies, very few students were familiar with units of measurement convenient for submacroscopic objects. Only a few students in this study thought of using decimals or scientific notation in tandem with a known unit like meter or millimeter in order to express the size of submacroscopic objects. Many students did not think of the millimeter, and some continued to use the centimeter or inch even after they were provided with the millimeter. Other studies have shown a lack of familiarity with metric units in general, and with the millimeter specifically (e.g., Jones et al., 2007). The lower anchor for middle school students should thus only assume familiarity with the inch and probably centimeter, but not smaller units. An examination of science and mathematics curriculum materials for the elementary grades might shed additional light on this subject, but lies outside the scope of this dissertation.

Even though undergraduates knew of units like the nanometer, micrometer, and picometer, most high school students did not, and therefore this study does not conclusively inform the upper anchor. The undergraduates (at a highly ranked research university) are a select group and not representative of the middle and high school students in this study. The fact that students do not know of units smaller than the millimeter does not necessarily imply that they cannot learn them, and this is an empirical question that can be addressed through a teaching experiment. Additional evidence concerning what it is reasonable to expect middle and high school students to learn is presented in Chapter 5.

Ordering by Size

The mean scores of students' responses for ordering 10 cards are shown in Figure 8. There is a gradual trend towards more accurate ordering with increased schooling, except for the 11th grade dip (see Appendix G).
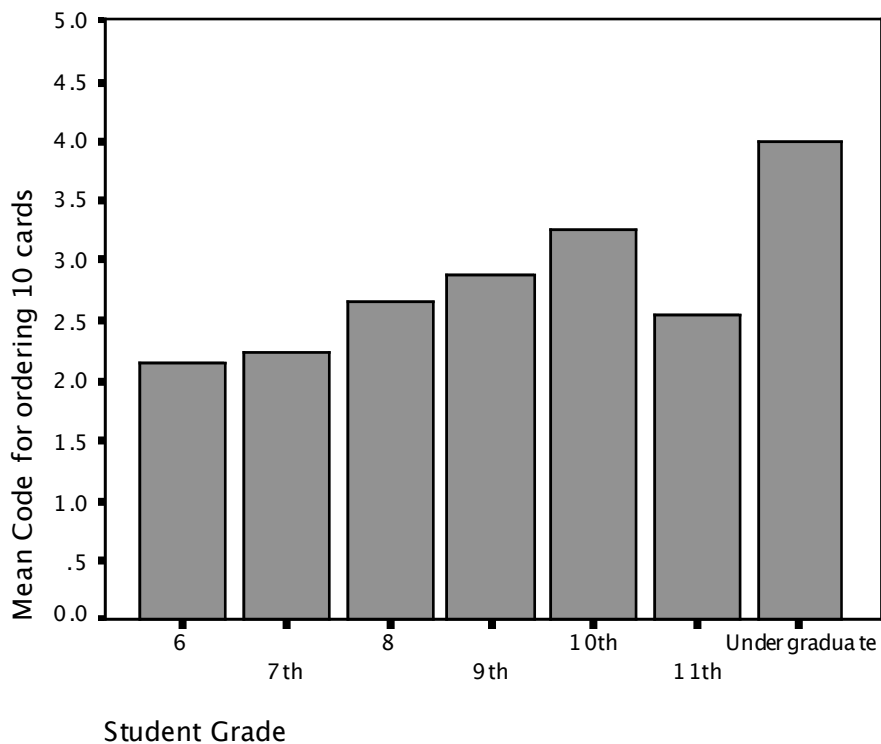
Figure 8: Mean score for 10-card ordering, by grade

The clustered bar graph in Figure 9 reflects this trend with increasing percentages of darker (more accurate) bars. The students represented by the white bars interspersed macroscopic and submacroscopic objects; usually, these students placed the pinhead as the smallest or second smallest object. All other students correctly ordered the macroscopic objects, separately from the submacroscopic objects. Only three students (one in each grade band) were able to order correctly without guessing; these are depicted in black. The black crosshatched bars represent the students who were able to order correctly but not justify their answer, along with those who correctly ordered the macroscopic objects, and the atom and the cell (as smallest and largest submacroscopic object, respectively) but had errors in ordering molecule, virus, and mitochondrion. The percentage of students at this category increases strongly with additional schooling (although the undergraduates are additionally a select group, in addition to having more schooling). The dark gray areas represent students who ranked the atom as the smallest

object, but did not have the red blood cell as the largest of the submacroscopic objects. The dark gray crosshatched areas represent the students who had cell and atom both out of order, but did have the atom ranked smaller than the cell; with additional years of schooling, a decreasing proportion of students fell into this category. The students represented in light gray ranked the atom larger than the cell. Most students at every age group were able to correctly rank the macroscopic objects (the rest mainly interspersed submacroscopic objects between macroscopic objects). Overall, 12% of students ranked the cell as smaller than the atom (including some students who interspersed macroscopic and submacroscopic objects). Over 15% of middle school students interspersed macroscopic and submacroscopic objects when ordering, with another 30% not ranking the atom as the smallest object from among the 10.
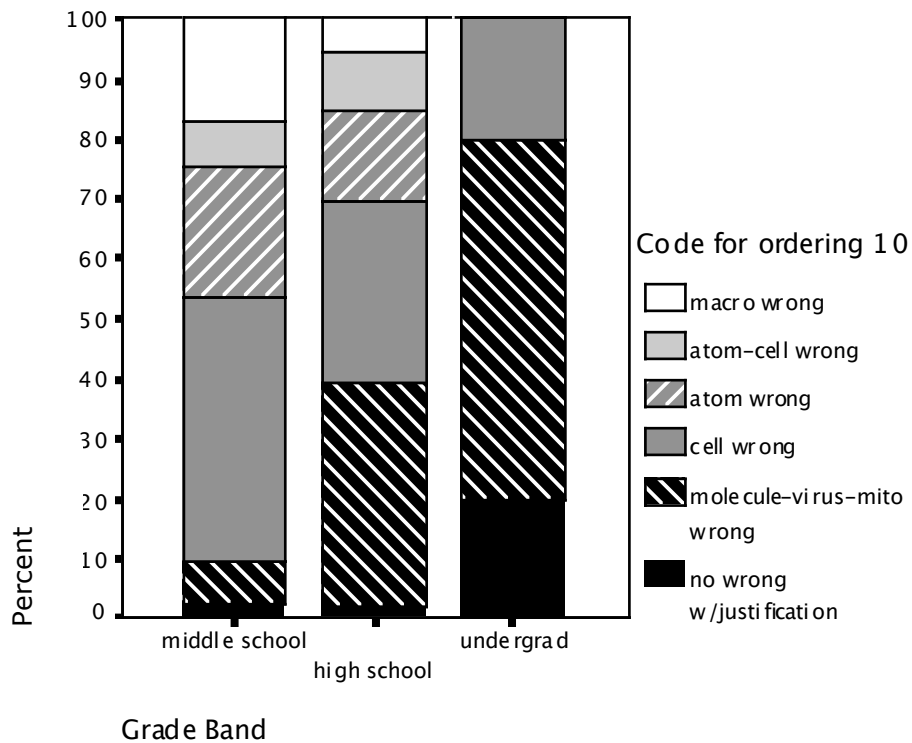
Figure 9: Percentage of students who ordered at each category, 10-card ordering task

Discussion

The students in this study performed as well on more difficult tasks, or more accurately on tasks of similar difficulty, than those in previous, survey-based studies (Castellini et al., 2007; Waldron et al., 2006). For instance, while I found that 60% of respondents correctly identified atom as the smallest object but erred in ordering the remaining objects, Castellini and colleagues (2007) reported 45%. Castellini and colleagues found that only 45% of respondents were able to correctly order by size four macroscopic objects (housefly, dust, eyelash, and grain of salt – it is not clear if the survey specified the dimension of the eyelash to be considered), whereas all students ordered pinhead, ant, human, mountain, and Earth correctly in my study (albeit, with some students interspersing submacroscopic objects). The interview format and the inclusion of macroscopic objects in the ordering task resulted in unexpected and informative findings that show that even some high school students believe that the smallest object that exists is the smallest object that can be seen.

Implications for the Learning Progression

Around 30% of students did not rank the atom as the smallest object from among the 10. The literature reports similar findings. Over 15% of middle school students in this study interspersed macroscopic and submacroscopic objects when ordering. Along with the expectation mentioned above that students might not know objects that are too small to see, this implies that we can expect minimal ability to order macroscopic and submacroscopic objects accurately at the beginning of middle school. Most of the undergraduates in this study were able to accurately order atom, cell, and macroscopic objects, so this might be a reasonable expectation for the upper anchor, with improved curriculum. However, it remains to be seen what can be accomplished with a focused curriculum on size and scale.

Grouping by Size

The responses by grade band on the 10-card grouping task are summarized in Figure 10. I do not present a graph showing the mean performance by grade band because my coding for grouping is non-hierarchical, except for the distinction between correct and inaccurate grouping. The figure shows that the proportion of students grouping correctly increases

with years of schooling. Among the students who grouped correctly, very few placed macroscopic and submacroscopic objects in the same group (shown as "mixed" in light gray). Around 55% of all students were able to group correctly; this percentage was around 40% for middle school students in the sample.
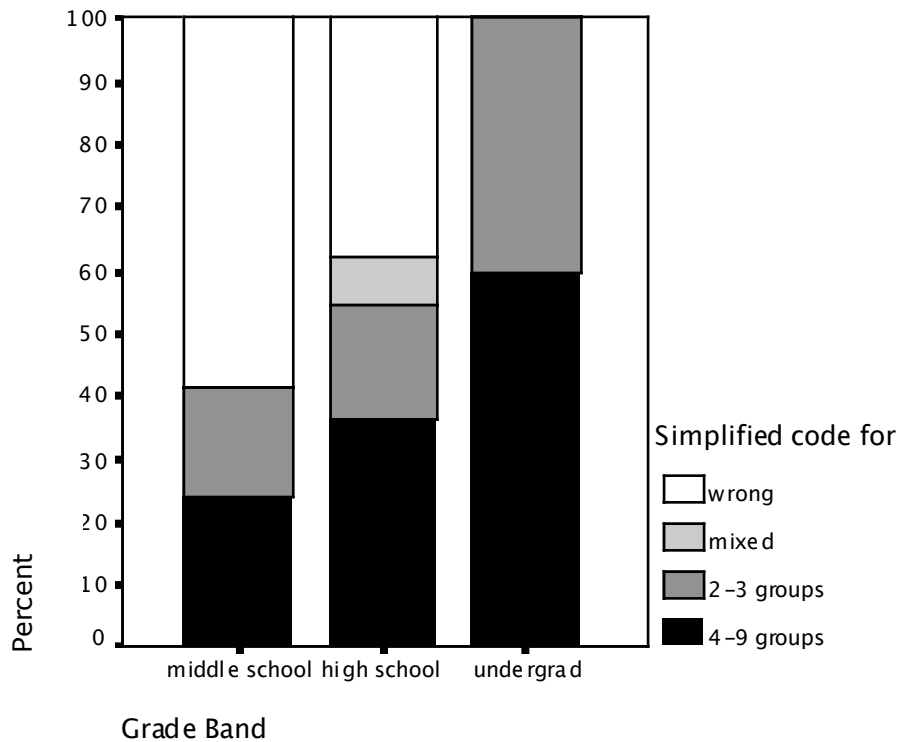
Figure 10: Percentage of students who grouped at each category, 10-card grouping task

Discussion

Among the students who grouped correctly, making 4-9 groups was about twice as common as making only two or three groups, for *all* grade bands. Prior research found that *in the aggregate,* older participants had larger distinctions between objects of different sizes, leading to the interpretation that "In general, the older the participants, the more distinct size categories they conceptualized" (Tretter, Jones, Andre, et al., 2006, p. 293). Since I did not code for the number of groups students made if their groups were inaccurate, these two results are not necessarily inconsistent, particularly if the students who grouped inaccurately in my study tended to make fewer groups. Further research on grouping is required to determine whether more knowledgeable learners *individually* tend

79

to create more groups. It may be that grouping by size also is sensitive to the disposition to lump or split (McKusick, 1969).

Implications for the Learning Progression

Accurate grouping of the 10 objects can be as simple as placing the objects that can be seen with the naked eye in one group, and the objects that are too small to be seen in another. Almost 60% of the middle school students in the sample were unable to group accurately, however. Along with the lower anchor expectation that students may not know of any submacroscopic objects, and the finding that 15% of students interspersed macroscopic and submacroscopic objects while ordering, the lower anchor cannot include any expectations for grouping. Given my finding that older students who group accurately *do not* in general make more groups, the upper anchor can only include the expectation that students group correctly objects that are included in the science curriculum and that are potential landmark objects.

<div align="center">

Relative Scale

</div>

The mean scores on relative scale estimation for four objects are shown by grade in Figure 11. Even undergraduates averaged fewer than 2.5 accurate responses (of 4), and there is no clear trend to improve over successive grades as there is in most other tasks; indeed, the sixth graders performed better than the other middle and high school students, as they had for units. All sixth grade students in this study were participants in the summer camp; their high level of knowledge may be due to their self-selection and may not be typical of all sixth grade students.
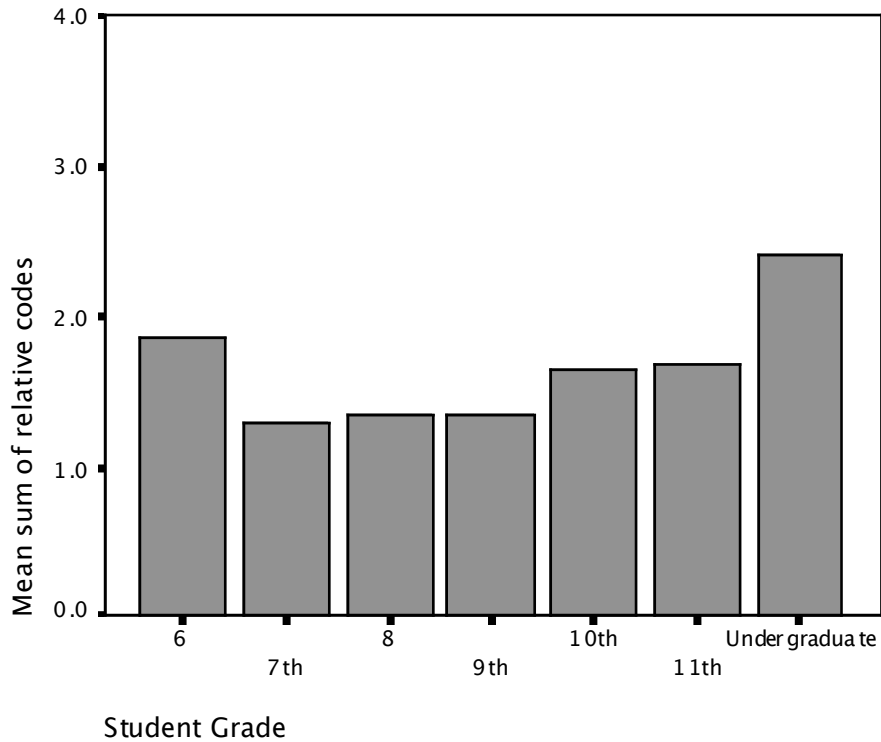
Figure 11: Mean sum of accurate relative scale tasks by grade, out of four possible

Not all objects were equally easy to estimate for students. Table 5 shows that, over the entire sample, the size of red blood cell and human relative to the pinhead were most often correctly estimated, atom and earth the least. This pattern held true for middle and high school students; undergraduates had the most trouble with the size of the atom.

Table 5: *Percent accuracy for relative scale estimation by grade band*

Object for relative scale estimation	Percent correct overall	Percent correct, 6-8th grade	Percent correct, 9-11th grade	Percent correct, undergraduates
Atom	19	22	16	20
Red blood cell	54	58	51	60
Human	55	56	50	100
Earth	27	11	35	60

In Figure 12 I show the distribution of student estimates for the relative scale of the atom. I use a base-10 logarithm scale due to the enormous range of estimates. The reference line indicates the normative value, 7 (the atom is 10^7 times smaller than the pinhead). The mean value is five, showing that students' estimates were (logarithmically) centered on 100,000; thus, students tended to overestimate the size of the atom and underestimate its relative scale compared to the pinhead. The large peaks at 3, 6, and 9 show that students tended to estimate "round" numbers such as 1000, one million, and one billion; 32% of students answered one of these three round numbers. Students were more likely to estimate a number that is close to but not exactly 1000 than a number that is close to but not exactly one million, or one billion; there is a broader spread around three than around six or nine.
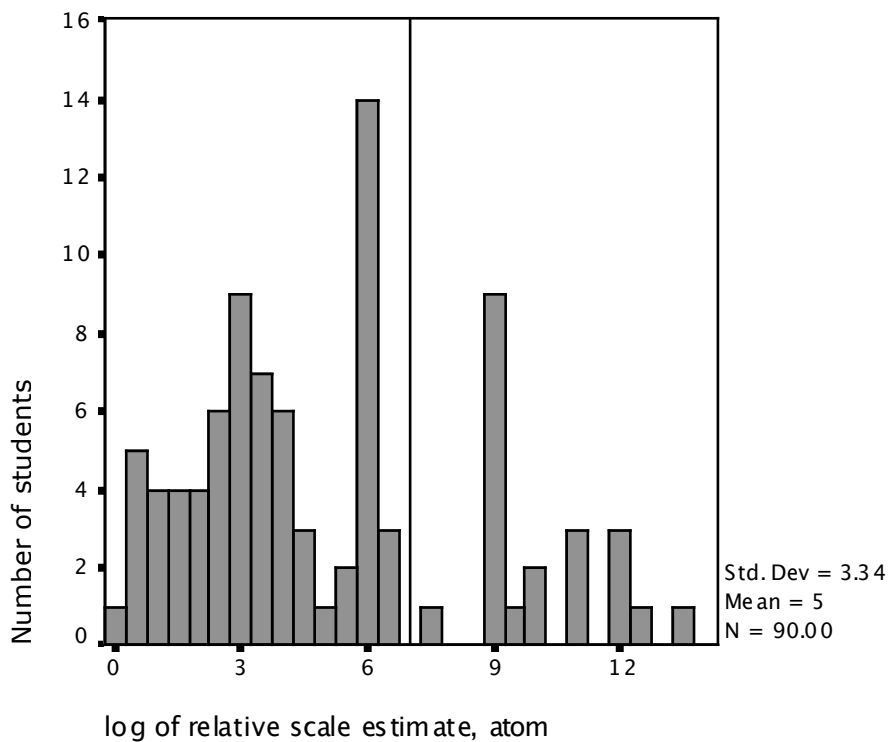
Figure 12: Distribution of base-10 logarithm of responses for relative scale estimation task for atom. The line indicates the normative value.

The corresponding histograms for cell, human, and Earth are included in Appendix H.

82

The estimates for atom and cell were roughly centered on the normative values, but students tended to underestimate the size of the Earth. Student estimates followed the pattern of falling on round numbers.

Discussion

The sizes of objects that are nearer human scale were estimated more accurately; prior research has found the same, for size of objects (Tretter, Jones, & Minogue, 2006) as well as other variables (e.g., age of geologic events, Trend, 2001; Dahl et al., 2005). The finding that students do not have a good idea of the size of the atom (often overestimating its size) is common in the literature (e.g., Brook et al., 1984; Griffiths & Preston, 1992). The findings show that most students do not have an accurate idea of the relative scale of potential landmark objects. Without landmarks, students may not be able to create a mental measurement line (Joram et al., 1998) or "relational web" (Tretter, Jones, Andre, et al., 2006) for size and scale, which will allow them to build a sense of scale for the minute size of molecules (AAAS, 1993).

Implications for the Learning Progression

Although the general trend is to overestimate the size of small objects and underestimate the size of large objects, not all students follow the trend. These findings inform the learning progression by further characterizing the prior knowledge of students: they are likely to have a wide variety of ideas about the size of landmark objects, and may tend to estimate round numbers. Middle school students on average were able to estimate the relative scale of one object from among atom, cell, human, and Earth, compared to the head of a pin, within a factor of 10. The human and the red blood cell were most commonly estimated correctly. For the lower anchor at middle school we cannot assume that students will be able to accurately estimate relative scale for any given object, but they may have reasonable ideas for some objects close to human scale. The low accuracy in pre-college students implies that in the absence of more effective curriculum, we can only include the expectation of accurate relative scale estimation for one or two objects (human, cell) in the upper anchor.

The mean scores on absolute size estimation for objects are shown by grade in Figure 13. There appears to be a weak trend towards greater accuracy with increasing science courses, with the 11th grade dip as seen previously. However, even undergraduates averaged fewer than 4 accurate responses (of 5).
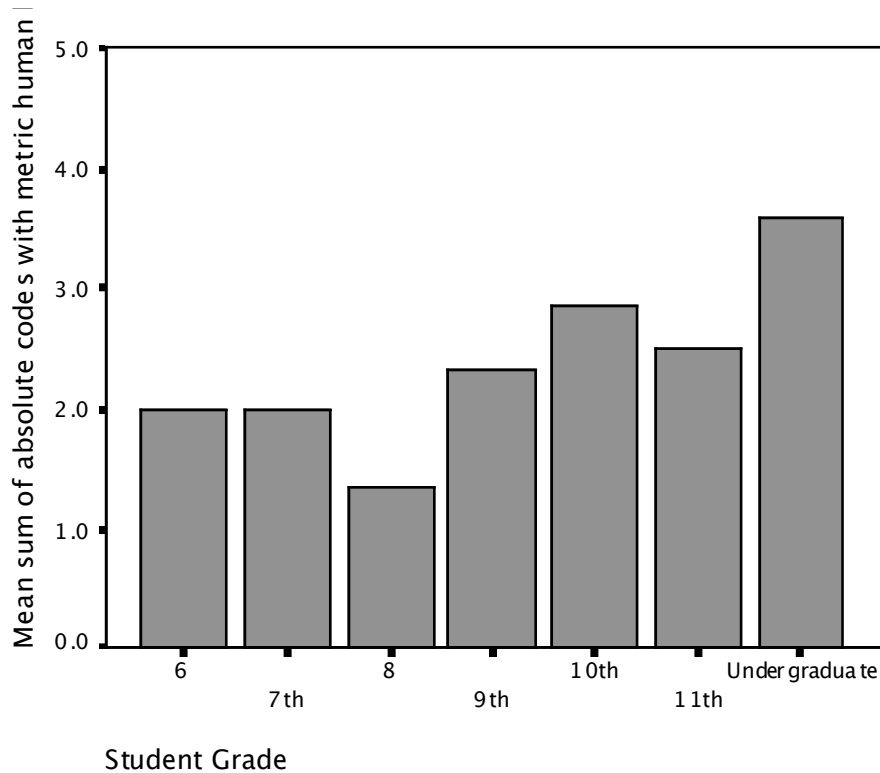
Figure 13: Mean sum of accurate absolute size tasks by science course, out of five possible

Not all objects were equally easy to estimate for students. Table 6 shows that estimating the height of the human (in any units) was near ceiling for all groups, at 87-100%. This high success rate with English units led me to also code for accuracy of estimate of the height of the human in metric units. During the interview, we had prompted students who answered in English units to *also* provide the answer in metric if possible, so the information was available in almost all cases. Success rate with metric was at 60% of the success rate with any units for middle school, 75% for high school, and 100% for undergraduates, bespeaking an increasing familiarity with the metric system as students

progress through the school system. This finding is reasonable, because the use of the English units in everyday life in the United States is near absolute; thus, students' exposure to metric probably comes primarily through science courses in school. Overall, and in every age group, the atom was the most difficult object to estimate, with 15-25% correct. Red blood cell and earth were about equally difficult to estimate, and intermediate in difficulty between the atom and the human. Estimating the height of the human in metric units was the second easiest task, after estimating human in any units.

Table 6: *Accuracy (within 10X) of estimates for absolute size, for atom, red blood cell, human, and earth*

Object for absolute size estimation	Percent correct overall	Percent correct, 6-8th grade	Percent correct, 9-11th grade	Percent correct, undergraduates
Atom	20	15	24	20
Red blood cell	31	22	35	60
Human	92	88	95	100
Earth	31	24	32	80
Human, in metric	65	53	71	100

In Figure 14 I show the distribution of the base-10 logarithm of student estimates for the absolute size of the atom expressed in millimeters. The reference line indicates the normative value, -7 (the atom is 10^{-7} mm in diameter). The mean value is -7, showing that students' estimates were (logarithmically) centered on 10^{-7} mm, the normative value – although there were no responses at exactly this value. However, this mean value is influenced strongly by three responses at large negative values (which correspond to excessively small sizes); thus, students in general tended to overestimate the size of the atom. The peaks at 3, 6, and 9 show the tendency to estimate sizes with "round" numbers such as one thousandth, one millionth, and one billionth of a millimeter. The bars at values greater than 0 represent student estimates *larger* than one mm for the atom, which included ¼ cm, 3 mm, and 12 in.

The histograms showing the distribution of student estimates for cell, human, and Earth are included in Appendix I. The estimates for atom and cell were roughly centered on the

normative values, but students tended to underestimate the size of the Earth, just as they did with their relative scale estimates. Student estimates followed the pattern of falling on round numbers. The distribution of responses for human is tightly clustered around the normative value.
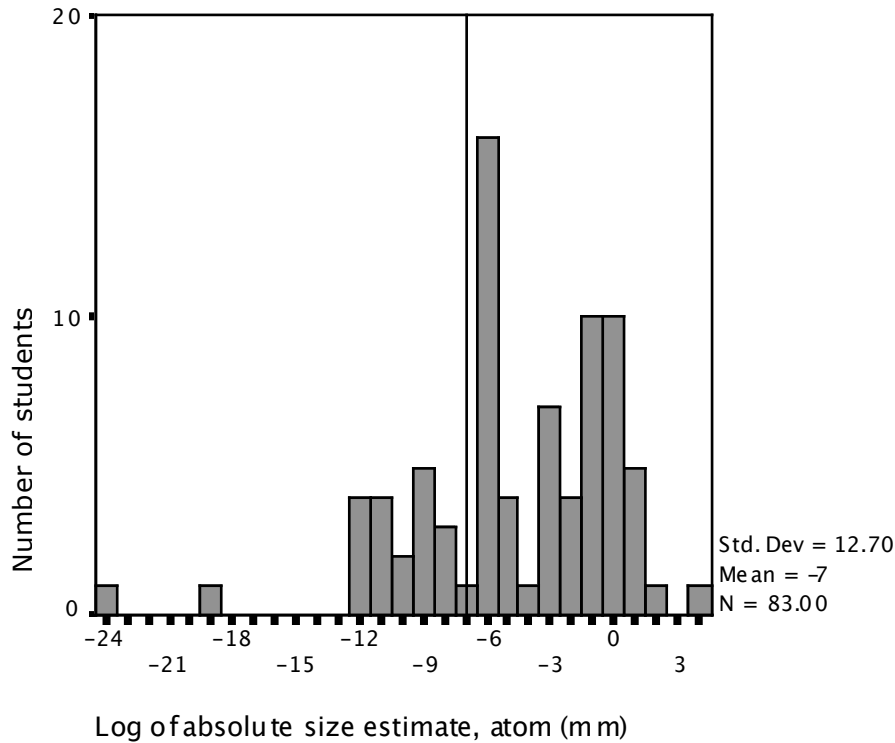
Figure 14: Distribution of base-10 logarithm of responses for absolute size estimation task for atom. The line indicates the normative value.

Discussion

The absolute sizes of objects that are nearer human scale were estimated more accurately, and students again tended to overestimate the size of the atom, as noted above for relative scale. The tightly clustered responses for human again show that this is the fundamental landmark for size (Tretter, Jones, Andre, et al., 2006). The least accurate estimates for atom included small whole numbers and fractions, often ½ and ¼, of millimeter,

86

centimeter, or inch. The fractions ½ and ¼ are the numbers between zero and one that students are usually expected to know at the beginning of their schooling (Post et al., 1993) so it is surprising that these middle and high school students could not come up with more appropriate fractions or decimals. Thus, these very naïve answers may have to do with lack of mathematical content knowledge (of fractions and decimals) as much as with flawed conceptions of the size of the atom. This is further explored in Chapters 5 and 6. Tretter, Jones, and Minogue (2006) also encountered students who interpreted fractions such as 1/100 m as 100 m (12% of middle school students, 5% of gifted high school seniors).

This lack of knowledge about the absolute size of landmark objects can be expected to hinder students' development of an overall understanding of size and scale. It also shows that students are not developing the "particularly important senses of scale" (AAAS, 1993) for the size of molecules (and probably not for the cosmos either, if they do not have an idea of the size of Earth).

Implications for the Learning Progression

Knowledge of the absolute height of a human can be assumed at the beginning of middle school and thus included in the lower anchor. This height is more familiar to pre-university students in English units. Only 53% of middle school students were able to estimate the height of the human in metric units, so this is not included in the lower anchor. Here, too, we can expect a wide range of estimates for any given object (except human, in English units) and overestimates for the size of the atom. The upper anchor can also include absolute height of the human in metric units, and possibly atom or cell. It remains to be seen what students can accomplish after experienced a focused curriculum on size and scale. The instructional activities in the learning progression will thus have to target the construction of submacroscopic landmarks including the atom and cells.

The overestimation of the atom and underestimation of the Earth results in a "compression" of the actual range of sizes of objects. This finding means that an important goal of the learning progression will be to "flesh out" or expand students'

understanding of the range of sizes that exist. Students need to be exposed to appropriate and engaging instructional activities that allow them to construct an idea of the very large relative scale differences between even very small objects.

Overall Factual Knowledge

I calculated a score for overall level of factual knowledge as described in Chapter 3. Figure 15 shows that pre-college students all have a level around one half of the total possible score, and undergraduates at a selective research university are considerably higher. Considering that the criteria for accuracy in relative scale and absolute size estimates is to be within 10 times of the normative value, this figure shows that students in middle and high school do not have a strong knowledge of the size and scale of important scientific objects.
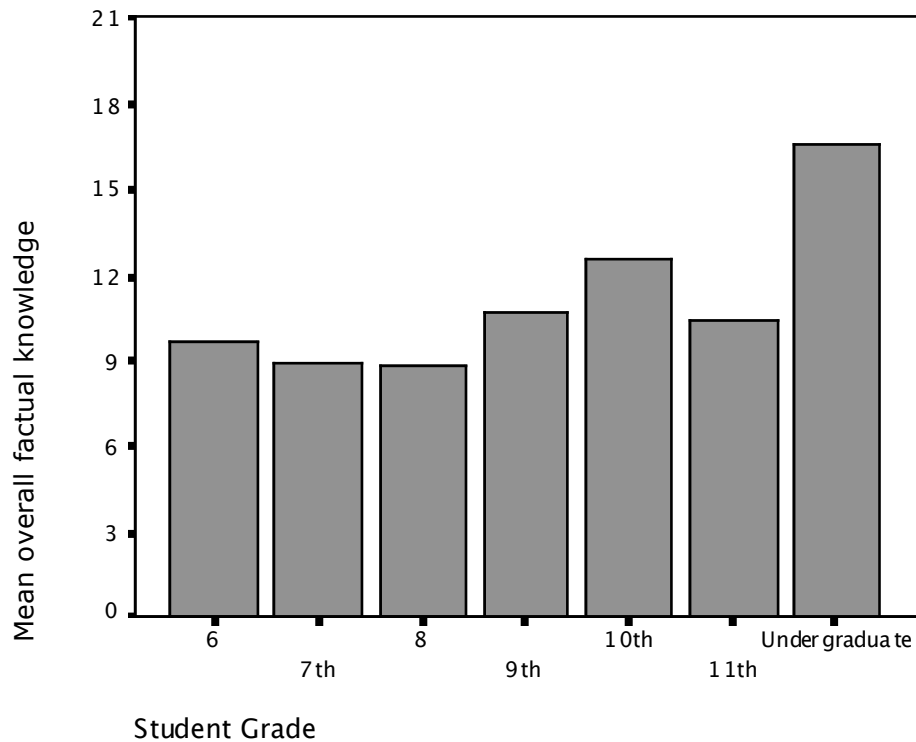
Figure 15: Mean levels of overall factual knowledge

Discussion

There is to my knowledge no measure in the literature to characterize overall factual knowledge of the size and scale of important objects in science. Thus, it is difficult to contextualize these findings beyond saying that current factual knowledge of the size and scale of important scientific objects and units is weak.

Implications for the Learning Progression

A student with solid landmarks of atom, cell, and human (as well as knowledge of the size everyday objects) would be able to respond with atom as the smallest object, nanometer or perhaps angstrom for the unit, order atom smaller than cell and cell smaller than the macroscopic objects, group the submacroscopic objects separately from the macroscopic objects, and generate accurate estimates for relative scale and absolute size for the two submacroscopic landmarks and human. This would result in an overall factual knowledge score of 15, which is better than over 85% of the pre-university students in this study. Thus, a curriculum guided by a learning progression that created strong landmarks would result in important advances in the overall factual knowledge score. More important is what students could do with these landmarks in place: create a mental measurement line (Joram et al., 1998) that would be helpful in understanding myriad scientific concepts that depend partially on size and scale.

Patterns in Factual Knowledge

In this section, I look for patterns in student responses *across* aspects and *within* aspects, thus addressing research question 1B: What patterns of factual knowledge are observed in students? Patterns in student responses may help define qualitatively different levels of understanding for the learning progression, through networks of related understandings. I also compare the difficulty of the different tasks, in order to inform the instructional activities that can be part of the learning progression.

Patterns Across Aspects and/or Tasks

In order to look for patterns in factual knowledge across tasks, I calculated and examined the correlation coefficients between the codes for each factual knowledge task in the

interview (except grouping, due to its non-hierarchical coding scheme). See Table 7. I expected to find strong correlations between student performance on relative and absolute estimation tasks for a given object, since these are logically related (e.g., the relative scale of an atom compared to a pinhead can be calculated if the absolute size of the atom and the pinhead are known). However, the results displayed in the shaded cells in Table 7 show only weak to moderate correlations of between 0.23 and 0.49. In the case of the atom, for example, 75 of 96 students followed the expected pattern, 66 by getting both wrong and nine by getting both right. However, 21 students got one right but the other wrong. A fine-grained analysis looking at consistency between absolute size and relative scale estimates for the same object, at the individual level, is presented later in this chapter.

The only other correlation that was statistically significant and above 0.4 was that between ordering and the smallest object known (0.493, $p < 0.01$). Thus, there were no consistent patterns of factual knowledge across aspects that can be used to define qualitatively different levels of understanding, in this sample. Batt and colleagues (2008) use the smallest object a student can recall to define a "think score" that they relate to a relative scale task (and other tasks related to models). The think score serves to "establish if an individual readily considers submicroscopic objects in their normal thinking" (p. 1144). The correlations displayed in Table 7 reveal that smallest object is not strongly correlated to very many other pieces of knowledge or tasks related to submacroscopic objects; ordering, absolute size of atom, absolute size of cell, and smallest unit known all have a greater number of significant correlations than smallest object. Additionally, responses to the prompt for smallest known object seem to be highly context and format dependent. In sum, the think score defined by smallest known object is not an ideal basis for levels on a learning progression. However, given the lack of extensive significant correlations, none of the other questions in isolation is a promising candidate, either.

Table 7: *Statistically significant Kendall's tau-b or phi intercorrelations between tasks*

	1	2	3	4	5	6	7	8	9	10	11	12
1 Smallest object	--											
2 Smallest unit	.38**	--										
3 Ordering	.49**	.36**	--									
4 Atom relative			.23*	--								
5 Cell relative			.38*		--							
6 Human relative		.27*		.24*		--						
7 Earth relative		.24*				.21*	--					
8 Atom absolute	.20*	.30**	.35**	.33**	.21*			--				
9 Cell absolute	.30*	.26*	.33**	.28**	.49**			.26**	-			
10 Human absolute						.34**				--		
11 Earth absolute	.242**		.31**				.23*				--	
12 Human absolute, metric	.224*		.25*			.43**	.25*			.42**		--

*significant at the p < 0.05 level   ** significant at the p < 0.01 level

I did not find strong correlations across aspects of size and scale to help define levels. It seems that factual knowledge does not develop in coherent networks, with students' current curriculum.

Patterns Within Aspects

As I mentioned in Chapter 3, the coding rubrics for smallest object, unit for smallest

object, and ordering by size already impose hierarchical levels based on theory, and the coding for grouping is non-hierarchical. Thus, I examine only relative scale and absolute size for patterns within aspects.

Relative Scale

Since the cell and the human were the two objects that were easiest for students to estimate accurately, in terms of relative scale, I investigated whether there was a strong pattern where students first learn about the size of these two objects and only then learn about the atom and the earth (see Table 5). I found that there is no clear pattern of development of accuracy for these four objects. Almost 18% of students had a pattern of answers involving accurate estimates for atom, earth, or both *without* also having accurate estimates for both the cell and the human. Thus, it seems that students are learning about the sizes of these four objects in no strict order.

There is to my knowledge no prior research that bears on this question of patterns in growth of accuracy in relative scale estimation. Tretter and colleagues found greater accuracy in relative scale estimation (in terms of body lengths) for sizes closer to humans (Tretter, Jones, & Minogue, 2006), but did not examine how these developed over time. The absence of a pattern means levels cannot be defined on this basis.

Absolute Size

Using the same procedure as I described above for a relative scale, and following the proportion of accurate answers for each object (see Table 6), I found that 27% of students are not consistent with a path where they first learned to accurately estimate the absolute size of the human in any units, then the human in metric, then cell and Earth in indistinct order, and finally the atom.

There is to my knowledge no prior research on the question of patterns in growth of accuracy and absolute size estimation, beyond the general idea that objects closer to human scale are easier to estimate. The absence of a pattern means levels cannot be defined on this basis.

Difficulty of Aspects

Outlining the relative difficulty of the different tasks used in this study can suggest sequencing for the learning progression, by suggesting what may be introduced earlier in the curriculum.

Units and Objects

While most students knew of objects too small to see (91%), very few knew of convenient units with which to express their size (only 25%). A McNemar test showed that the difference is statistically significant ($p < 0.001$).

The finding that students are less familiar with submacroscopic units than submacroscopic objects appears to contrast with Waldron and colleagues' findings that "Respondents of all ages were more successful ordering units of measure [millimeter, micrometer, and nanometer] than in putting 'germ', 'molecule' and 'atom' in correct size order" (2006, p. 573). This discrepancy may be due to the difference between recalling and recognizing objects or units, or because one task involved ordering three objects or units, while the other task only involved one.

The greater familiarity with objects than units means that instructional activities that form part of the learning progression will have to devote efforts to introducing units, possibly linking them to better-known objects such as the cell and atom. The use of units smaller than a millimeter can be circumvented altogether by the use of decimals, fractions, or scientific notation in conjunction with the millimeter or meter. However, student knowledge of these forms of notation is not necessarily solid either. Thus, it will be worthwhile to explore whether students can learn about submacroscopic units given a well-designed curriculum.

Ordering and Grouping

While only 8% of students were able to order all 10 cards correctly, 55% of students were able to group correctly. A McNemar test showed that this difference is statistically significant ($p < 0.001$). Grouping is a far more forgiving task than ordering. For instance, there are 120 different ways of arranging the five submacroscopic objects, only one of

which is coded as correct in ordering. If the student however decides to group them all together, then every one of those 120 permutations will be coded as correct in the grouping task. To my knowledge there is no prior research comparing the relative difficulty of ordering and grouping submacroscopic objects.

Helping students to develop an understanding of different scale "worlds" may be a productive educational strategy. An initial, coarse-grained understanding that there are objects that can be seen with the naked eye, objects that can be seen with an optical microscope, and objects that are too small to be seen with an optical microscope might be a good initial goal for students, later to be complemented with ordering of objects *within* each world (e.g., eukaryotic cells are often larger than one-celled organisms like bacteria; the red blood cell is smaller than most other types of eukaryotic cells).

Relative Scale and Absolute Size

Table 8 shows that overall accuracy on the absolute task was higher than on the relative task for human and lower for the cell ($p < 0.001$), and similar for earth and atom. There was no statistically significant difference between overall success rates for relative scale and absolute sizes estimation tasks, using a t-test and Wilcoxon signed-ranks test.

Table 8: *Accuracy (within 10X) of estimates for absolute size, for atom, red blood cell, human, and earth*

Object	Absolute, percent correct	Relative, percent correct	p-value of McNemar test
Atom	20	19	ns
Red blood cell	31	56	< 0.001
Human	92	55	< 0.001
Earth	31	27	ns

In Figure 16 I compare the distributions of student estimates for relative scale and absolute size for all four objects, using a box and whiskers plot. I display the absolute values for ease of representation. (The values smaller than one for absolute size in millimeters of the submacroscopic objects result in negative numbers when taking the logarithm.) The box represents the two middle quartiles, and the whiskers the bottom and top quartiles of estimates. For clarity, the graph does not display the few values that are

94

very much larger or smaller than the majority of responses. Figure 16 summarizes the distribution of student estimates shown in greater detail in Figures 12 and 14 for atom, and in Appendices H and I for the red blood cell, human, and Earth. This figure reveals that student estimates for the absolute size of an atom displayed very large variability, and for the absolute size of a human very low. For both absolute and relative estimates, objects that are farther from human scale display greater variability. Absolute estimates for atom and cell show greater spread than for relative scale, but absolute size estimates have a smaller spread than relative scale for human, and about the same for earth.

Tretter, Jones, and Minogue (2006) likewise found that respondents coming up with objects of a certain size range expressed in metric units (absolute) or body lengths (relative scale) did not display a consistent difference in accuracy. Tretter, Jones, Andre, and colleagues (2006) found larger standard deviations in the responses of younger students, on size and scale tasks; this trend mirrors the one found in this study, where larger measures of spread are found for tasks that are more difficult.
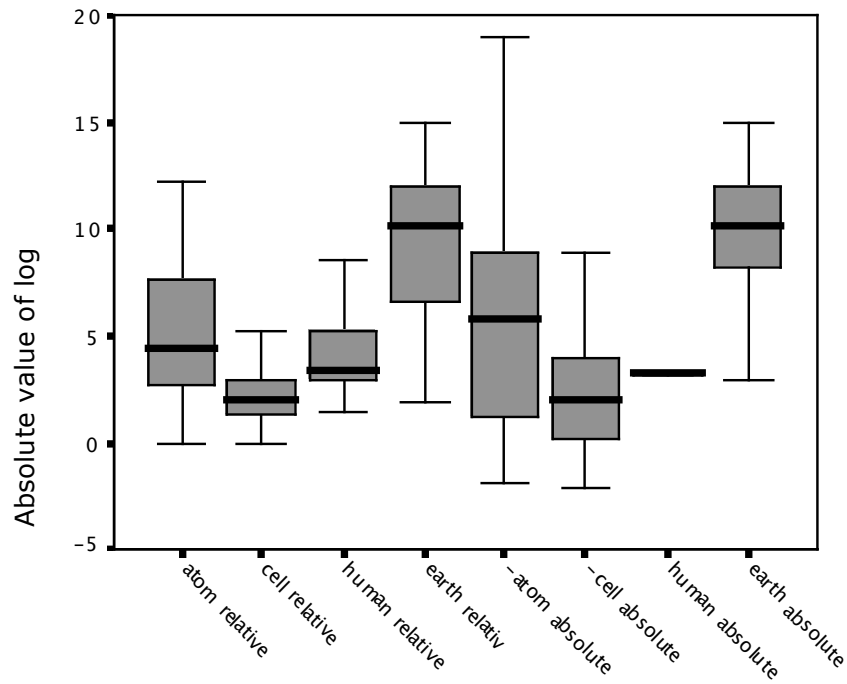
Figure 16: Distribution of student relative and absolute estimates for atom, cell, human, and Earth

The difference in variability of relative scale and absolute size estimates for human exemplify students' tendency not to link the two aspects, but generate separate estimates. Following the strategy I observed in experts (the two university professors) and most undergraduates, of calculating relative scale through known or estimated absolute sizes, a respondent would have identical success rates on the relative and absolute tasks, and the distributions would also be identical. The undergraduates are close to a pattern of consistency, with identical success rates for relative and absolute estimates for three of the four objects (see tables 5 and 6), indicating that they mainly follow this strategy of explicitly linking relative scale and absolute size. The middle and high school students, however, have very different success rates on the relative and absolute tasks, indicating that they are not linking relative scale and absolute size. A close examination of just these strategies and connections is discussed in the section on consistency below.

From these findings it is not clear whether instruction with relative scale or absolute size should come first. These findings again show that the height of the human is a very well established landmark in absolute size, though not in relative scale. Thus, it might be an interesting strategy to begin an exploration of relative scale using the human. On the other hand, students are as successful estimating the relative scale of the red blood cell as of human, and their estimates are clustered more tightly than for human, indicating that studying the relative scale of cells could be productive as well.

Students' Consistency of Knowledge by Year in School

In this section, I address research question 1C: What connections (as inferred from consistency) do students make across aspects of size and scale, by year in school? These data can inform the lower and upper anchors of the learning progression.

As I mentioned in Chapter 3, I code for consistency across ordering and grouping, ordering and relative scale, and ordering and absolute size with zero for absent or one for present; and I code relative-absolute consistency as zero for absent, one for partial, and two for full consistency. This makes a total of five possible points for consistency. The

mean scores are shown in figure 17. There appears to be a weak trend to improve throughout middle and high school (with the 11th grade dip previously mentioned – see Appendix G). Pre-university students are far from fully consistent, ranging from 2 to slightly above 3; the five undergraduates at a selective research university are close to fully consistent.
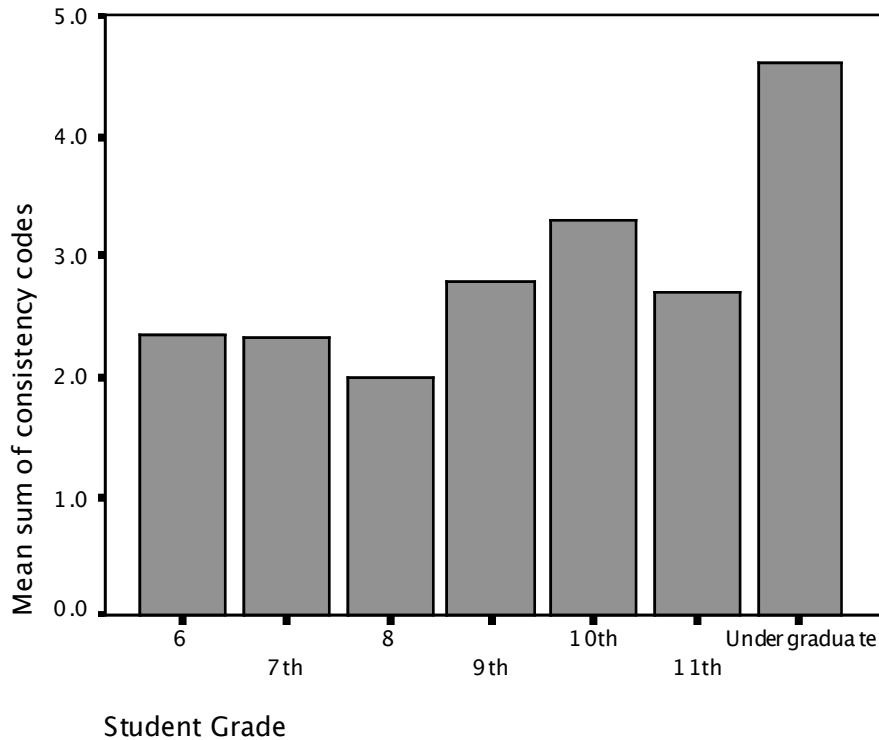
Figure 17: Mean consistency score by grade in school

The clustered bar graph shown in Figure 18 depicts the same pattern in finer-grained detail. There is an increasing amount of dark bars (representing more consistent responses) in students with more years of schooling. Only 1/5 of middle school students and around one-third of high school students, but all five undergraduates, had a sum of consistency codes of four or five, showing that they had at least partial consistency between relative scale and absolute size. Since this relationship is a matter of proportional reasoning, mathematical knowledge is crucial to successfully solving the size and scale tasks posed to students (this is discussed further in Chapter 6).
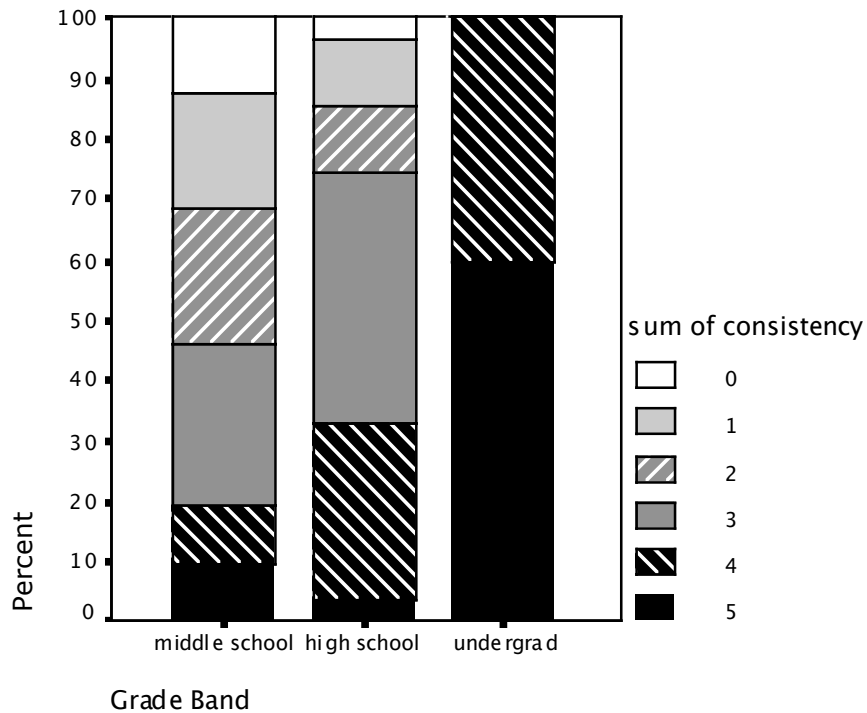
Figure 18: Percentage of students who were at each score for consistency, by grade band

Not all types of consistency were equally difficult for students. Table 9 shows the percentage of students who were consistent, for each type of consistency, overall and by grade band. Order-group consistency was the most common and full relative-absolute consistency the least common, for all grade bands, and overall. This finding is discussed more fully in the section below on patterns in consistency.

Table 9: *Percentage of students who were consistent, for each type of consistency*

Type of consistency	Overall	6-8th grade	9-11th grade	Undergraduates
Order-group	87	81	91	100
Order-relative	85	78	87	100
Order-absolute	70	54	77	100
Relative-absolute, partial or full*	33	22	36	100
Relative-absolute, full	10	10	6	60

* Figures listed are for students who reached *at least* partial consistency for relative-absolute

Discussion

Tretter and colleagues presented respondents with parallel tasks for absolute size and relative scale (using body lengths) where one possible appropriate strategy was to use the same object for both tasks. They found increasing percentages of students following the strategy with increasing age:

> In some cases, participants listed the same object for micrometer and one-millionth body length, and in other cases for nanometer and one-billionth body length. This strategy would be appropriate since there is less than a factor of two difference between a meter and a person's body length, which is an insignificant difference at the scale being requested. The percentages of each group listing common objects at one or the other microscopic scale were: elementary = 8%, middle = 10%, high = 21%, gifted seniors = 11%, and experts = 55%. (Tretter, Jones, & Minogue, 2006, p. 1073)

This is one instance in which that study examines individual knowledge rather than aggregate, and it raises interesting questions. Using the same object for relative scale and absolute size tasks may show that the respondent understands the connection between these two aspects. The proportion of respondents who are consistent and thus are apparently using this connection tends to increase with age, as it does in my study.

Several factors may influence whether students are consistent across relative scale and absolute size of the objects being considered, among them familiarity of the objects. After the end of the interview, I probed one student who thought that "comparing" and "actual size" were not related (#1008). With scaffolding and in the context of everyday objects, the link between relative scale and absolute size was much clearer to him:

> I: If I have an object that's like a fingernail, and it's like, for my thumb, I guess, 2 cm, and the table's length is 180 cm, can you tell me how many times larger the table is than the fingernail?
> R: 90.
> I: And you were able to get this number from here? How did you get 90?
> R: Division…

After this, I returned to the interview task, and the student was able to make an estimate for absolute size for the cell that was now consistent with his relative scale estimate:

I: Now, when you come back and look at this, you know that the pin head is 1 mm, and the cell is one million times smaller. Would that help you figure out the size of the cell?

R: What goes lower than a mm?

I: Micrometer, nanometer, etc.

R: Then probably a micrometer, like you said, or smaller.

I: Could you state the size of the cell as a fraction of a mm?

R: I guess.

I: What fraction would it be, like 1/3, or a trillionth, or a hundredth, what do you think, if it's one million times smaller.

R: A millionth

I: So before you said that these numbers and these numbers weren't related. What do you think now? Do you still think that they're kind of separate?

R: I think they're related now, in some ways.

The task with the table and fingernail has several differences with the interview tasks. First of all, it involves familiar objects and units of measurement. Secondly, it begins with known absolute sizes rather than estimates for relative scale. Third, the "if-then" wording strongly signals that there is a relationship. It is hard to determine which of these factors made this problem more approachable.

My cross-sectional study raises the question of how students learn to link relative scale and absolute size. Asking respondents who did link these two aspects, however, was not particularly illuminating. For instance, student #0099 (in 11th grade) explained her thinking:

I: Did you use these numbers [relative scale] to think about these [absolute size]?...

R: Once you give [an object] a number [of times bigger than the pinhead], it corresponds, it comes out to be the same thing [as absolute size]…

I: Do you think there was a time, when you were smaller, in middle school or elementary, when you didn't know numbers like this had to be connected to numbers like that, or do you think you ALWAYS knew how to do this?

R: Well, probably, every skill is learned, but I can't remember a specific time like learning it.

I: It seems obvious now?

R: Yeah.

This shows that retrospective self-reports as are employed in several prior research studies on size and scale (e.g., Jones & Taylor, 2009) do not get at the heart of conceptual problems in learning. Neither may cross-sectional studies like this one; longitudinal studies or teaching experiments provide stronger evidence for the characterization of learning (Duschl et al., 2007).

Implications for the Learning Progression

Around 81% of middle school students' answers were consistent between ordering and grouping, and 78% between ordering and relative scale. For the consistency between ordering and absolute size, the percentage drops to 54%, and to 10% for consistency between relative scale and absolute size. Assuming that all middle school students will be consistent between ordering and grouping and between ordering and relative scale might severely handicap the 20% of students who actually are not consistent. At least with objects that are too small to see with the naked eye, the lower anchor does not include any expectation of consistency across aspects of size and scale. Given that only one-third of high school students are even partially consistent across relative scale and absolute size, the upper anchor can only include order-group, order-relative, and order-absolute consistency without the development of improved curriculum.

Patterns in Consistency

I conducted an analysis similar to the one described earlier for relative scale and absolute size, looking for a possible developmental pattern in student learning, in order to address research question 1D: What patterns of consistency are observed in students? Using the classification routine, I found that 92% of students in fact are consistent with a developmental path in which the connections across aspects of size and scale (as inferred by consistency across tasks) are made in the order of difficulty indicated above. See Table 10. The 92% of students who fit on this pattern correspond to the 87 students listed in the top section of the table. The students labeled "exception" in table 10 were able to make difficult connections while failing to make easier connections (as inferred by consistency). The student that appears to contradict the progression most strongly, #9007, decided to place each object in its own group, as they were all of distinct sizes in his

opinion. The strategy was coded as being unable to group, and thus was also coded as lacking consistency between grouping and ordering.

Table 10: *Number of students displaying each pattern of consistency*

Consistency Level	Number of Students	Order-Group	Order-Relative	Order-Absolute	Relative-Absolute (Partial)	Relative-Absolute (Full)
5	9	Y	Y	Y	Y	Y
4	21	Y	Y	Y	Y	
3	33	Y	Y	Y		
2	10	Y	Y			
1	10	Y				
0	4					
Total	87					
Exceptions						
#0046, 1001, 8003	3		Y			
#0066, 8004	2		Y	Y		
#9015	1	Y	Y		Y	
#9007	1		Y	Y	Y	Y
#8013	1	Y			Y	
	Total: 8					

Note: The 6 students with missing data are consistent with the progression.

Discussion

The two aspects that are easiest for learners to be consistent across are ordering and grouping; as the two non-quantitative aspects of size and scale, these aspects may be more accessible to learners, and thus easier to connect. The next more difficult connection is that between order and relative scale. Relative scale is a quantitative measure that however does not depend on measurement units, possibly making it easier to connect to ordering than absolute size. The connection between ordering and absolute size is next in difficulty. The final connection to be established is that between relative scale and absolute size, two quantitative measures whose coordination requires proportional reasoning. This cross-sectional analysis strongly suggests that learners develop consistency across aspects of size and scale in a predictable order.

While some researchers have identified proportional reasoning as a prerequisite for students to be able to understand scale (Tretter, Jones, and Minogue, 2006), a more complex picture emerges from this research. It appears that the gradual coordination and connection of different aspects of size and scale, starting with non-quantitative aspects and gradually incorporating numbers and then units, precede and may build up to the form of proportional reasoning required to connect relative scale and absolute size. Other researchers have traced the origins of ratio and proportional reasoning to qualitative ways of thinking, albeit in substantively different ways than the one surfacing in the present study. Resnick and Singer (1993) proposed the theory that

> early abilities to reason nonnumerically about the relations among amounts of physical material provides a child with a set of relational schemas and eventually apply to numerically quantified material and later to numbers as mathematical objects. (p. 109)

Perceptually-based size comparisons and the use of terms such as big and small characterize the "protoquantitative compare" schema in very young children (the other protoquantitative schemas are "increase/decrease" and "part-whole"). The protoquantitative compare schema in my view corresponds to the qualitative aspect of size, ordering. Resnick and Singer (1993) also defined the relation of "fittingness", or "the idea that two things go together because their sizes or amounts are appropriate for one another" (p. 112) - in my view, analogous to grouping. Finally, while these authors examine whether and how children understand how size-ordered series of different objects co-vary (e.g., small, medium, and large bears correspond to small, medium, and large beds), they do not explore how children relate the same objects across ordering and grouping, as the present study does.

Case and colleagues (Case, 1996; Case, Stephenson, Bleiker & Okamoto, 1996) presented the theory that there is a "central conceptual structure" for spatial thought. They propose that children's ability to represent space grows through the gradual integration of two separate structures. One structure allows students to represent familiar objects on a two-dimensional surface, along with their more salient parts, "and the

adjacency and inclusion relations that obtain among them." (Case, 1996, p. 11). The second structure is an object-location schema that represents the location of an object within the scene that it forms part of. While the present study is concerned with students' thinking about objects that cannot be directly perceived and may not be very familiar to students, there are parallels worth mentioning. Adjacency and inclusion seem analogous to ordering and grouping, respectively; and while these two are relative, locating an object within its scene or background is more of an absolute measure. Also worthy of note is the emphasis on the coordination of two structures, similar to my focus on consistency across aspects.

Implications for the Learning Progression

The finding of a strong pattern in consistency has clear implications for the learning progression, since it outlines a long-term process of gradual coordination of successively more complex aspects of size and scale that results in greater conceptual understanding. This finding allows me to characterize six qualitatively distinct, hierarchical levels (named Levels 0-5; see Table 10) of understanding for size and scale, depending on each student's level of consistency. Level 0 is characterized by a lack of consistency across aspects, level 1 by order-group consistency only, and so on, up to level 5 featuring full consistency. As I show in Chapter 6, these levels are strongly and significantly correlated to factual knowledge, more strongly than factual knowledge is correlated to year in school. Therefore, I next generate a first iteration of a learning progression, with levels defined by consistency and with associated descriptions of typical content knowledge at each level.

The First Iteration of a Learning Progression (LP1)

As mentioned earlier, the strongest evidence for a learning progression would come from a longitudinal study or teaching experiment (Duschl et al., 2007). However, even a tentative learning progression generated from a cross-sectional data presented above will be useful in guiding the development of curriculum for a teaching experiment. In this section I present the first iteration of a learning progression (LP1) for one-dimensional

aspects of size and scale. This learning progression is composed of levels defined by consistency and descriptions of typical content knowledge, along with suggestions of instructional experiences presented earlier in this chapter in piecemeal fashion. The second iteration of the learning progression (LP2) presented in Chapter 7 is more sophisticated and includes descriptions of what students are able to do with the knowledge characteristic of each level. It also includes the logical and mathematical knowledge necessary at each stage, as detected by the task analyses and presented in Chapter 6. LP2 also includes instructional activities that were found to be effective in the teaching experiment presented in Chapter 5, or suggested activities, to help students at each level advance along the learning progression.

Next, I describe LP1. I generated bar graphs for mean factual knowledge of the different aspects of size and scale by the level on the progression (see Appendix J). These bar graphs show what factual knowledge is typically associated with each level of consistency. They show that the mean level of factual knowledge monotonically increases with level of consistency (except for the smallest object known, where Level 4 students outperformed Level 5 students). Each level's description thus includes consistency and factual knowledge.

Level 0: the Lower Anchor

Students at this level do not link any aspects of size and scale, as shown by their inconsistency across aspects. Their only size landmark is the height of the human in English units, and they do not know of submacroscopic objects. Thus, they cannot order or group accurately any assortment of objects that includes both macroscopic and submacroscopic objects, tending to treat small macroscopic objects as smaller than submacroscopic objects. They can usually order familiar macroscopic objects correctly, although they overestimate vertical distances relative to horizontal ones (Tretter, Jones, Andre, et al., 2006). They may be unfamiliar with units smaller than inch or centimeter, to the degree that they do not recognize the usefulness of the millimeter in expressing the size of objects smaller than a centimeter or inch.

Level 1: Order-Group Consistency

Students at this level connect ordering and grouping, as shown by their consistency across tasks. This may open the *possibility* of creating "worlds" of different size ranges, since objects whose place in ordering is known cannot go into just any group. Their only size landmark is the human, in English units and possibly also metric. They may know of an object too small to see, often "germs" or cells, but may require prompting to recall the object. They do not know of units of measurement below the millimeter and have little idea of the relative scale or absolute size of submacroscopic objects. Their ordering and grouping reflects their knowledge that a cell is smaller than macroscopic objects, but they tend to place less-familiar submacroscopic objects as larger than small macroscopic objects like the head of a pin or ant.

Level 2: Order-Relative Scale Consistency

In addition to being consistent across ordering and grouping, students at this level are also consistent across ordering and relative scale. This means that even if they often estimate relative scales that are inaccurate, the estimates are larger for objects more distant in size from the reference object. This may open up the *possibility* of establishing coordinated landmarks along a mental measurement line; however, their factual knowledge is still weak. Their only size landmark is the human, in English units and possibly also metric. They may know of cells and atoms but do not have a good idea of their absolute size or relative scale, nor knowledge of units smaller than a millimeter. They may additionally know that atoms are smaller than cells. Students may now have an idea of the relative scale of a cell, human, or both, relative to a pinhead. They still order or group other, less-familiar submacroscopic objects inaccurately. The students at this level tend to know the absolute size of a human in English but still not in metric units.

Level 3: Order-Absolute Size Consistency

In addition to being consistent across ordering and grouping, and ordering and relative scale, students at this level are also consistent across order and absolute size. This means that even though they may generate estimates for absolute sizes that are inaccurate, their estimates for the absolute size of smaller objects are smaller than for larger objects. Landmarks are still limited to the human, in English and now also in metric units.

Students may now have an idea of the relative scale of a cell, human, or both, relative to a pinhead, but cannot relate this "sense" of size to an absolute size. Some students may know of subatomic particles, and most know that atoms are smaller than molecules, viruses, cells, and parts of cells.

Level 4: Partial Relative Scale-Absolute Size Consistency

Students at this level additionally begin to connect relative scale and absolute size, generating consistent estimates for some objects. This allows them to generate relative scale from known absolute sizes, that is, from landmarks. Students at this level may also know the absolute size of the atom, cell, or both, allowing the extension of a student's mental measurement line (Joram et al., 1998) into the submacroscopic world. However, students connect relative scale and absolute size only part of the time. Most students at this level know of subatomic particles, but are still unfamiliar with units smaller than the millimeter. Like level 3 students, they still may not know how the absolute size of cells compares to that of viruses, and this affects their ordering and grouping accuracy. Students at this level may have an idea of the relative scale of atom, cell, and human relative to a pinhead, and/or may calculate these from absolute scale.

Level 5: Full Relative Scale-Absolute Size Consistency

Students at this level connect relative scale and absolute size in all cases. This means that they use proportional reasoning to calculate relative scale from absolute size, or to generate a consistent absolute size from an estimated or known relative scale. Students at this level have cells, atoms, and human as landmarks. Their consistency affords accuracy in relative scale estimation equal to absolute size estimation. They can calculate the size of objects that are not landmarks, if they know or can estimate how many times larger one is than the other. They may still have trouble accurately ordering and grouping due to uncertainty about the size of viruses compared to the cell, parts of cells, or both. Students at level 5 know of subatomic particles (or could, with improved curricula). I believe that Level 5 may be a reasonable expectation for 12th graders that experience improved curriculum, based on my findings from Study One and the literature.

Students' position along the learning progression should initially be assessed in an

interactive manner that allows for probing and is mindful for context, as in the interview protocol used in this dissertation. Not until enough knowledge has been generated about the learning progression that we can characterize students' likely answers at each level will the development of a reliable and valid paper and pencil instrument be feasible.

LP1 guided the development of the curriculum for the teaching experiment described in the next chapter.

<center>Limitations of the Cross-Sectional Study</center>

It is likely that the types of consistency in a given learner would be different had I employed different objects; particularly, they might be more advanced had I used everyday, familiar objects. Earlier in this chapter I provide an example of one student (#1008) who was able to use absolute sizes to calculate relative scale for a table relative to a fingernail, but not in the interview; in fact, he stated the two aspects were not related. However, I surmise that the order in which learners build the connections (as reflected in consistent answers across aspects) is robust, and is determined by the relative difficulty of qualitative vs. quantitative representations of size, and of including units against using only pure numbers.

Another limitation of the study is that the effect of interviewer on student responses was not evaluated. This was due to the absence of a reliable external measure related to students' knowledge of size and scale against which to compare students' performance as a function of interviewer. However, prior to interviewing the interviewers read and discussed literature on interviewing students in science education (Bell, Osborne, & Tasker, 1985; Novak & Gowin, 1984; White & Gunstone, 1992), practiced the interview, discussed it, and mutually observed one another; this, along with a detailed interview protocol with suggested prompts, ensured some level of consistency across interviewers. Additionally, the mean factual knowledge and consistency of knowledge scores for the 54% of students I interviewed as compared to those interviewed by the other interviewers were not significantly different using a t-test (p values of 0.2 and 0.4 respectively).

Despite pilot testing, the interview protocol changed slightly between the first and subsequent rounds of interviewing. For instance, students were not probed for the smallest object of which they knew in early rounds. This may have a small effect on the reported results.

Chapter Summary and General Discussion

I systematically studied students' consistency across aspects, a dimension of student knowledge that can be measured independently of the accuracy of students' factual knowledge (Vosniadou, 2003), and which previous studies on size and scale had detected but not explored (Tretter, Jones, & Minogue, 2006). I hypothesize that consistency reflects mental connections across aspects in learners. Over 90% of students fit along a pathway in which consistency between ordering and grouping develops first, followed by consistency between ordering and relative scale, then ordering and absolute size, and finally relative scale and absolute size. This path goes from the coordination or linking of qualitative aspects, to the coordination of a qualitative aspect with a quantitative aspect - first without units, and then with units, and finally the coordination of two quantitative aspects. Consistency is correlated to factual knowledge, and helps define six hierarchical, distinct levels on the learning progression. The data from the cross-sectional study helped characterize the lower anchor, and tentatively, the upper anchor. I then built LP1 by adding students' typical factual knowledge to each level of consistency.

Students in general have little factual knowledge of the size of key, scientifically important objects in all four aspects of size and scale, and older students know more than younger students. Student estimates of relative scale or absolute size are more accurate and more tightly grouped for objects that are nearer human size. The absolute size of a human is an established landmark for most middle school students, often in English units and less frequently in metric units. The atom and the cell are good candidates for submacroscopic landmark objects for students, as they were often mentioned as the smallest known object. These findings are consistent with or extend prior research (Tretter, Jones, Andre, et al., 2006; Batt et al., 2008; Castellini et al., 2007; Waldron et al., 2006).

In sum, the study of students' consistency across aspects of size and scale has resulted in a plausible account of how students gradually build a more connected conceptual understanding of size and scale. This learning progression includes a description of six qualitatively distinct, hierarchical levels of understanding characterized by patterns of consistency and factual knowledge. This account is compatible with but not identical to theories of early mathematical development, and may be useful not only for science educators and researchers, but also for mathematics education researchers concerned with the development of proportional reasoning. The data from the cross-sectional study helped characterize the lower anchor, and tentatively, the upper anchor. It remains to be seen how much students can progress with a focused curriculum in size and scale based on the first iteration of the learning progression. This learning progression guided the development of a focused curriculum in size and scale for the teaching experiment in my second study, presented in Chapter 5.

CHAPTER 5: RESULTS AND DISCUSSION FOR STUDY TWO

In the previous chapter, I described how the cross-sectional Study One allowed me to build LP1 for size and scale, organized by levels of consistency and supplemented with the typical factual knowledge that students have at each level. In this chapter, I describe how LP1 guided my collaborators and me in developing a 12-hour focused curriculum for size and scale that we implemented during a summer nanoscience camp for middle school students. I then describe the results of the camp, focusing on both the efficacy of camp and implications for LP2.

Study Two is a teaching experiment. It provides evidence to confirm or help revise the learning progression, based on whether students follow LP1. It helps define the upper anchor of the learning progression for the middle school grade band by examining what is feasible for students to learn, given the opportunity to experience carefully designed instructional activities based on the first iteration of the learning progression. If the middle school participants in the teaching experiment end up with levels of factual knowledge or consistency that go beyond those of current high school students, then the teaching experiment will also help define the upper anchor for high school. Study Two also sheds light on how students learn, and on the effectiveness of instructional activities, helping to characterize LP2.

The Teaching Experiment

Curriculum

I led a team that included doctoral students in Science Education; Science, Literacy, and Culture; and Learning Technologies in developing a 12-hour instructional unit for middle

school students. We followed a construct-centered design (CCD) approach (Pellegrino et al., 2008 – described in Chapter 3). We decided to include activities that would be useful for learners at any level in LP1. This effectively meant a sequence of activities that could in theory lead learners from Level 0, the lower anchor, to Level 5, the upper anchor. Of course, we did not actually expect to achieve this advance in any given student with just 12 hours of instruction over 8 days. If this were possible, then we would not require a learning progression. (The length of the curriculum was dictated by the constraints of the 2-week summer camp, which included another curricular strand and student projects.) We made some modifications to our Level 5 learning goals based on the grade levels of our campers; for instance, we decided not to include subatomic particles or molecules.

Thus, the learning goals of the curriculum (though not for all individual students) were for full consistency with accurate knowledge of the size of the atom and the cell as landmarks. These landmarks would allow ordering of atom < cell < macroscopic objects, good relative scale and absolute size estimation for atom and cell, and knowledge of the atom as smallest object.

As I mentioned in Chapter 4 and will analyze more closely in Chapter 6, order-absolute consistency and accurate absolute size estimation of submacroscopic objects requires either a command of fractions, decimals, or scientific notation on the one hand, or units smaller than a millimeter on the other hand. We opted to address this conceptual bottleneck through the use of measurement units for two reasons. First, because it is interesting to investigate whether middle school students can learn about and benefit from learning about such units. The fact that pre-university students in Study One did not know of the micrometer and nanometer does not necessarily imply that they cannot learn them. This is an empirical question that is amenable to being investigated through a teaching experiment. Second, we opted to use measurement units because the context of camp (nanoscience) necessitated explaining the "nano-" prefix in any case.

In Chapter 4 I presented the finding that relative scale estimations of the cell and the human relative to a pinhead are of similar difficulty to pre-university learners. We

decided to stress activities that would help campers learn about the relative scale of cells rather than a human because of the context of the camp. With all of these considerations in mind, we generated the learning goals related to size and scale for the camp:

> Students will know that there are unseen worlds that are too small to be seen with the naked eye. These include the micrometer, nanometer, and sub nanometer worlds. Students will know of several submacroscopic objects, including the atom and the cell, and have an idea of their relative and absolute sizes.

Students' progress towards these learning goals can be measured by the existing interview protocol described in Chapter 3.

Next, we designed a series of instructional activities designed to help students achieve the learning goals. We followed best practices consistent with current learning theory, providing opportunities for students to actively build upon their prior knowledge (guided by LP1), within rich and meaningful contexts, through direct and tool-mediated interaction with their physical and social environment. Some strategies we employed included multimodality (the use of visual, auditory, and kinesthetic modes – see Tang, 2009; Tang & Delgado, in preparation), modeling, and collaborative learning in small groups. Descriptions of the lessons and how they addressed the needs of learners at different levels on the learning progression are shown in Table 11. We looked for existing instructional activities but found that most were not suitable for our learners. Many rely on logarithmic or powers of 10 representations not appropriate for middle school; for instance, the Nanosense Size Matters unit for high school students (Schank, Wise, Stanford & Rosenquist, 2009) and the Powers of Ten video (Jones et al., 2007). The *Benchmarks* (AAAS, 1993) consider that middle school students should be able to "express numbers like 100, 1000, and 1,000,000 as powers of ten" (9A/2, 6-8), but do not include numbers like 1/100 or 1/1,000,000. Similarly, the math standards (NCTM, 1989) expect middle school students to use exponential notation for very large numbers. One promising activity involved a 50,000:1 scale model where the thickness of a dime is the reference object, and is represented as a 50-m long hallway; submacroscopic objects are modeled in 1-D or 2-D (Tretter, 2006). A nanometer is still not visible at this scale, so

we modified and extended this activity to have a 100-m field represent the thickness of a hair at one million to one scale, where a nanometer would be represented as 1 mm and an atom would be the thickness of a hair. We also included 3-D models of proteins, DNA, and viruses, and incorporated hands-on activities in which students measure the objects with a nanometer (a 1-mm thick wire). We also designed a computer simulation that shows successive magnifications of objects like the Cells Alive: How Big Is A…? (Sullivan, n.d.; http://www.cellsalive.com/howbig.htm) online simulation but with additional capabilities and different objects. We designed both of these activities specifically to link the aspects of size and scale. Other activities we designed borrowed from standard classroom activities such as visualizing and sketching one's own cheek cells or swabbing surfaces for bacteria and incubating these on growth medium in Petri dishes. Others were original, as far as we know. Titles and descriptions of these lessons are displayed in Table 11.

Context

The instructional unit was contextualized via the driving question, "How can nanotechnology keep me from getting sick?" and focused on bacteria and viruses. Students learned about a middle school basketball star who died after contracting methicillin-resistant Staphylococcus aureus, a drug-resistant bacterium that has long been a problem in hospitals and has lately been emerging into community spaces like gymnasiums. Students were asked to imagine that they had been asked to form part of a scientific advisory board to help the school administration avoid outbreaks. In particular, they were to evaluate whether to purchase toilets incorporating a nanotechnology that produces ultra smooth finishes, claimed to reduce the buildup of germs. This context was rich and generative because it led to an examination of the size of surface features and of bacteria and viruses, and meaningful because the tragic scenario had involved a student much like them.

Table 11: *Lesson description and alignment with learning progression*

Title	Description	Part of LP1 addressed
S&S 1: What I can't see CAN hurt me	Students learn about dangerous bacteria that can be transferred from surfaces. Students swab various surfaces onto Petri dishes with growth medium to determine what surfaces had bacteria. They model the role of surface roughness in keeping a surface bacteria-free using sandpaper of different grades to represent magnified surfaces, and grains of salt to represent bacteria. Student groups create posters and present their explanation of the role of roughness to the class.	Level 0, 1: learning/reviewing mm, introduction/reinforcement of the submacroscopic world. Provides a rich and meaningful context for students. The need to know the size of bacteria to design or select surfaces is established
S&S 2: How small is small?	Students observe sand, dust, hair under microscope with mm ruler overlaid. They sketch, then estimate the size (in mm). Students try to estimate diameter of a hair by viewing projected slides with 1 and 10 hairs lined up. They use a tracing of the projected slide with 1 hair and the mm to determine the diameter, repeatedly tracing hair across the mm to get ~1/10 mm. Students observe their Petri dishes and determine what surfaces had bacteria.	Level 0, 1: learning/reviewing mm Levels 0, 1: relative scale (e.g., sand is 3X < distance between marks on ruler Levels 0-3: easy fractions (1/10). Levels 3, 4: linking relative scale and absolute size to calculate diameter of hair. Also provides opportunity to go over foundational ideas of measurement.
S&S 3: Measuring the size of skin cells and Staph A	To evaluate nanotoilet claim, students must determine the diameter of bacterium. Students prepare slides of their own cheek cells and a hair, observe relative scale of cell vs. hair (~3X smaller), and calculate diameter of cheek cells (~1/30 mm). Micrometers are introduced as 1000 μm = 1 mm, diameter of hair and Staph A calculated (~100 μm, ~30 μm). Hair, skin cell, and Staph A bacterium are successively projected and traced at 2000X magnification onto large sheets of butcher paper. Students trace skin cell across hair to determine diameter of skin cell, then Staph A bacterium across skin cell and/or hair to determine diameter of bacterium. Students evaluate the nanotoilet's claim.	Level 0, 1: introduction to submacroscopic world Level 1: linking ordering and grouping: larger macro objects belong to different group than smaller submacro objects. Level 2: demonstrating large relative scale differences in the submacroworld. Linking order and relative scale (Staph A < skin cell, and more Staph A fit across hair than skin cells). Level 3: introducing the micrometer; cell, hair as landmark. Absolute size–order link (Staph A < skin cell < hair, sizes = 1 μm, 15 μm, 100 μm). Level 4: linking relative scale and absolute size (calculating absolute size of Staph A from its being 15X smaller than skin cell of 15 μm) Level 5: introducing the μm

S&S 4: Virus – Small Size, Big Threat	Students learn about viruses that can also be transferred from surfaces. They use one computer simulation that automatically lines up (iterates) the smaller object across the larger one to determine the diameter of a virus, and evaluate the nanotoilet's claim. Nanometers are introduced here as 1000 nm = 1 µm. They use this simulation to determine absolute size of objects from hair to atom. A second simulation provides the relative scale of various objects down to atom relative to pinhead. Students evaluate the nanotoilet's claim.	Level 0, 1: introduction to submacroscopic world Level 1: Order-group consistency: ordered objects in simulation belong to successive groups: macroscopic, microscopic, too small for optical microscope. Level 2: demonstrating large relative scale differences in the submacroworld. Linking order and relative scale (atom < virus < pinhead, more atoms fit across pinhead than viruses). Level 3: introducing the nanometer, and linking order and absolute size (atom < DNA < rhinovirus, and sizes are 0.1 nm, 2 nm, 20 nm). Developing atom and cell as landmarks. Level 4: linking relative scale and absolute size (calculating absolute size of atom from its being 20X smaller than DNA, which is 2 nm wide) Level 5: introducing the nm
S&S 5: A million to one	Students interact with a 1,000,000:1 scale model where the thickness of hair is the length of a football field and students the size of bacteria. They use 3-D models of DNA, proteins, and viruses; 2-D model of a bacterium; and 1-D models of cells and thickness of hair to visualize the micro- and nanoworlds at once. They measure the absolute size of the smaller objects using a 1-mm thick wire that represents a nanometer, the medium objects with a small object, and so on.	Level 0, 1: introduction to submacroscopic world Level 1: Order-group consistency: ordered objects belong to successive groups: macro, microscopic, too small to be seen with optical microscope. Level 2: demonstrating large relative scale differences in the submacroworld. Linking order and relative scale (skin cell < cheek cell, more bacteria fit across cheek cell than skin cell). Level 3: introducing the nanometer, and linking order and absolute size (e.g., DNA < cheek cell < hair, and sizes are 2 nm; 30,000 nm; 100,000 nm). Developing atom, cell, hair as landmarks. Level 4: linking relative scale and absolute size (calculating absolute size of hair thickness from its being 100X bigger than Staph A 1 µm) Level 5: introducing the nm

Building Upon Prior Knowledge and Articulation with LP1

Students first explored and measured familiar, small macroscopic objects (e.g., grains of salt and sand, dust, hair) using a millimeter ruler and the low power on an optical microscope. Students then used the thickness of a hair as a reference or scale with which to measure microworld (1-100 μm) objects under the microscope, e.g., their own cheek cells and prepared slides of Staphylococcus aureus. This approach was aimed at helping Level 0 and 1 students, who do not know or do not think of submacroscopic objects. By observing the relative scales and absolute sizes of an ordered sequence of objects, students could construct understanding of the link between ordering and relative scale (Level 2) and between ordering and absolute size (Level 3). By calculating the absolute size of an object from relative scale and the known absolute size of a reference object, students could build their understanding of the connection between relative scale and absolute size (Level 4). At the same time, all students had an opportunity to build the cell as submacroscopic landmark and the thickness of a hair as a landmark object at the boundary of the macroscopic and submacroscopic worlds. Next, the students interacted with even smaller objects at the nanoscale using custom-built computer simulations that we designed. This allowed students to build the atom as a landmark. We followed a similar procedure with units. Students first used short sections of thin, plastic millimeter rulers and the low power of an optical microscope to measure the sizes of grains of salt, grains of sand, and a hair. This provided students at Levels 0 and 1 an opportunity to learn about or review this metric unit. The micrometer was introduced on a need-to-know basis as the students measured the size of their own cheek cells by comparing them to a hair, after the thickness of the hair had been established at 1/10 mm. The nanometer was also introduced on a need-to-know basis when students studied objects smaller than bacteria. These units are useful for students at all levels, but particularly at Level 3 and beyond. Thus, the instructional unit followed the principle of building upon prior knowledge, in this case guided by the learning progression.

Interaction With the Physical World

Since most of the objects of study were too small to observe directly, students used amplification and magnification to study the objects. Student swabbed surfaces and

incubated Petri dishes to amplify the number of bacteria. They were able to visually determine the presence of bacteria on some surfaces in this way. They observed objects under the microscope, preparing the slides themselves in some cases. Students also observed videos concerning bacteria and viruses.

Interaction With the Social World

Students worked in small groups, and the instructors (another science education doctoral student and I) strove to foment student-centered, "interanimated" (Scott, Mortimer, & Aguiar, 2006) discourse patterns in the classroom, with the aim of scaffolding students' social construction of knowledge. Students communicated their findings to other groups, and to the community (campers, parents, and friends) during the final session.

Tools

Students used Petri dishes and incubators, optical microscopes, computer simulations, and observed real-time images from a projecting optical microscope that provided a greater magnification than their own optical microscopes.

Models

Students modeled the effect of surface roughness on bacterial adhesion using different grades of sandpaper to represent magnified surfaces, and grains of salt to represent bacteria (see Tang, 2009; Tang & Delgado, in preparation). We used a projecting optical microscope to trace on butcher paper the outlines of a hair, a skin cell, and a Staphylococcus A bacterium all at ~2000X magnification. Students interacted with a computer simulation that modeled successive magnifications by steps of 10X up to ten million, and portrayed objects from a hair down to an atom. These macro-to-nano activities were followed by a lesson in which students went from nano to macro, using two- and three-dimensional million-to-one scale models of objects from DNA (2 nm) to Staph A (1 μm), and one-dimensional representations of the size of various types of cells and the thickness of a hair (the length of a football field, at this magnification).

Multimodality

In addition to the kinesthetic and visual modes of learning described in the activities

above, students worked with a simulation that used visual and auditory modes as well as time to represent the size of submacroscopic objects (Song & Quintana, 2009). Students engaged with and created visual representations of one, two, and three dimensions.

Establishing Consistency Across Aspects

We did not engage in direct instruction about the four aspects of size and scale or how they were related (mainly due to time constraints). However, students engaged in activities that illustrated or relied upon these connections. For instance, the million to one scale activity displayed objects from the atom to the thickness of the hair in a single representation on the football field. Students had observed some of the objects under the optical microscope, but some were too small. This was an opportunity to link ordering (the objects on the football field) to grouping (microworld objects that can be seen under the optical microscope, and nano world objects that are too small). We used a projecting optical microscope to trace on butcher paper the outlines of a hair, a skin cell, and a Staphylococcus A bacterium all at ~2000X magnification. Students then determined the relative scales by repeatedly tracing the skin cell across the hair, and the Staph A bacterium across the skin cell. In this activity, students could observe that the sizes of the skin cell and the hair in terms of the bacterium were related to the order of sizes, with a larger number of bacteria fitting across the larger object; this activity links relative scale to ordering. In the million to one scale activity, students measured the sizes of the smaller objects using a 1-mm diameter wire that represented one nanometer, and used those objects to measure the next larger objects. This activity relied on the connection between relative scale and absolute size, and also resulted in students recording larger and larger sizes for objects that were ordered closer to the large end (thus linking absolute size and ordering).

Articulation With Learning Goals and LP1

Students were introduced to objects too small to see with the unaided eye, and they had experiences that directly linked the unseen world to some of the smallest visible objects. They were introduced to the micrometer and nanometer, and experienced the microworld through the optical microscope and the nanoworld through computer simulations and physical models. Throughout the activities, students learned about several

submacroscopic objects, including the atom and the cell. They had various opportunities to learn about the sizes of these objects in both relative and absolute terms, and to explore the connection of size across aspects. See Table 11 for a description of how specific activities addressed students at different levels on the learning progression.

Enactment

We enacted the instructional sequence during a free, two-week summer science camp for 32 students who had finished 6th through 8thgrades, as described in Chapter 3. The camp was held at a research university close to the school district. There were 19 sixth graders in one group, 13 seventh and eighth graders in the other. The two main instructors for the size and scale strand were a PhD student in Science Education and me; we had both taught during the previous year's camp, were involved in the development of the curriculum, and tested several of the instructional activities with a small group of middle school students prior to the actual camp. Two additional co-teachers assisted with instruction and facilitated small-group discussions. Lessons included many small-group and whole-group discussions, using student-centered discussion formats. The six two-hour lessons were enacted during the first six days of camp. A second strand dealt with size-dependent properties, and occasionally used or reinforced the concepts from the size and scale strand. Other activities included field trips to a clean room and 3-D modeling lab, and a small-group project.

Findings

Students' Development of Factual Knowledge

Next, I describe the learning gains and final state of factual knowledge for the students who experienced a focused curriculum on size and scale, thus addressing research question 2A: What is students' factual knowledge of the size of important objects in science, after a focused curricular experience on size and scale?

Smallest Object

Efficacy of camp.

Practically all campers were able to mention the atom as the smallest object at the end of camp. The average code for smallest object for the summer campers improved from 1.38 (SD = 0.71) to1.87 (SD = 0.54). The difference is statistically significant, as shown by a McNemar-Bowker test (p < 0.001). As shown in Table 12, this mean of 1.87 was composed of 20 responses of atom, one subatomic, two cell, and one macroscopic.

Table 12: *Responses for smallest object pre- and post-camp (middle school)*

Code	Object	Pre-Camp	Post-Camp
0	Macroscopic	2	1
1	Cell/microorganism	12	2
2	Atom	9	20
3	Sub-atomic particle	1	1

In order to contextualize these gains, I compared the pre- and post-camp mean codes for the campers to the mean codes by grade for students in Study One who *had not* experienced a focused curriculum for size and scale, in Figure 19. This figure shows that after camp, the students' mean code for smallest object is intermediate between that of middle school students and that of high school students in Study One who had not experienced a focused curriculum. Since there was no instruction on subatomic particles, this is still a positive outcome, near ceiling (recall that the atom is coded two). Many studies have shown that students do not know much about the size of an atom and may confuse it with the size of a cell (e.g., Brook et al., 1984; Griffiths & Preston, 1992; Flores, 2003; Harrison & Treagust, 1996). The interdisciplinary approach of our curriculum, integrating the study of objects ranging in size from the atom through small macroscopic objects, may have aided students' differentiation of the atom and the cell. Disciplinary boundaries between biology and chemistry in traditional school curricula may be hindering this differentiation by not providing opportunities for students to consider the sizes of both objects at the same time.
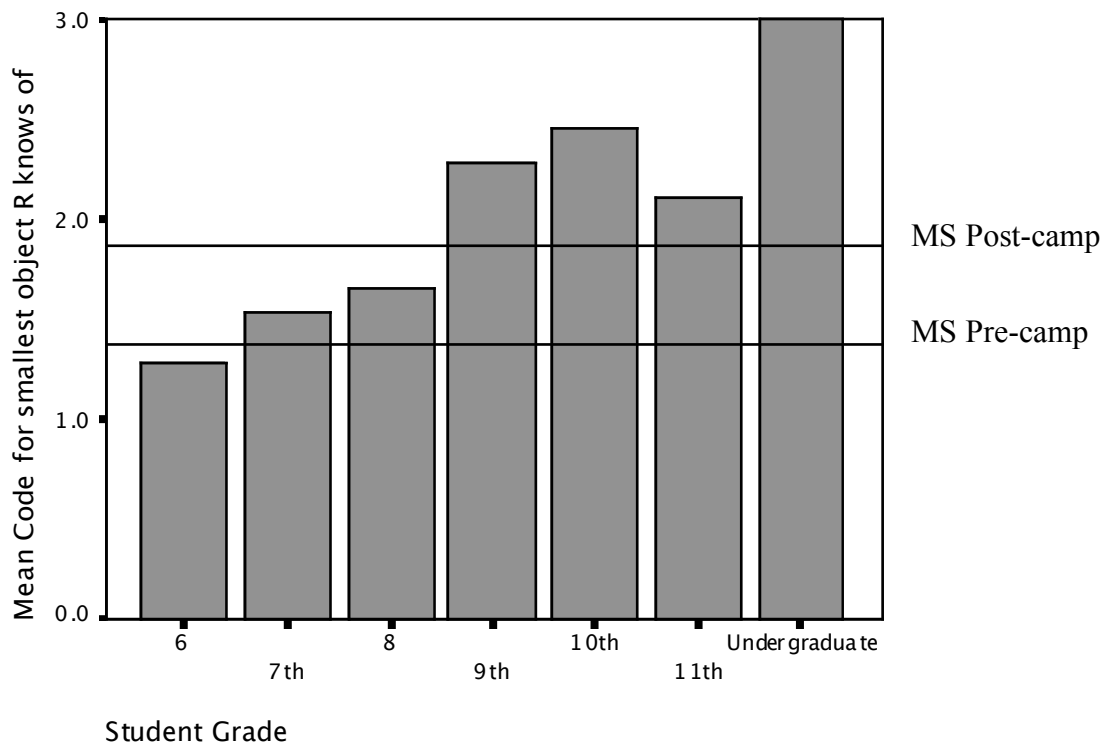
Figure 19: Pre-and post-camp mean score for smallest object of which respondent can think (lines), compared to mean score by grade for non-campers (bars)

Implications for the learning progression.

My findings indicate that it may be reasonable to expect middle school students to understand that the atom is very small, smaller than microscopic objects like a cell or macroscopic objects like a grain of salt, and begin to differentiate atom and cell, if they experience adequate instructional activities. Given that the five undergraduates knew of subatomic particles, and that over half (57%) of high school students who had taken or were taking Chemistry responded with a subatomic particle, it is a reasonable expectation for students to know of subatomic particles by the end of high school. This finding also supports the idea that the atom and cell can become landmark objects for students. Students at all levels on the learning progression were able to learn about atoms' tiny size relative to other objects (in qualitative terms, at least).

Smallest Unit

Efficacy of camp.

Practically all campers were able to mention submacroscopic units at the end of camp. The average code for the unit for smallest object (or smallest unit known) for the summer campers improved from 1.13 (SD = 0.45) to1.92 (SD = 0.28). The difference is statistically significant as shown by a McNemar test (p < 0.001; the sole code of 0 was recoded to 1). This mean of 1.92 was composed of 20 responses of nanometer, one each of micrometer and picometer, and 2 of millimeter – see Table 13.

Table 13: *Responses for smallest unit pre- and post-camp (middle school)*

Code	Unit	Pre-Camp	Post-Camp
0	Do not know	1	0
1	Macroscopic (e.g., mm)	19	2
2	Submacroscopic (e.g., nm)	4	22

In order to contextualize these gains, I compare the pre- and post-camp mean codes for the campers to the mean codes by grade for students in Study One who had not experienced a focused curriculum for size and scale, in Figure 20. This figure shows that campers outperformed middle and high school students in Study One who had not experienced a focused curriculum, and performed almost at the level of the undergraduates. It must be noted that the campers started out with a high level of knowledge.
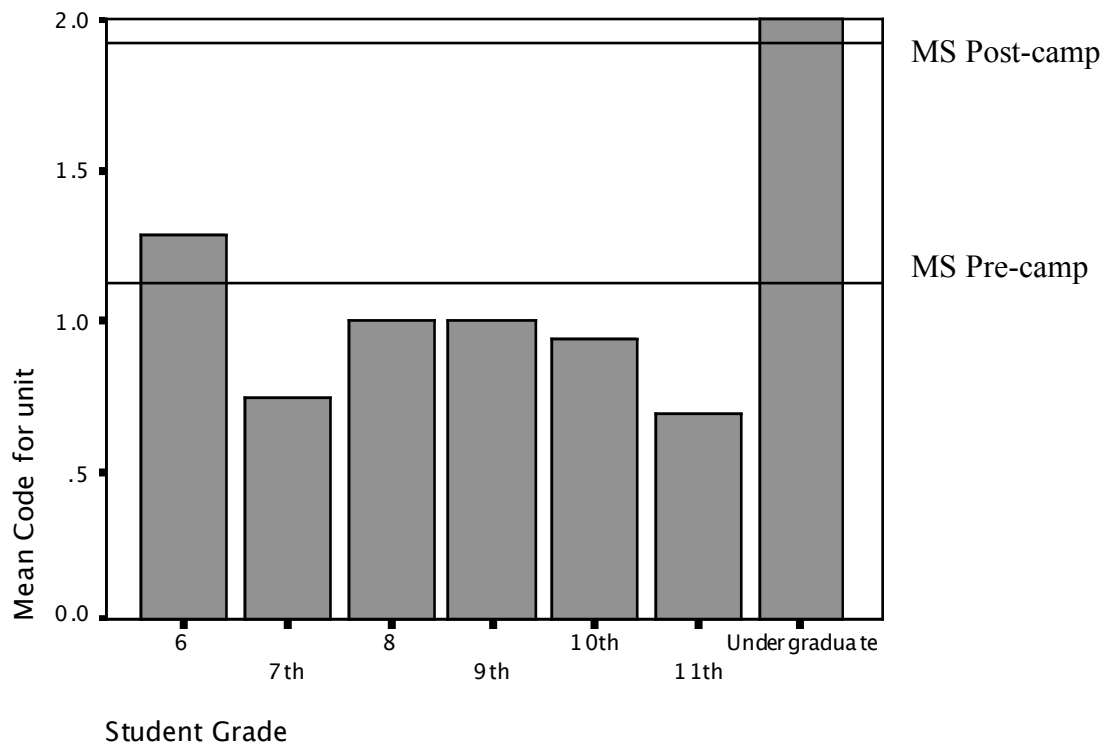
Figure 20: Pre-and post-camp mean score for smallest unit of which respondent can think (lines), compared to mean score by grade for non-campers (bars)

Implications for the learning progression.

Most students were able to learn about atoms' tiny size relative to other objects. Given that middle school students who experienced a focused intervention on size and scale were able to learn about the nanometer, it is reasonable to expect high school students to know about these units of measurement as well; thus, the upper anchor can include knowledge of micrometers and nanometers. This finding also implies that early instruction about units may be a feasible strategy to help students comprehend the size of the cell and the atom, developing these as landmarks, and helping to build consistency between ordering and absolute size.

Ordering

Efficacy of camp.

A t-test showed that there is a statistically significant increase in the mean code for

124

ordering 10 objects for the summer campers, from 2.04 (SD = 1.33) to 2.63 (SD = 1.69), $p < 0.05$. Table 14 shows that seven students were at code 0 or 1 pre-and post-camp; five of these students stayed at these low codes, while two increased to a code of 4 and one decreased (from a code of 2 to a code of 0).

Table 14: *Responses for ordering 10 objects pre- and post-camp (middle school)*

Code	Ordering	Pre-Camp	Post-Camp
0	Macroscopic objects interspersed with submacro	5	6
1	Macro OK, but cell < atom	2	0
2	Macro OK, cell > atom, but atom not smallest	6	1
3	Macro OK, atom smallest, but cell not largest submacroscopic object	10	9
4	Macro OK, atom smallest, cell largest submacroscopic object, but molecule/virus/ mitochondria wrong, or cannot justify correct order	0	6
5	All correct with justification of ordering	1	2

In order to contextualize these gains, I compare the pre- and post-camp mean codes for the campers to the mean codes by grade for students who *had not* experienced a focused curriculum for size and scale, in Figure 21. This figure shows that campers performed at a level roughly intermediate between middle and high school students in Study One who had not experienced a focused curriculum.

Implications for the learning progression.

Four of the six students whose ordering was coded zero after camp were at Level 0 or 1 on LP1 before starting camp. The other two students did not conform to LP1, but had only two types of consistency prior to camp. This is one instance of correlated results across consistency of knowledge and accuracy of factual knowledge that resulted in a strong and statistically significant correlation discussed further in Chapter 6. This correlation suggests that connectedness of knowledge may play a role in students' meaningful accumulation of factual knowledge, although of course other possible causes would need to be ruled out before attributing causality to this relationship. Two other students who were at Level 1 prior to camp did however improve on their ordering, from
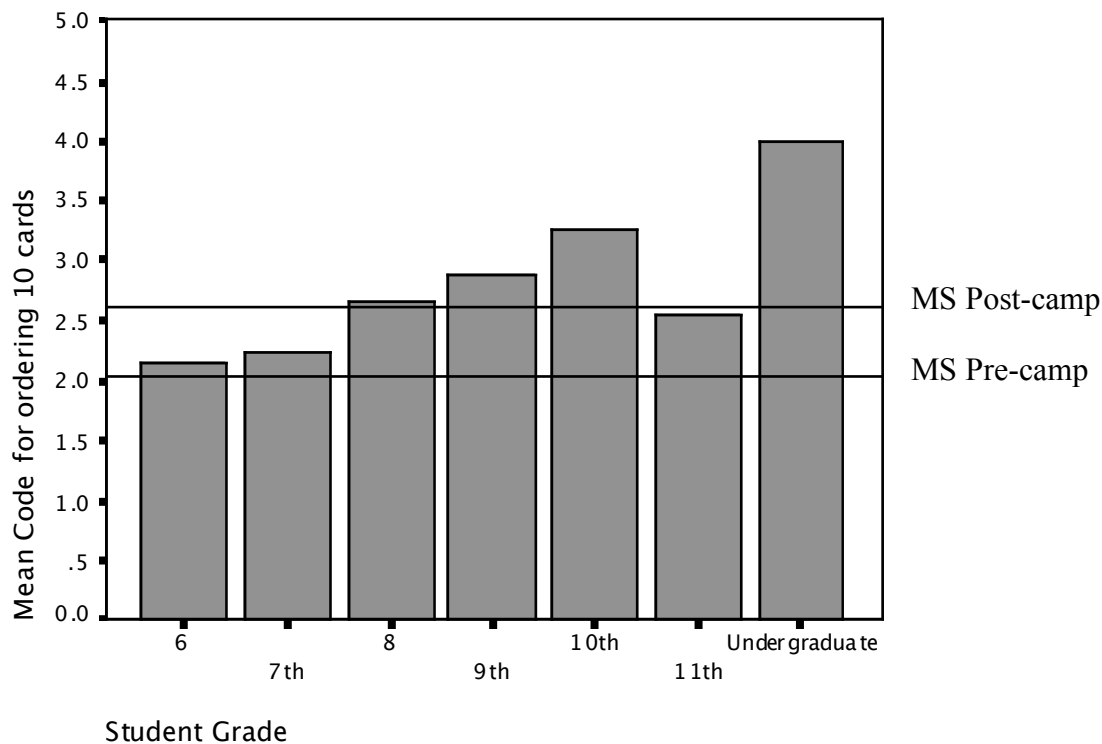
Figure 21: Pre-and post-camp mean score for ordering 10 objects (lines), compared to mean score by grade for non-campers (bars)

1 to 4 and from 3 to 4. Thus, starting out at a low level did not preclude improved ordering. The student who regressed ranked the cell larger than the pinhead at the end of camp, though not at the beginning. I do not have an explanation for this student's lower performance on ordering after the camp.

These results show that students at the end of middle school can be expected to have enough of an understanding of the size of atoms and cells to rank these correctly relative to each other and to macroscopic objects - *if* they experience an appropriate curriculum that helps students develop their consistency of knowledge. Thus, adding this expectation to the upper anchor is contingent on consistency, discussed further below.

Grouping

Efficacy of camp.

The coding scheme for grouping is not hierarchical, except for the difference between correct and inaccurate grouping. The percentage of summer campers grouping the 10 objects correctly did not increase significantly. Their specific strategies did not change very much, as shown in Table 15. I also coded for accuracy of grouping without the two submacroscopic objects not studied in camp (mitochondria and molecule), and the increase in accuracy was not statistically significant according to the McNemar test. As I will present in greater detail in Chapter 6, there is a tendency for students to group unfamiliar objects separately, and this might be a factor behind these results, as the grouping task included objects that may have been unfamiliar to the students.

Table 15: *Responses for smallest object pre- and post-camp (middle school)*

Code	Grouping	Pre-Camp	Post-Camp
0	Incorrect	14	12
1	Mixed		1
2	Two correct groups	1	1
3	Three correct groups	2	2
4	Four correct groups	2	3
5	5-9 correct groups	5	4

Implications for the learning progression.

This study does not shed light on what is feasible to expect from students at the end of high school, in terms of grouping important objects in science. However, it does inform the assessment of student understanding, which is also part of the learning progression. Including objects that are unfamiliar to students in a grouping task may affect the grouping accuracy even for the objects that are more familiar. Thus, assessment will either have to be closely tied to instruction, or be interpreted in light of these findings.

Relative Scale

Efficacy of camp.

Mean performance in the estimation of relative scale, summed over the four objects (atom, red blood cell, human, Earth) did not improve in a statistically significant way, as shown by a t-test. Students did not have opportunity to work with the relative scale of the macroscopic objects, human and earth, so I did not expect improvement for those objects. However, the change in students' estimates for the relative scale of the red blood cell, increasing from 61% to 78% accurate, had a one-tailed p-value of 0.06 using a McNemar test. Students had more opportunities to think about the relative scale of cells compared to small macroscopic objects (e.g., thickness of a hair) during the camp than about the relative scale of the atom; the many orders of magnitude in relative scale difference between small macroscopic objects and the atom precluded presenting them together on a single screen in the computer simulation, or viewing them together under the microscope. Students were able to experience representations of atom and hair in a single representation, simultaneously, only in the million to one scale football field activity. At this scale, an atom is the thickness of a hair, so the size was mentioned but not represented by an object (as it would be too small to easily see). A computer simulation (Song & Quintana, 2009) that the students used showed how many of a given object could be lined up across a pinhead, and one of the objects was the atom. This simulation correlated size to time (objects were placed across the pinhead at a constant rate). Most students focused on the amount of time required rather than the total number of atoms. Thus, it is not surprising that the only change approaching significance is in the campers' relative estimates of cells. Conceivably, it might be more productive to engage students in thinking about the relative scale difference between atoms and cells, and then between cells and macroscopic objects, rather than between atoms and objects large enough to see in a single step. This suggestion is included in the *Benchmarks* (AAAS, 1993): "A million becomes meaningful, however, as a thousand thousands, once a thousand becomes comprehensible." (p. 276).

It is worth noting that schooling seems to have little effect on the accuracy of students'

ideas of relative scale for submacroscopic objects, as can be seen by the relatively unchanging percentages correct for atom (16-23%) and red blood cell (52-61%) at the different grade bands in the absence of a focused curriculum (see Table 5 above, in Chapter 4). Thus, the improvement in accuracy for the red blood cell at a one-tailed p-value of 0.06 is encouraging, even if statistical significance is not quite reached at the 95% confidence level. The students' 78% accuracy in estimating the relative scale of the red blood cell after camp is considerably higher than even the undergraduates in the sample from Study One. See Figure 22.

Figure 22: Pre-and post-camp mean score for accuracy of estimates of relative scale of cell (lines), compared to mean score by grade for non-campers (bars)

Implications for the learning progression.

When provided with appropriate opportunities to learn, middle school students appear to be able to understand the relative scale of cells compared to a small macroscopic object, but have greater difficulty with atoms. As I will explain below, however, students in the summer camp *did* greatly increase their accuracy in estimating the absolute size of the atom. If these students also were able to link relative scale and absolute size, then they

could *calculate* the relative scale of the atom. Thus, even if students do not directly remember or have a feel for the relative scale of the atom compared to a pinhead, they would still have access to calculating it. While the study shows that expecting students to have an understanding of the relative scale of the cell is feasible to include in the upper anchor, we do not yet know how feasible it is to expect the same for the atom.

Absolute Size

Efficacy of camp.

The average sum of the codes for estimates of the absolute size of the four objects (including human in any units and in metric) for the summer campers improved from 1.87 (SD = 1.1) to 2.96 (SD = 1.16) (out of 5 possible) for an effect size of 0.95 (p = 0.001 for t-test for difference in means). See Table 16.

Table 16: *Statistically significant changes in mean accuracy of absolute size estimation, from pre- to post-camp*

Object	Pre-camp	Post-camp	Test and p value
Human (metric)	61% accuracy	87% accuracy	McNemar *
Atom	8% accuracy	58% accuracy	McNemar ***
Red blood cell	17% accuracy	46% accuracy	McNemar *
All objects (including human in metric)	1.87 mean score of 5 possible	2.96 mean score of 5 possible	t-test *** ES = 0.95

$* p < 0.05$ $***p \leq 0.001$ (t-test comparing means)

In order to contextualize these gains, I compare the pre- and post-camp mean sum of codes (for the four objects, including metric human) for the campers to the mean sum of codes by grade for students in Study One who had not experienced a focused curriculum for size and scale, in Figure 23. This figure shows that campers outperformed middle and high school students in Study One who had not experienced a focused curriculum; however, the campers did not reach the level of the undergraduates.

When analyzing by individual object using a McNemar test, the pre-post changes for
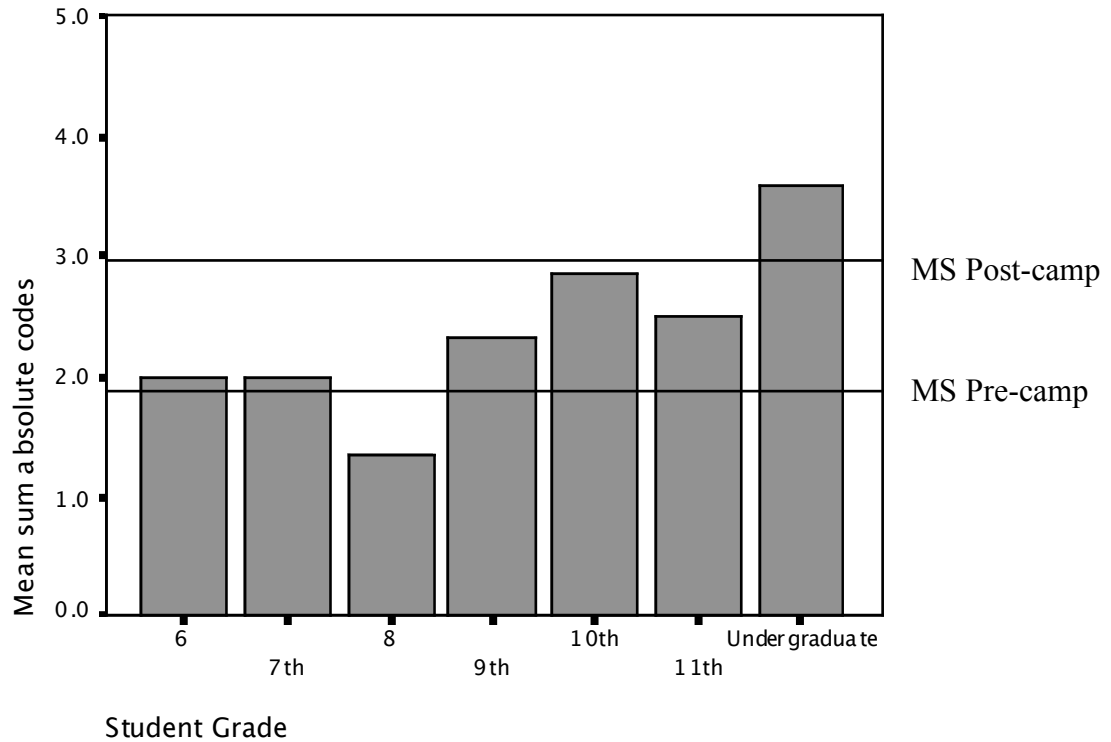
Figure 23: Pre-and post-camp mean score for accuracy of estimates of absolute size of four objects (lines), compared to mean score by grade for non-campers (bars)

human in any units and earth were not statistically significant, which is not surprising as they were not the focus of the instructional intervention. However, the mean accuracy for estimates for the absolute size (height) of the human in metric units improved significantly ($p < 0.05$) from 61% to 87%. This increase may have been caused by our use of the metric system in camp; as campers used millimeters to describe the size of small macroscopic objects we reminded them of the relationship between the millimeter and the meter. The accuracy for atom increased from 8% to 58% ($p < 0.001$), and the accuracy for red blood cell increased from 17% to 46% accurate for the cell ($p < 0.05$).

The campers' level of performance for cell is in between high school and undergraduates, and far better than any age group for atom. This is depicted graphically in Figures 24 through 26.
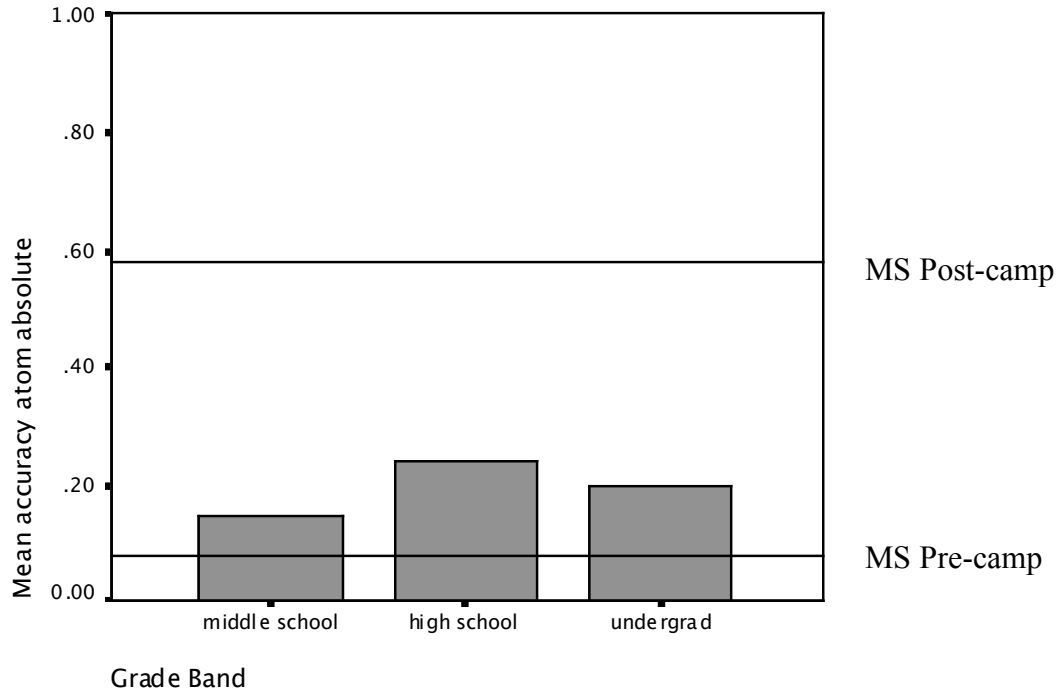
Figure 24: Pre-and post-camp mean score for accuracy of estimates of absolute size of atom (lines), compared to mean score by grade for non-campers (bars)
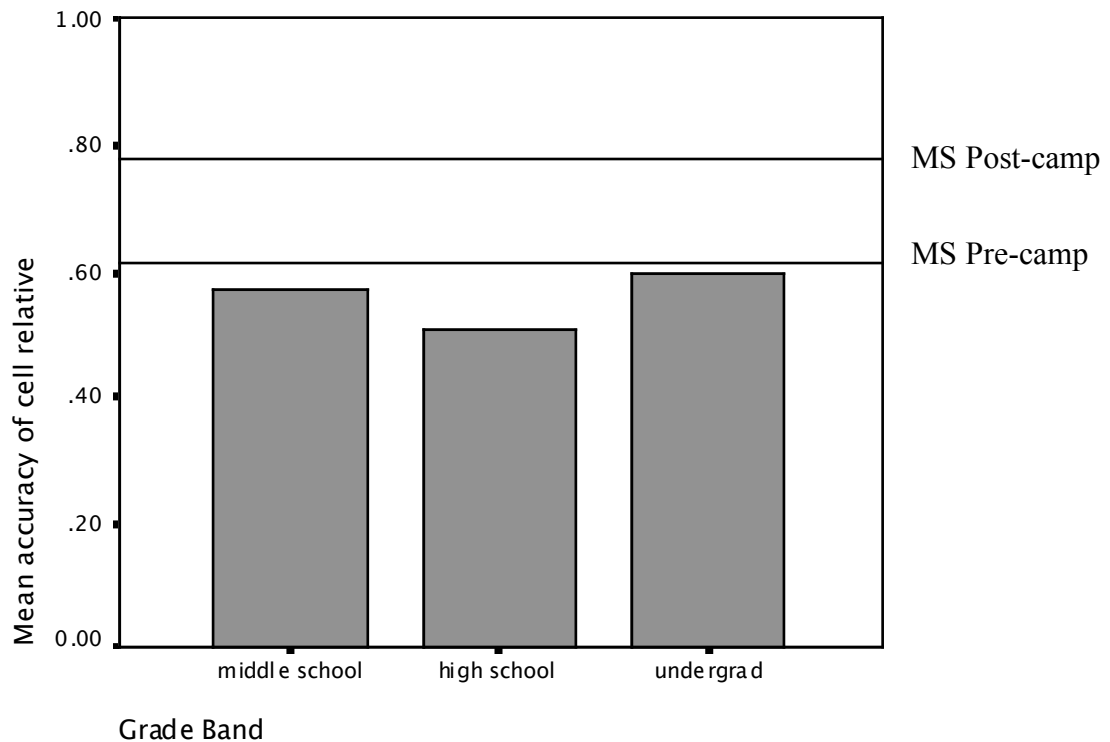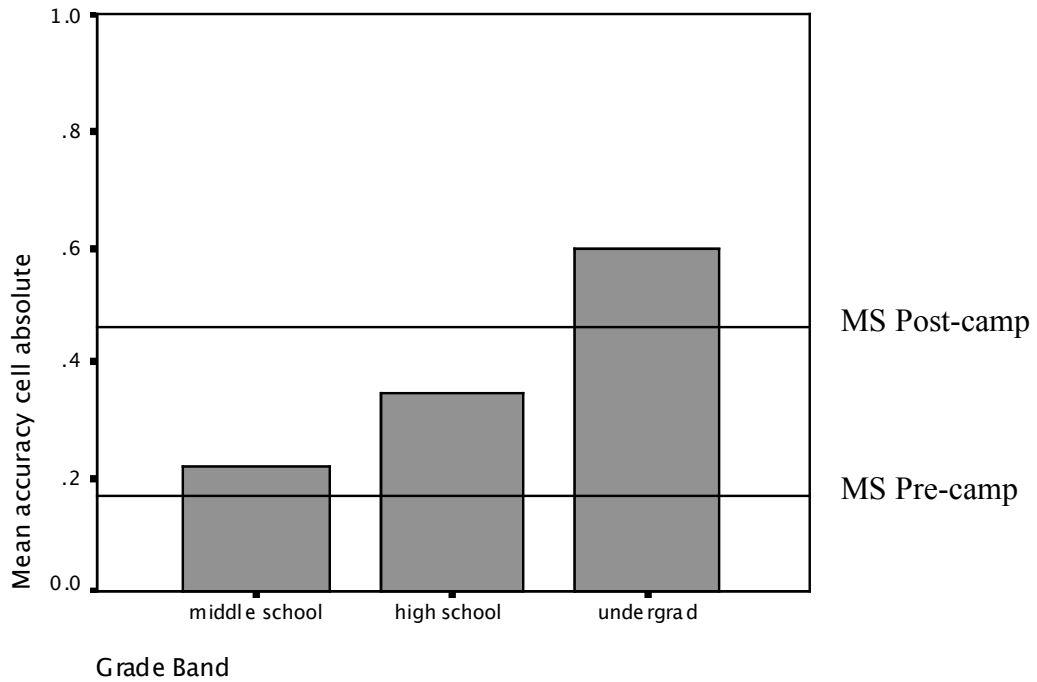

Figure 25: Pre-and post-camp mean score for accuracy of estimates of absolute size of cell (lines), compared to mean score by grade for non-campers (bars)
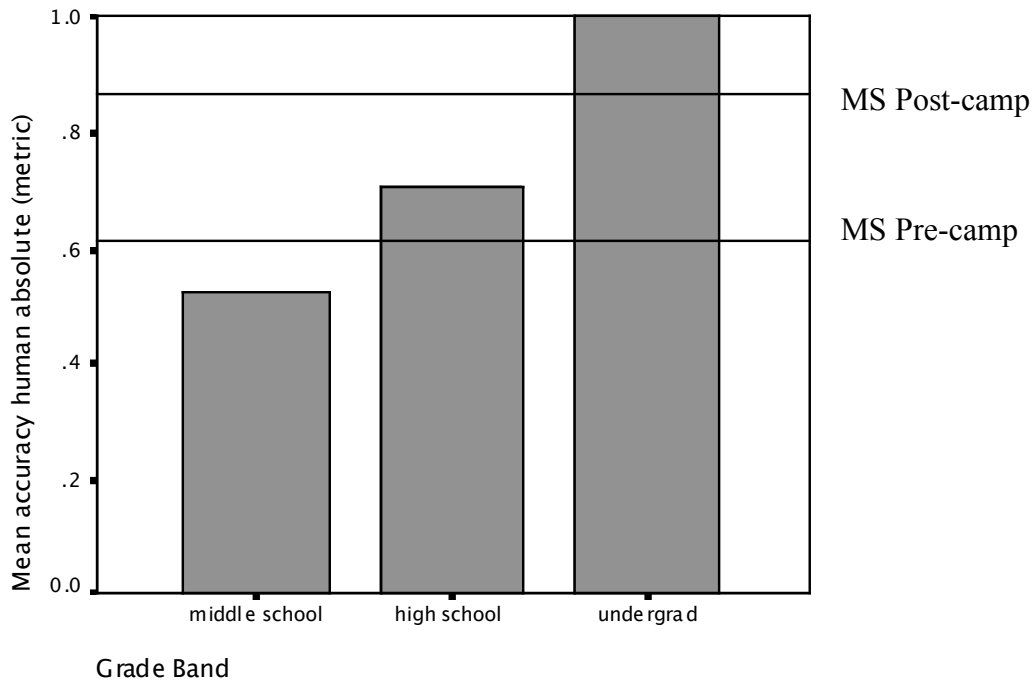
Figure 26: Pre-and post-camp mean score for accuracy of estimates of absolute size of human in metric units (lines), compared to mean score by grade for non-campers (bars)

Despite having introduced students to micrometers and nanometers as fractions of a millimeter, students overwhelmingly resorted to micrometers and nanometers when expressing the size of a cell or an atom in the post-camp assessment, rather than using the corresponding fractions of a millimeter. The few students who *did* use fractions of a millimeter mainly estimated grossly inaccurate sizes such as ½ mm for an atom.

Implications for the learning progression.

The middle school students were able to greatly improve their accuracy in estimating the absolute size of cells and atoms, to around 50%. Presumably with opportunities to learn about the size of the Earth, a similar improvement could be achieved. Thus, the upper anchor for end of high school could reasonably include the expectation that students be able to estimate the size of atoms, cells, humans, and the Earth within a factor of 10, thus refining their knowledge of important landmark objects.

The camp was successful in helping students improve their understanding of the absolute

size of landmark submacroscopic objects, thus informing the learning progression's instructional component. The campers had many opportunities to think about the absolute size of cells and atoms during the camp, in a variety of activities employing different modalities and instructional strategies.

Overall Factual Knowledge

    Efficacy of camp.

The overall score for factual knowledge (calculated as described in Chapter 4) increased from 9.1 (SD = 3.53) to 12.2 (SD = 3.71) (out of 21 possible) with an effect size of 0.83 ($p < 0.001$ for t-test comparing means). In order to contextualize these gains, I compared the pre- and post-camp mean overall factual knowledge score for the campers to the mean score by grade for students in Study One who had not experienced a focused curriculum for size and scale, in Figure 27.

Figure 27: Pre-and post-camp mean overall score for factual knowledge (lines), compared to mean score by grade for non-campers (bars)

This figure shows that campers outperformed most middle and high school students in Study One who had not experienced a focused curriculum; the campers advanced half of the way to the level of the undergraduates.

Summary and Discussion of Gains in Factual Knowledge

The summer camp was fairly effective in helping students increase their knowledge of submacroscopic objects and units that were part of the curriculum, though not the macroscopic objects that were not stressed or covered. Students improved significantly in terms of the smallest object of which they could think, units smaller than the millimeter, ordering macroscopic and submacroscopic objects, the size of the cell in both relative and absolute terms, and the absolute size of the atom. Campers did not change their grouping strategies or accuracy significantly, possibly due to the inclusion of unfamiliar objects in this task. These changes are summarized in Table 17.

Table 17: *Factual knowledge items with statistically significant changes*

Item	Pre-Camp Mean	Post-Camp Mean	Statistical significance of change
Smallest object	1.38	1.87	***
Smallest unit	1.13	1.92	***
Ordering	2.04	2.63	*
Cell relative scale	61%	78%	0.06[a]
Atom absolute size	8%	58%	***
Cell absolute size	17%	46%	*
All four objects, absolute size	1.87	2.96	Effect size = 0.93***
Overall factual knowledge (of 21 points possible)	9.1	12.2	Effect size = 0.83***

Notes: *** p < 0.001   * p < 0.05   ~ p < 0.10
a: one-tailed

At the end of camp, the middle school students reached a level intermediate between middle school and high school students who had not been exposed to a focused curriculum for the smallest object of which they knew, and for ordering; a level intermediate between high school students and the five undergraduates for the smallest unit of measurement and absolute size of the cell; and performed better than the

undergraduates on estimating the relative scale of a cell and absolute size of the atom. In Chapter 4, I explained that a student with solid landmarks of atom, cell, and human (as well as knowledge of the size everyday objects) would be able to reach an overall factual knowledge score of 15, which is better than over 85% of the pre-university students who did not experience a focused curriculum on size and scale. The curriculum guided by a learning progression made progress towards the creation of landmarks and helped students advance half of the distance from their initial performance of 9.1 on the overall factual knowledge score to 15, reaching a mean level of 12.2.

The strong improvements in both absolute size estimation and in knowledge about measurement units are almost certainly related and suggest that learning about submacroscopic units may play a powerful role in students' construction of knowledge about the size of objects. Students who did not use the new units tended to estimate grossly inaccurate sizes. I surmise that the micrometer and nanometer allowed students to off-load the requirement for fractions or decimals onto the inscriptional system of measurement units, thus distributing intelligence onto the environment (Pea, 1988; Martin & Schwartz, 2005). Student #1007 (from Study One) explicitly mentioned her perceived need for units smaller than a millimeter "I don't know how to [estimate absolute size for] the cell because it'd have to be smaller than a pinhead, and I don't think there's any measurements under a mm." The student estimated 1/10 mm for the cell.

It might be argued that students were merely associating nanometer with atom and micrometer with cell. Students are building their understanding of units in part based on what they know about objects, and building their understanding of objects in part based on what they know about units. Therefore, these associations would seem to be an important first step. The additional understanding that a micrometer is 1/1000 of a millimeter and a nanometer is 1/1000 of a micrometer would result in a fairly normative understanding of the size of the landmark objects. I did not directly measure students' grasp of these relationships between units, but observed that some students did seem to understand them well during the activities.

Due to time constraints, we did not include explicit instruction about the nature of the four aspects of size and scale or their connections. Thus, an open question for future iterations of the teaching experiment is the effect on factual knowledge gains of paying meta-level attention to the aspects.

<p align="center">*Students' Development of Consistency*</p>

Next, I describe the learning gains and final state of consistency of knowledge for the students who experienced a focused curriculum on size and scale, thus addressing research question 2B: What is students' consistency across aspects of size and scale, after a focused curricular experience on size and scale?

Order-Group Consistency

Order-group consistency did not change in a statistically significant way according to a McNemar test. However, this type of consistency was near ceiling (78% pre-camp, 87% post-camp), and this may have limited the impact of camp on student knowledge. Thus, it is reasonable to expect that at the end of middle school, all students will be consistent between ordering and grouping.

Order-Relative Scale Consistency

Order-relative consistency improved from 71% to 92% ($p < 0.05$ one-tailed McNemar). This means that more students were able to assign relative scales larger for objects farther in size from the reference object at the end of camp than before camp. For example, during the pre-camp interview, student #20 correctly ranked atom < cell < pinhead; however, she estimated that the red blood cell was five times smaller than the pinhead while the atom was only two to three times smaller than the pinhead. After camp, she estimated the cell at 10 times smaller and the atom 50 times smaller than the pinhead – values that are still not accurate but that now are consistent with the size order.

In order to contextualize these gains, I compared the mean pre- and post-camp order-relative consistency for the campers to that of students in Study One who had no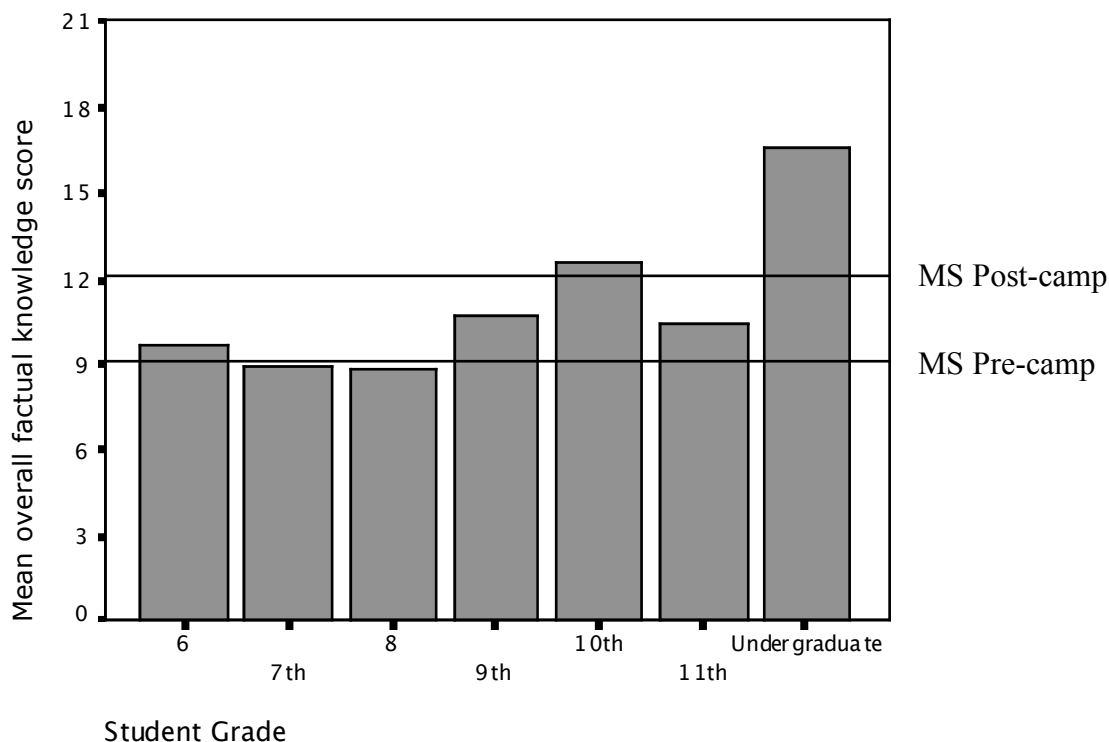t experienced a focused curriculum for size and scale, by grade band, in Figure 28. This figure shows that campers outperformed middle and high school students in Study One

who had not experienced a focused curriculum, and came close to reaching the level of the undergraduates.
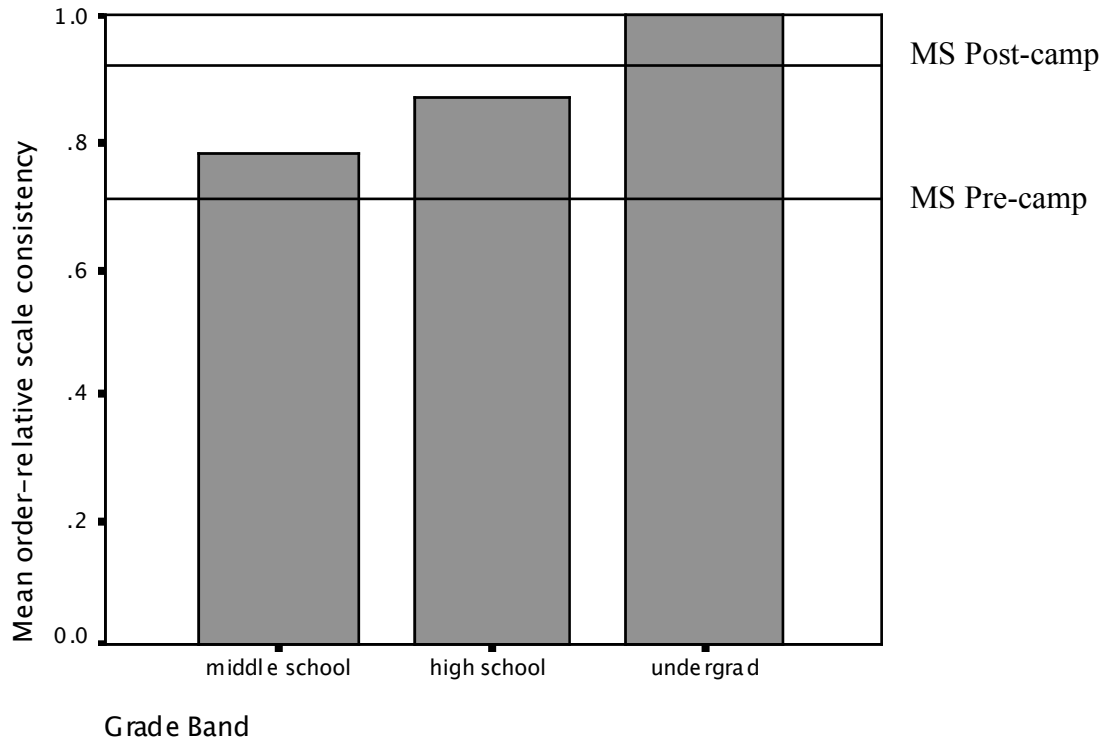
Figure 28: Pre-and post-camp mean score for consistency between ordering and relative scale (lines), compared to mean score by grade for non-campers (bars)

It appears that that the instructional activities involving measurement of one or more object (e.g., skin cell and hair) using another object (e.g., Staph A bacterium) helped students observe that smaller objects must be a larger *number of times smaller* than bigger objects, when both are compared to the same object. This concept was also reinforced by one of the computer simulations (see Song & Quintana, 2009). These instructional activities used transitivity (e.g., the thickness of the hair is smaller than the pinhead, and the cheek cell is smaller than the thickness of the hair, therefore the cheek cell is smaller than a pinhead – even if cheek cell and pinhead cannot both be visualized at the same time) and conservation (the use of the thickness of the hair as a reference

object to compare to objects from the pinhead to the atom). Thus, these activities employed but did not explicitly discuss the prerequisite reasoning abilities identified by Hiebert (1981) to understand the inverse relationship between unit size and number. The inverse relationship between unit size and unit number corresponds to order-relative scale consistency, as mentioned in Chapter 2. I did not measure students' transitivity or conservation directly, but it appears that the activities were helpful for students with a weak grasp of these principles, as many improved in the consistency across aspects (which requires transitivity or conservation – see Chapter 6).

Since 92% of the middle school students who attended the camp displayed consistency between ordering and relative scale after camp, it is reasonable to expect that at the end of high school, all students will be consistent between ordering and grouping. The activities seemed effective in helping students develop this type of consistency. In future iterations, it might be interesting to measure knowledge of conservation, transitivity, and the inverse relationship pre-and post-camp to see if and how these develop, and empirically test the relationship between these principles and order-relative scale consistency.

Order-Absolute Size Consistency

Order-absolute size consistency increased from 42% pre-camp to 67% post-camp, but this change was not statistically significant at the 95% confidence level using a McNemar test. Given that 2/3 of the middle school campers were consistent between ordering and absolute size after the camp, it is reasonable to expect that at the end of high school, all students will be consistent. It appears that learning about micrometers and nanometers help students express sizes smaller than 1 mm, which helped them become consistent.

Relative Scale-Absolute Size Consistency

I found no statistically significant change for the consistency between estimates for relative scale and absolute size coded across all four objects, using a McNemar-Bowker test. However, upon analyzing this type of consistency by object, we found that it improved significantly for human ($p < 0.05$, one-tailed McNemar). Several participants stated that estimating consistent numbers for absolute and relative was not important, as

they were not sure of the exact size of the objects in the first place (so both estimates were bound to be wrong anyway); having a good idea of the absolute size of a human may have led more students to calculate rather than estimate the number of times bigger a human is compared to a 1 mm pinhead. This is explored further in Chapter 6.

Since there was no statistically significant improvement in relative scale-absolute size consistency except for human, it is difficult to draw implications for the upper anchor from my empirical findings. However, relative scale-absolute size consistency is a matter of proportional reasoning, which is an important part of the middle school mathematics curriculum. Thus, it may be reasonable to expect consistency between relative scale and absolute size in all students by the end of high school, as a case of a general ability that should be in place.

Overall Consistency

Importantly, I found a statistically significant mean increase in overall consistency, from 2.19 (SD = 1.37) to 2.81 (SD = 1.03), with an effect size of 0.44 ($p < 0.05$). In other words, students on average increased their consistency across aspects of size and scale, from which I infer that they established new connections and increased their conceptual understanding. Since consistency is what defines the levels on LP1, students progressed along the learning progression.

In order to contextualize these gains, I compare the pre- and post-camp order-relative consistency for the campers to that of students in Study One who had not experienced a focused curriculum for size and scale, by grade band, in Figure 29. This figure shows that campers almost reached the mean level of the high school students in Study One who had not experienced a focused curriculum on size and scale.
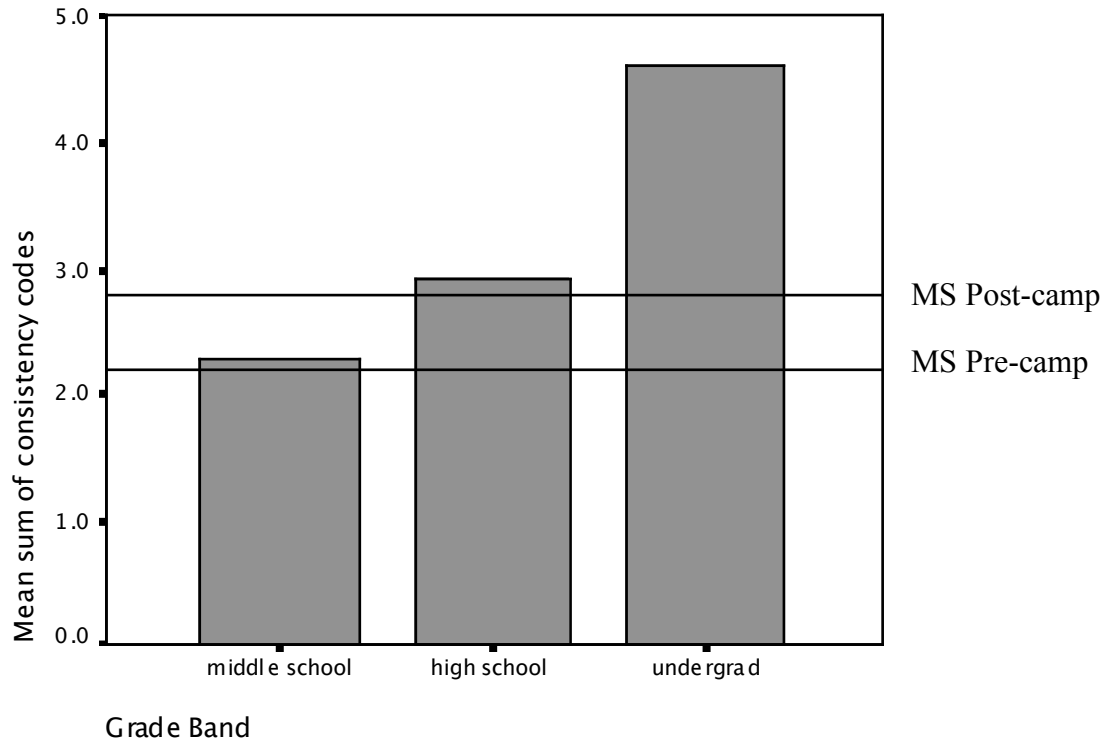
Figure 29: Pre-and post-camp mean score for overall consistency across aspects of size and scale (lines), compared to mean score by grade band for non-campers (bars)

Since the middle school students were still only partially consistent after experiencing the curriculum unit on size and scale, it is not clear what a reasonable upper anchor would be for the end of high school from this study. All but one student reached Level 2 or higher on LP1, and this is a reasonable expectation for the middle school upper anchor.

Summary and Discussion of Gains in Consistency

Students by the end of camp were near ceiling for order-group consistency. They experienced statistically significant improvements in their order-relative consistency, surpassing high school students who had not experienced a focused curriculum on size and scale. Their overall consistency level also improved significantly, to a level near that of high school students in the absence of a focused curriculum; thus, they seem to have progressed along LP1. The campers also improved their consistency between relative scale and absolute size for the human. These changes are summarized in Table 18.

141

Table 18: *p-values for statistically significant changes in consistency*

Item	Pre-camp	Post-camp	P value
order-relative consistency	71%	92%	*
relative-absolute consistency	17%	42%	
human			*
level of consistency	2.19	2.81	Effect Size = 0.44*

Note: *** significance < 0.001   * significance < 0.05   ~ significance < 0.10

While some students are able to spontaneously restructure their knowledge in response to learning activities, others may require additional instruction focusing on the principles that are of importance (Schwartz & Bransford, 1998; Klahr & Siegler, 1978). Specifically for the measurement estimation (of absolute size), Joram and colleagues (1998) caution that children may not go beyond knowing isolated reference points to build a measurement line if not pushed to do so by their teacher. Future iterations of our teaching experiment will incorporate meta-level discussion of the four aspects, the connections among them, and of general reasoning skills such as conservation and transitivity, in order to analyze whether these modifications lead to greater learning gains in consistency.

Patterns in Student Knowledge After Camp

In this section I address whether and how the patterns of factual knowledge and consistency in the students who experienced a curriculum focused on size and scale differ from those of students in Study One who had not experienced the focused curriculum, thus addressing research question 2C: How do the patterns of factual knowledge and consistency differ from those of students who have not experienced the focused curriculum, if at all?

Patterns in Factual Knowledge After Camp

Patterns across aspects of size and scale.

In the students who had experienced only the traditional curriculum, I found mainly weak statistically significant correlations between the different tasks for factual knowledge. The factual knowledge of summer campers after experiencing the curriculum features

moderate to strong correlations that are statistically significant, but fewer significant correlations overall. This may be due to the low power caused by the small number of participants – only correlations of 0.4 or higher are statistically significant. The correlations that were statistically significant are shown in Table 19.

Table 19: Statistically significant intercorrelations between tasks after camp

	1	2	3	4	5	6	7	8	9	10	11	12
1 Smallest object	--											
2 Smallest unit	.70**	--										
3 Ordering			--									
4 Atom relative				--								
5 Cell relative	.41*	.59**	.61**		--							
6 Human relative						--						
7 Earth relative							--					
8 Atom absolute			.57**		.61**			--				
9 Cell absolute			.57**		.47*	.47*		.78**	-			
10 Human absolute						.43*				--		
11 Earth absolute				.51*							--	
12 Human absolute, metric						-.47*						--

* p < 0.05   **p < 0.01

The shaded boxes show the cells for absolute size and relative scale for a given object. These correlations are a measure of consistency because relative scale can be calculated from absolute size. As can be seen, these correlations are not statistically significant for atom and earth, and are moderate and significant for cell and human. Consistency is discussed in greater detail below. The two strongest correlations are those between smallest object and smallest unit, and between absolute size of cell and absolute size of

atom. These were important learning goals for camp, and the fact that they are correlated in pairs suggests that students did not learn in disconnected bits and pieces; however, the fact that not all knowledge is correlated shows that students' learning is not entirely connected either. This may be due to students learning different things depending on their prior knowledge, for instance, depending on their level on the learning progression.

As one might expect, ordering is correlated to the absolute size of the cell and the atom, and also to the relative scale of the cell. This makes sense because correctly ordering the submacroscopic objects depends on knowing the sizes of the atom and the cell. Also worthy of note is the fact that the relative scale of a cell is correlated to many other tasks: absolute sizes of cell and atom, smallest unit, and smallest object. This may mean that the relative scale of the cell may be an important landmark, may be a part of effective instructional activities, and may be a good assessment item in future tests.

Patterns within aspects of size and scale.

I found that 22 of 23 students (96%) fit along a pattern in which students learn the relative scale of the cell or the human first (in indistinct order), and only then learn the relative scale of the atom or the earth. This path follows the order of difficulty detected in Study One (see Table 5 in Chapter 4), where 82% of students fit on this pattern. In other words, the trend where the relative scale of objects closer to human scale are easier to estimate was maintained throughout camp.

On the other hand, only 10 of 24 students (42%) fit along the old pattern in which students learn the absolute size of the human in any units first, then the human in metric, then the cell and Earth in indistinct order, and finally the atom. This is the trajectory suggested by the relative difficulty of estimating each object found in Study One (see Table 20). Instead, 19 of 24 students (79%) now fit along a path suggested by the difficulty of estimating the absolute size of the objects after camp: human in any units, then human in metric, then atom, then cell, and finally Earth. The fact that the relative or absolute size of the Earth is last to be learned is not surprising because it was not covered in camp. What is surprising is that more students were able to estimate the absolute size

of the atom than the cell, despite the atom being farther from human scale. All accurate answers were expressed in nanometers; all answers employing millimeters were grossly incorrect (e.g., ¼ mm, ½ mm, 1 mm). This again points to the importance of units in students' learning of the size and scale of landmark submacroscopic objects.

Table 20: *Proportion of students accurately estimating absolute size of atom, red blood cell, human, and Earth, before and after camp*

Object for absolute size estimation	Pre-camp	Post-camp
Atom	0.08	0.58
Red blood cell	0.17	0.46
Human	0.92	0.92
Earth	0.08	0.17
Human, in metric	0.61	0.83

Patterns in difficulty of aspects of size and scale.

After experiencing the focused curriculum on size and scale, practically all students were able to come up with submacroscopic objects and submacroscopic units. Thus, while the submacroscopic units are less familiar than submacroscopic objects in the absence of a focused curriculum, units can be learned by students, and then are useful in many other tasks.

Grouping continued to be easier than ordering even after camp: 20 of 23 students correctly grouped, while only 11 of 24 ordered correctly, in both cases considering eight objects (i.e., excluding mitochondria and molecule, which were not part of the summer camp's curriculum).

An important finding is that students' mean ability to estimate absolute sizes was significantly *higher* than their mean ability to estimate relative scales, by the end of camp, as shown by a t-test ($p < 0.01$). In comparison, prior to camp there was no significant difference. This analysis compares the relative and absolute estimates for four objects so the maximum total score is 4 in both cases (i.e., the estimate for human in metric units is not included). Prior to camp, the mean sums of relative codes (1.61, SD =

145

1.11) and of absolute codes (1.26, SD = 0.86) were not significantly different, but post-camp, absolute size estimation was significantly more accurate than relative (2.13, SD = 1.04, vs. 1.63, SD = 0.88), respectively. This indicates that appropriate instruction about measurement units and absolute sizes may be a powerful strategy to improve students' knowledge about the size of objects, including those which can become landmarks objects and help students characterize size regimes or "worlds". The micrometer and nanometer opened an avenue to learn about submacroscopic objects without fractions, decimals, or negative powers of 10. Even scientists may benefit from this strategy:

> At the other end of the spatial scale, small units, such as the nanometer may be conceptualized as a new unit (rather than as a fraction of a meter) and subsequently used by nanoscientists for investigations in the nanoworld. The ability to accurately compare and develop intuition about vast and tiny spatial scales may depend heavily on ability to unitize appropriately. (Tretter, Jones, & Minogue, 2006, p. 1063)

While I did not test students on their ability to convert between units or the numerical relationship between micrometers, nanometers, and millimeters, their ability to associate the nanometer with the atom, a micrometer with the cell, seems a promising start in developing better understanding of the submacroscopic world.

In summary, there is evidence of a growing integration of knowledge, with knowledge of units of measurement and absolute sizes an important component of students' learning gains. Additionally, I observed that students' learning gains were sensitive to instruction, in that students' knowledge of the size of objects that did not form part of the curriculum did not change as much as for objects that were included.

Summary and discussion of patterns in factual knowledge.

The summer camp students appeared to be creating a "relational web" for size and scale (Tretter, Jones, Andre, et al., 2006) as some correlations between pieces of knowledge grew much stronger. However, low power may have kept weak and moderate correlations from achieving statistical significance. Strong correlations between absolute sizes of cell and atom, and between smallest object and smallest unit may signal the importance of

units to student learning. The relative scale of the cell was highly correlated to other pieces of knowledge, possibly indicating its emergence as a reference point in students' "mental measurement line" (Joram, 1998).

Relative scale estimation maintained the previously observed pattern of greater accuracy for objects near human scale, but absolute size accuracy dramatically increased for both the atom and the cell in tandem with the use of micrometers and nanometers. The initial difference between knowledge of submacroscopic objects and submacroscopic units disappeared, and students became significantly more accurate at absolute size estimation compared to relative scale estimation at the end of camp. These findings again highlight the importance of units to students. In fact, students who continue to use the millimeter in their absolute size estimations were in most cases grossly inaccurate.

Patterns in Consistency After Camp

Almost all of the students (91%, or 21 of 23) continued to conform to the developmental path outlined in Chapter 4. Thus, the teaching experiment provides support for LP1 generated through the cross-sectional study described in Chapter 4. Of the 19 students who conformed both pre- and post-camp, five remained on the same level, seven increased one level, two each increased two and three levels, and three regressed one level (see Table 21.) There is a statistically significant Pearson correlation of -0.617 (p < 0.01) between initial level of consistency and consistency gain, showing that students at a lower initial level tended to gain more. (However, this effect disappears after controlling for initial *factual* knowledge – see Chapter 6). Of the students who did not conform to the learning progression, two stayed roughly the same, one improved, and one regressed. The students who regressed did improve their overall factual knowledge, however.

In future work, I would like to revise the teaching experiment to include explicit attention to the four aspects of size and scale and how they are related, while maintaining the opportunities to interact with a variety of important objects in science across various aspects of size and scale as the current iteration does. Paying explicit attention to the principles we wish students to understand after they have had an opportunity to interact

with the phenomena may be a fruitful strategy (see Schwartz & Bransford, 1998; Klahr & Siegler, 1978). It would be very interesting to see if students' gains on consistency improve with this modification.

Table 21: *Students' movement pre- to post-camp on consistency level*

Number of Students	Pre-Camp Level	Post-Camp Level	Change in Levels
1	0	0	0
2	0	2	+2
1	1	2	+1
1	1	4	+3
5	2	3	+1
1	2	5	+3
2	3	2	-1
3	3	3	0
1	3	4	+1
1	4	3	-1
1	4	4	0

Summary and discussion of patterns in consistency.

A major finding is that students continue to conform to LP1's sequence of levels defined by consistency. This implies that LP2 can continue to be organized along consistency. Students tended to advance along the learning progression during camp, with students at a lower level progressing more. As only one student reached Level 5 consistency, this trend does not seem to be due to a ceiling effect. An interesting question for a second iteration of the teaching experiment is to research whether meta-level attention to the four aspects and consistency will lead to greater gains or different patterns.

Limitations of the Teaching Experiment

A limitation of the study is that the effect of interviewer on student responses was not evaluated. This was due to the absence of a reliable external measure related to students' knowledge of size and scale against which to compare students' performance as a function of interviewer. However, prior to conducting interviews the new interviewer read literature on interviewing students in science education (White & Gunstone, 1992), and I observed him during two practice interviews. This, along with a detailed interview

148

protocol with suggested prompts, ensured some level of consistency across interviewers. Additionally, the mean factual knowledge and consistency of knowledge scores for the students I interviewed as compared to those interviewed by the other interviewer prior to camp were not significantly different (p values of 0.3 and 0.6 respectively).

Another limitation is that the small number of respondents in the analysis (24) resulted in low power, which may have led to the lack of statistical significance in some items.

Another limitation of this study is that of limited generalizability. Students who participated in the camp were self-selected (by their own interest, or that of their parents) and thus perhaps not typical. However, it must be noted that there were no aptitude requirements for the camp. Initially, students were admitted from among applicants based on a short essay, but every student who was placed on the wait list was eventually offered a spot in the camp as some students declined to participate due to transportation or family/health problems. The learning gains experienced by students might be different with a different population. However, all samples tended to fit on the progression of increasing consistency: students who self-selected for camp and students who did not; students from a low SES, diverse public school district, and students from a mid-high SES private school; middle school, high school students, and undergraduates. This would seem to indicate that this pathway is fairly robust. The fact that units appear to be crucial in the development of several types of knowledge may mean that learners who are more familiar with the metric system (i.e., students in other countries or recently immigrated to the United States) may develop differently.

Students were interviewed only two weeks apart, which may have led to two different types of problem. First, students may have learned from the first test, thus inflating learning gains. However, students did not receive feedback concerning their answers in the first interview, so this effect is unlikely. A second and opposite type of problem is the possibility of students remembering and repeating their answers from the first interview during the second interview; this would have the effect of lowering apparent learning gains. The following transcript illustrates this problem. (This is from a student who was

not included in the analysis due to his missing some of the lessons on size and scale; however, he most clearly illustrates this potential problem.)

(pre-interview)

> I: what's the smallest thing you can think of?
> R: a screw.
> I: can you think of anything that's too small to see?
> R: A piece to a glasses, to glasses, like another screw but a really really small one.
> I: can you think of anything too small to see…?
> R: [no]

(post-interview)

> I: what's the smallest thing you can think of?
> R: A head of a pin
> I: can you think of anything that's too small to see?
> R: Still the head of a pin probably, or a little screw, that's [inaudible] from glasses.
> I: and if you think of some of the things that we've done in camp can you think of something that's too small to see?
> R: Umm (20 s pause) [no]

The student then ordered the 10 objects with the pinhead being the smallest. At the end of the interview, I asked him about some activities from the camp.

> I: Now I just want to ask you some questions about some activities in the camp. Do you remember that simulation that made the clicking noise?
> R: Yup.
> I: You were lining things up across the head of a pin. What are some of the things that you lined up?
> R: We lined up atom across, umm, like a rhinosvirus [sic]
> I: Rhinovirus… so, are those things smaller than the head of a pin?
> R: Nnn, not really because like it only took a couple, for the rhinovirus it only took like a couple sec…umm, a minute, so yeah it might be small, see it's sort of, it's like smaller than the head of a pin mostly.

Another limitation is that, although the interview protocol tied in well to the curriculum, the fit was not perfect. For instance, the ordering and grouping tasks included molecule and mitochondria, which were not included in the curriculum. Even though I also coded these tasks without them, their presence may have distracted or confused students,

150

providing an inaccurate picture of what they knew at the end of camp.

Study Two measured what students knew at the end of a 12-hour curriculum, but did not assess how this knowledge behaved over time. Follow-up interviews could have revealed what knowledge was easily forgotten, and what knowledge (if any) continued to develop over time, as students advanced and suddenly realized that experiences in camp were relevant to their learning (as has been observed with museum experiences – Rennie & Johnston, 2004).

Another limitation of this dissertation is that I treated several variables as continuous, interval-level variables (for t-tests) even though they were in fact only ordinal-level variables with around six hierarchical levels. I resorted to this treatment because more appropriate statistical tests, like the McNemar-Bowker test, required conditions that were not met either (square matrices). However, the fact that this is a first investigation of a new construct signifies that the conclusions are tentative and the tools less than perfect. Future work can continue to improve on the methodology; meanwhile, the results should be seen as informative rather than definitive in nature.

<div align="center">Chapter Summary and General Discussion</div>

In this chapter I described how LP1 informed the development of the instructional unit employed in the teaching experiment. LP1 provided starting points and endpoints for the curriculum (the lower anchor and upper anchor) that helped define learning goals. I described how we searched for existing instructional activities to modify and created others to help students construct their knowledge in the sequence suggested by LP1. The major findings are support for the organization of the learning progression around consistency, evidence for the advantage of early instruction in units smaller than the millimeter, and support for the feasibility of building landmarks of atom and cell for the submacroscopic world at the middle school level. The findings reported in this chapter also show that the instructional activities used in the teaching experiment are effective, and suggest possible improvements – meta-level discussion of the aspects, consistency, and important logical and mathematical principles. The effectiveness of the curriculum

lends support to the learning progressions approach.

A key finding of the teaching experiment is that over 90% of students continue to conform to LP1's sequence of levels defined by consistency. This means that LP2 can continue to be organized along consistency. The teaching experiment also informs the upper anchors for middle school and high school.

A second major finding is that learning about units smaller than a millimeter greatly helped students learn about the submacroscopic world. Students who adopted the micrometer and nanometer were able to estimate the size of atoms and cells more accurately than students who used the millimeter; the new units helped them develop landmarks in the submacroscopic world. The relative scale of the cell was highly correlated to other pieces of knowledge, possibly indicating its emergence as a reference point in students' "mental measurement line" (Joram, 1998). Relative scale estimation maintained the previously observed pattern of greater accuracy for objects near human scale, but absolute size accuracy dramatically increased for both the atom and the cell in tandem with the use of micrometers and nanometers. Students became significantly more accurate at absolute size estimation compared to relative scale estimation at the end of camp. Units may be allowing students to off-load the requirement for fractions or decimals onto the inscriptional system (Pea, 1988) of measurement units.

A third major finding of this study is that a curriculum designed in a principled way following a learning progression was in fact able to achieve significant learning gains as well as informing future iterations of the learning progression itself. This lends support to seminal papers that consider that learning progressions may be important in improving science education (Smith et al., 2006; Duschl et al., 2007). Students improved their ordering, their absolute and relative ideas of the size of cells, and the absolute size of an atom, in some cases beyond the level of five undergraduates at a selective research university. Students were near ceiling for order-group consistency and improved significantly on order-relative consistency. The campers also improved their consistency between relative scale and absolute size for the human. An interesting question for a

152

second iteration of the teaching experiment is to research whether meta-level attention to the four aspects and consistency will lead to greater gains or different patterns. These additions may help students restructure their knowledge in response to learning activities (Schwartz & Bransford, 1998; Klahr & Siegler, 1978).

The question of whether consistency helps students meaningfully accumulate factual knowledge and/or accumulating factual knowledge aids the development of consistency is addressed in Chapter 6.

CHAPTER 6: RELATIONSHIP BETWEEN FACTUAL KNOWLEDGE AND
CONSISTENCY OF KNOWLEDGE

Study One resulted in LP1, organized around patterns of consistency across aspects of size and scale. LP1 included descriptions of typical factual knowledge at each level of consistency. However, the cross-sectional design of Study One did not allow me to investigate how factual knowledge and consistency of knowledge interact as each type of knowledge grows over the course of the learning progression. Study Two, a teaching experiment, showed some instances in which consistency appears to aid the growth of content knowledge and identified some circumstances in which factual knowledge can contribute to consistency. It also showed that additional logical or mathematical knowledge is necessary in some tasks. In this chapter, I further examine the relationship between factual knowledge and consistency of knowledge through a theoretical task analysis for each aspect and empirical analyses drawing on data from both studies. I also examine the role of logical and mathematical knowledge in thinking about size and scale using the four aspects.

The relationship between factual knowledge and consistency of knowledge is important because of its implications for instruction. We ultimately want students to have extensive, accurate, and well-connected knowledge (Chi & Ceci, 1987) about the size of important objects in science. An instructional approach inspired by Piaget (e.g., Piaget, 1983) would posit that domain-general logical structures have to be in place before truly understanding any specific content domain; in this approach, one would focus on building student understanding of the connections across aspects and the necessary logical or mathematical structures, and expect these to transfer broadly to multiple contexts. Hiebert (1981) found that students could learn most of the concepts and skills involved in measurement in the absence of logical structures such as conservation and

transitivity; however, he also found that the inverse relationship between unit size and unit number (involved in order-relative scale consistency) did depend on these logical structures. It is thus critical to further analyze the relationship between consistency of knowledge and factual knowledge to inform instructional strategies and in order to identify logical and mathematical knowledge necessary for the growth of understanding that size and scale. A careful task analysis (Hiebert, 1981) may reveal the nature of the relationship between factual knowledge and consistency. Analysis from empirical data is essential; as Hiebert notes: "Task analyses carried out from a child's perspective rather than the adult's may help to illuminate the cognitive demands of individual tasks." (1981, p. 209).

In this chapter I first report on a theoretical task analysis identifying and describing different strategies to solve size and scale problems involving the four aspects, then use the data from my two studies in an empirical task analysis that studies the strategies that students actually use. In other words, first I examine whether connections across aspects *can* be used in solving factual knowledge tasks, and then I examine whether students *do* in fact use these connections. Likewise, I examine whether accurate factual knowledge *could* and *does* play a role in determining whether students make connections across aspects. This analysis characterizes the relationship between factual knowledge and consistency one task at a time and in a qualitative way.

Next, I use the cross-sectional data of Study One to examine more globally the relationship between factual knowledge and consistency of knowledge, using a correlational analysis. The quantitative nature of this analysis allows me to assess the weight of the relationship between consistency of knowledge and accuracy of factual knowledge in a way that the task analysis does not. However, the cross-sectional nature of the data collected in Study One provides no information about the relationship in terms of possible mechanisms or patterns of development. I use the data from Study Two in a regression analysis to gauge the relative importance of factual knowledge and consistency of knowledge to subsequent learning gains of factual knowledge, and to subsequent gains of consistency of knowledge.

In order to unpack the four aspects of size and scale, define the knowledge required to carry out the interview tasks, and examine the relationship between factual knowledge and consistency of knowledge, I carried out a detailed task analysis with two phases. First, I generated all the strategies of which I could think or which I found in the literature to carry out *generic* factual knowledge tasks for each aspect of size and scale (e.g., to determine or estimate the relative scale of any object), thus conducting a theoretical task analysis. Second, I conducted an empirical examination of students' actual strategies on the *specific* tasks I posed to them in their interviews. Both task analyses identified logical and mathematical knowledge required to successfully carry out the tasks as well as the factual knowledge about the size of objects needed for various strategies. This analysis also shed light on the relationship between factual knowledge and consistency.

For every aspect, I present a flow diagram that summarizes the generic, theoretical task analysis, as well as a prose description of each strategy generated. The order of the strategies is arbitrary; there is no implication that strategy 1 is in any way prior to or more common than strategy 2. Across this entire section, I use consistent color coding on the flow diagrams: pink for ordering, lilac for grouping, yellow for relative scale, and blue for absolute size. I examine absolute size first, because it is the only aspect that involves a single object. I examine relative scale second, because it involves only two objects; then ordering, which involves iterative cycles of comparisons of two objects; followed by grouping, where various objects must be considered simultaneously. After considering these four aspects, I examine my interview tasks of smallest object and smallest unit, which are in the ultimate analysis instances of ordering (or the comparison that makes up ordering). I then present my findings concerning the actual strategies that students used, i.e., the specific, empirical task analysis. I include examples of successful and problematic applications by students of the strategies generated in the theoretical task analysis, as well as descriptions and analyses of strategies students used that I had not foreseen.

Theoretical Task Analysis

As mentioned in chapter 2, absolute sizes are magnitudes that are established in relation to conventionally defined units. Unlike ordering, grouping, and relative scale, absolute size involves only one object and uses conventionally defined units of measurement; this standardization facilitates communication (Lehrer, 2003). Figure 30 shows that determining absolute size can be approached in several ways.

      Strategy 1: absolute size by recall.

This strategy ends in the blue oval at the top left of Figure 30, and consists of recalling the known absolute size of an object. For instance, most American adults can probably recall the fact that the height of the average human male is around 5'10". A familiar object of a size accurately known in terms of standard units is known as a "reference point" (Joram et al., 1998) or "landmark" (Tretter, Jones, Andre, et al., 2006). This strategy does not require any connection to other aspects of size and scale. Generating the known absolute size in the first instance required measurement, which involves a web of foundational ideas (e.g., Lehrer, 2003). A learner may use landmarks in establishing a "relational web of scale sizes" (Tretter, Jones, Andre, et al., 2006, p. 307).

      Strategy 2: absolute size through known relative scale.

This strategy ends in the yellow oval at the top left of Figure 30. If the relative scale of an object is known, and also the absolute size of the reference object, then the absolute size of the object can be calculated by multiplying. For instance, if a school bus is known to be twice as long as an SUV, and the SUV is known to be 5 m in length, then the length of the school bus can be calculated by multiplying 5 m times 2. This strategy exemplifies the connection between relative scale and absolute size. This estimation strategy is called "prior knowledge" by Joram and colleagues (1998). A special case of this strategy is when the relative scale is precisely 1, that is, when an object is about as long as another object (present or recalled) of known size. This strategy is called "comparison" by Joram and colleagues (1998). If the object is present, it functions as a "perceptual standard"
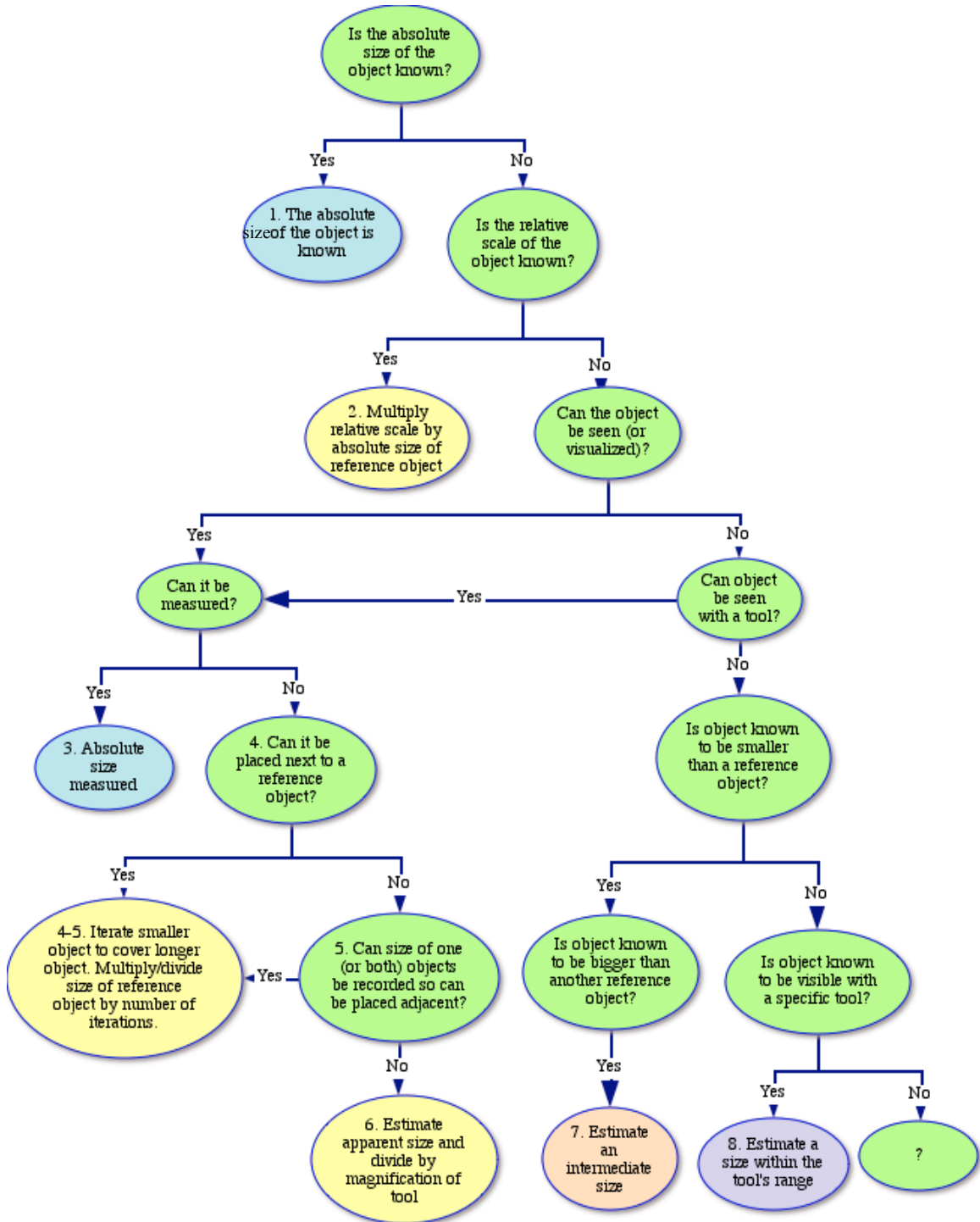
Legend: Ordering Pink, Grouping Purple, Relative Scale Yellow, Absolute Size Blue

Figure 30: Flowchart for strategies for absolute size

(Gelman & Ebeling, 1989). A variant is using one of these strategies to estimate the size of a *part* of the object, then multiplying the number of parts; this strategy is called "decomposition/recomposition" by Joram and colleagues (1998).

Strategy 3: absolute size by measurement.

This strategy ends in the blue oval at the left of Figure 30 and consists of using a measurement tool, such as a ruler, to determine the absolute size of an object. This strategy does not depend on the connection to any other aspect of size and scale. This strategy can be employed even when using a tool to see an object, for instance, placing a plastic ruler next to a tiny ant when using a magnifying lens, or using microscope slides with grids of known dimensions. Once could also record the length of an object (for instance, with two marks on a length of string) and then measure the distance between the marks. This requires conservation of length (Piaget & Inhelder, 1971; Piaget et al., 1960; Clements & Stephan, 2004). The foundational principles of measurement according to various authors (Wiedtke, 1990; Lehrer, 2003; Clements & Stephan, 2004) are discussed in Chapter 2.

If the object is not physically present and visible, but can be visualized mentally, this strategy can still be used by mentally lining up a measuring device next to the object. This estimation strategy is called "mental instrument application" by Joram and colleagues (1998). Alternatively, if a ruler or other measuring device cannot be visualized, the learner can mentally segment the object into standard units (e.g., feet) and count the units. This estimation strategy is called "unit iteration" by Joram and colleagues (1998).

Strategy 4: absolute size through direct iteration.

This strategy ends in the oval shown in yellow at the left of Figure 30. It involves the use of relative scale to calculate absolute size, and thus relies on the connection between these two aspects. A reference object of known size (a "perceptual standard" - Gelman & Ebeling, 1989), such as a hand span of 8 inches, can be iterated across a larger object. The absolute size of the object is the product of the number of iterations and the absolute

size of the reference object. If the reference object is larger than the object of interest, then the object of interest can be iterated across the reference object instead. The absolute size of the object is then the absolute size of the reference object divided by the number of iterations. Iteration is considered a fundamental principle of measurement (Wiedtke, 1990; Lehrer, 2003; Clements & Stephan, 2004).

If the object is not physically present and visible, but can be visualized mentally, this strategy can still be used. An object equivalent in size to a standard unit (e.g., a shoe that represents a foot) can be iterated (a strategy called "recall reference point" by Joram et al., 1998)

Strategy 5: absolute size through indirect iteration.

This strategy culminates in the same yellow oval as does strategy 4, but arrives through a different pathway. When the two objects cannot be placed next to each other, one or both sizes may be recorded (perhaps as two marks on a length of string) so that the iteration may take place. Iteration is considered a fundamental principle of measurement (Wiedtke, 1990; Lehrer, 2003; Clements & Stephan, 2004). The portability of the size of an object depends on conservation of length (Piaget & Inhelder, 1971; Piaget et al., 1960; Clements & Stephan, 2004). This strategy depends on the connection between relative scale and absolute size.

Strategy 6: absolute size of objects through estimation and magnification factor.

The strategy ends in the yellow oval at the bottom of Figure 30. In the case of objects that are visible only through a tool such as an optical microscope, one can estimate the *apparent* size of an object in the field of view (using any estimation strategy – see Joram et al., 1998) and then divide by the magnification. For instance, if a human hair appears to be 1 cm across when viewing it at 100X magnification, we can estimate the thickness at 1/100 cm or 100 μm. This strategy depends on the connection between relative scale and absolute size.

Strategy 7: absolute size using the known absolute size of larger and smaller objects.

This strategy ends in the pink oval at the bottom of Figure 30. It depends on knowing the absolute size of larger and smaller objects, and then estimating an intermediate value. For instance, if one knows that the diameter of a buckyball is around 0.7 nm, and the thickness of DNA around 2 nm, and furthermore that a single-walled nanotube is intermediate in thickness between a buckyball and DNA, then one can estimate a value in between 0.7 and 2 nm for the nanotube. This estimation strategy is called "squeezing" by Joram and colleagues (1998), and can be used for objects that are visible or can be visualized. The strategy depends on the connection between ordering and absolute size, as it relies on knowing objects smaller and larger than the object of interest.

Strategy 8: absolute size through tool range.

The strategy ends in the purple oval at the bottom of Figure 30. In the case that the only information about the size of an object is that it can be seen with a specific tool, we can still narrow down its range of possible sizes to that of the tool. For instance, if we know that an object is too small to be seen with an optical microscope, but it can be seen with an electron microscope, then we have narrowed its range of possible sizes to approximately 0.2 to 200 nm – still a huge range, but much smaller than the total possible range of values (which is essentially infinite). The strategy depends on the connection between absolute size and grouping, as we use group membership of an object (e.g., the object is a member of the group of objects that can be seen with an electron microscope) to estimate its absolute size.

Summary of theoretical task analysis for absolute size.

The absolute size of an object can be previously known, measured at the time, or obtained from other absolute sizes. Obtaining the absolute size of the object of interest from other absolute sizes requires the connection between absolute size and other aspects of size and scale. The connection between absolute size and relative scale is used in calculating absolute size from a part of the object or a second object that is iterated across the object of interest. This process involves understanding the additivity of lengths or partitioning (Lehrer, 2003; Stephan & Clements, 2003). Absolute size can also be

calculated via relative scale from a scaled image of the object. Absolute sizes calculated via relative scale can have good precision. Lower precision estimates of absolute size can be obtained through the connection between ordering and absolute size (squeezing – Joram et al., 1998) or the connection between ordering and grouping (tool range). The difference in precision is due to the qualitative nature of ordering and grouping, which do not involve an exact characterization of size.

Next, I present my empirical findings for the specific absolute size task I employed in my studies.

Empirical Task Analysis

   Strategy 1: absolute size by recall.

This strategy was almost universally employed (successfully) for the height of the human. English units were used in accurate answers more often than metric units. Much less frequently, students accurately recalled the diameter of the earth, the red blood cell, or the atom. This strategy became much more widespread and accurate for atom and cell, among the middle school students who experienced the focused curriculum for size and scale in the summer camp described in Chapter 5. The campers thus seemed to be creating additional landmarks for the submacroscopic world.

   Strategy 2: absolute size through known relative scale.

Some students used this strategy to estimate the diameter of the Earth. Student #9018 said: "From here to Chicago is around 200 [mi], so [the diameter of Earth is] 4000 mi." Similarly, student #0099 estimated the distance from the location of the interview to Boston at 3000 miles (the actual distance is around one-fourth of that) and from that estimate concluded that the diameter of the Earth is around 10,000 miles. Both students used their estimate of the relative scale of the diameter of the Earth (e.g., 20 times more than the distance from their location to Chicago) and multiplied by the known (or estimated) absolute size (distance). This strategy corresponds to Joram and colleagues' decomposition/recomposition estimation strategy (1998).

However, many more students did *not* use the strategy, and when asked about the relationship between relative scale and absolute size, several stated that there was no relationship at all. For instance, student #1008 was asked, "OK. You have some numbers about [the objects] up here [relative scale]. Might those help you think about the [absolute] sizes here?" and responded "Yeah, but you were comparing these, so there's sort of a difference". Student #0080 expressed a similar idea:

> I: OK. Do you think [relative scale and absolute size] are related?
> R: Yes…wait a little…NO.
> I: No? OK. Umm, why do you think they're not related? Like, how are they different?
> R: Because this is their actual size and this is them comparing them to other objects.

Other students fell in between these two extremes, stating that consistency across relative scale and absolute size was important only if you know the absolute size (but not if you are guessing); or that being consistent for some objects but not others was not a problem. Obviously, the relationship between absolute size and relative scale, which is a matter of proportional reasoning, is a relationship that most students do not understand well. However, it appears that having a good idea of the absolute size of an object can lead students to *try* to be consistent. For instance, student #9030, when asked about whether estimates for absolute size should be consistent with estimates for relative scale, said "If I had the actual numbers for relative size, then yeah, it would be important to use those, but since I just kind of made those up…" Student #0144 had a similar explanation: "[My estimates for relative scale and absolute size] would have to be related if they are for the same human…they probably should be, but because my guesses are probably way off, it doesn't have to fully follow one thing". This is an instance of factual knowledge supporting consistency across aspects of size and scale.

Since the interview protocol asked students to estimate relative scale before absolute size, students had the opportunity of calculating absolute science from their previous relative scale estimates. While experts and most undergraduates had calculated the relative scale from known or estimated absolute sizes in the first place, and now had only to recall their

original absolute size estimates, students who had not employed this strategy were calculating absolute size for the first time. Some students used their relative scale estimates to generate absolute size estimates. For instance, student #1007 estimated that the red blood cell was 10 times smaller than the head of the pin and the atom 50 times smaller. She then estimated absolute sizes of 1/10 mm and 1/50 mm for the cell and atom, respectively. She stated that she had used the relative scales in coming up with the absolute sizes. In fact, after estimating that the human was 100,000 times larger than the head of the pin, she estimated the height of the human at 100,000 mm. Many students in this situation realized that their estimate for absolute size (generated from their earlier relative scale estimate) was incongruent with what they knew of the size of the object, and then changed their relative scale answer to be consistent with the known absolute size (e.g., #0097, 9009, 9008, 9024). Student #9005 changed the estimate for the relative scale of the human (originally one billion) after being asked if the relative scale estimates were related to the absolute size estimates:

> I: Are the numbers up here related to the numbers down here?
> R: Kind of. How many mm in a foot?
> I: Around 300.
> R: So 300 x 5…
> I: There's a calculator here.
> R: I was way off. 3000 times, pinhead to a human would be around 3000 times (changes answer).

The coordination of actions based on hindsight, or "retroaction", was identified by Inhelder and Piaget (1969) as a step in developing coordination between classification (grouping) and seriation (ordering), and can be seen occurring in these cases as well.

Other students also attempted to generate absolute sizes based on their relative scale estimates but were unsuccessful due to problems converting from one numerical format to another. Student #9037 estimated that the atom was one billion times smaller and then estimated an absolute size of 10^{-24} m. When probed as to whether these two numbers were related, he said, "They should, ideally 10^{-24} m would be one billion times smaller than a mm." Upon being asked if this was the case, he answered. "Why not?" Another student

(#0087) had a very similar explanation:

> R: [Absolute size and relative scale] correspond, but I personally, like they go together and everything. They have something to do with each other...I put from a pinhead to earth I put 1 mm and a million kilometers, and I said billions cubed [for relative scale], so…
> I: So, is this [1 million km] equivalent to billions cubed mm?
> R: [Laughs], yeah, I hope so!

Despite the math standards' (NCTM, 1989) expectation that middle school students be able to convert between units (within a single system – measurement 2, 6-8), we found that many students even in high school could not do so.

While any size can be expressed in any unit, expressing the size of submacroscopic objects in the units that students tend to know requires the ability to write very small numbers between zero and one. Many students struggled to find fractions, decimals, or powers of 10 to express the size of these objects in millimeters, centimeters, or inches. Student #0033 estimated that the Earth was 100 trillion times larger than the pinhead, and estimated an absolute size of 100 trillion mm, but had trouble with the smaller objects:

> R: I don't know the rest [sizes for cell, atom]. I can't even guess.
> I: OK. Umm, do you think you could use these numbers here along with the measurement of the pinhead to think about some of these sizes? (90 s pause)
> I: You said for instance that the cell was 5 or 6 times smaller than the pinhead.
> R: (sighs). Yeah. I don't know how to put that, though.

Student #0088 estimated that the atom was 5 million times smaller than the pinhead, and stated that relative scale and absolute size were related, but estimated the absolute size of the atom as 5×10^{-6} (instead of $1/5 \times 10^{-6}$, or 2×10^{-7}, which would be the inverse of 5 million). Student #0046 used *negative* numbers of millimeters for sizes smaller than 1 mm. Student #0047 explicitly considered her relative scale estimate for the cell, of one million times smaller than a pinhead, in producing her absolute size estimate – which was 1,000,000 mm rather than one-millionth of a millimeter. These cases all demonstrate the importance of mathematical content knowledge in these size and scale tasks. More specifically, using the connection between absolute size and relative scale may require

the ability to calculate the inverse of a number and convert different formats of numbers or units. The conceptual understanding that relative scale and absolute size are connected is also required before students attempt to calculate absolute size from relative scale (or vice versa). If middle school students were in fact able to understand that numbers can be written in different formats (*Benchmarks* 9A/2, 6-8; AAAS, 1993) and work flexibly with fractions and decimals (NCTM number and operation 1, 6-8; number and operation 2, 6-8; NCTM, 1989), then they would have some of the tools required for consistency between absolute size and relative scale. However, these mathematical tools are often not in place. This is effectively a bottleneck obstructing students' learning about the absolute size of potential landmark objects. Students will thus have to learn to write very small numbers or learn about micrometers and nanometers, as these units will allow them to express very small sizes with more convenient and familiar numbers.

Strategy 3: absolute size by measurement.

Physically measuring objects would only have been feasible with the height of the human, as it was the only object physically present. Since no measurement devices were available, students would have had to visualize the measurement instrument. A few students compared the height of a human to their recollection of a yardstick, following the strategy called mental instrument application by Joram and colleagues (1998).

Strategy 4: absolute size through direct iteration.

This strategy was not used observed.

Strategy 5: absolute size through indirect iteration.

This strategy was not observed.

Strategy 6: absolute size of objects through estimation and magnification factor.

While the red blood cell could have been estimated in this fashion by the summer campers, by recalling the apparent diameter when viewed through an optical microscope and dividing by the magnification, this strategy was not observed. The same strategy could have been employed by recalling the apparent size of objects in the computer simulation or million to one scale activity and dividing by the magnification factor;

166

however this strategy was not observed either.

Strategy 7: absolute size using the known absolute size of larger and smaller objects.

This strategy was not observed.

Strategy 8: absolute size through tool range.

The strategy was not observed.

Additional strategies.

Some students specifically did not generate consistent estimates for relative scale and absolute size as a guessing strategy in an attempt to maximize their possibilities of being right (given that there is literally an infinite number of possible estimates, the strategy does not actually work). Student #9032 stated her strategy as: "So I figured I'd have two guesses, and multiply my chances of being correct." Student #9022 pursued a similar, two-guess strategy in order "to better the chance of one being right".

Some students (e.g., #1004) first estimated the absolute size of the cell, and then simply tried to guess *any* smaller size for the atom. Some of these students had trouble doing this, however:

> R: What about if I put 0.5 mm [for the atom]?... It's about the same [size as the cell – estimated at 0.5 mm]. Maybe a little smaller but I don't know what to put for smaller.
> I: So you don't know how to put a smaller number?
> R: Would it be like 3 or something? No, that'd be bigger.

Summary and discussion of empirical task analysis for absolute size.

Many students used the strategy of recalling absolute sizes for objects that are landmarks for them. Some students did not understand the relationship between absolute size and relative scale at a conceptual level, and thus did not attempt to use the strategies (2, 4, 5, and 6) that utilize this connection. Other students attempted to do so but were stymied by lack of the mathematical content knowledge required, including the ability to calculate

the inverse of a number, and to convert among units or formats of numbers. Students were more likely to use strategies involving the connection between absolute size and relative scale for *well-known* objects. This study thus describes how landmarks are actually involved in building the relational web hypothesized by others (Tretter, Jones, Andre, et al., 2006).

Some students who lacked knowledge to guide their estimation of the size of an atom nevertheless understood the need for them to estimate a smaller absolute size than for the cell, that is, for order-absolute size consistency. Some of these students failed to be consistent due to insufficient mathematical content knowledge of fractions, decimals, or other ways of expressing small numbers between 0 and 1. The tasks generated the need for such numbers because the reference object used is 1 mm in diameter. Alternatively, students could have used units like the micrometer and nanometer that would allow them to use whole numbers instead, but only a few students in addition to the undergraduates and summer campers post-camp knew of such units (see Chapters 4 and 5). A large number of students in Study One asked the interviewer for smaller units, pointing to their conceptual importance in students' thinking (we did not provide these).

Students did not resort to the strategy involving the estimation of apparent size of a magnified object. For students to use the strategy successfully, they would need to conceptually grasp that this connection is possible, and also be able to estimate the apparent size of an object. My study did not allow me to distinguish between these two possible reasons for students not using the strategy.

Neither the "squeezing" (Joram et al., 1998) strategy that relies on the order-absolute connection, nor the tool range strategy that employs the group-absolute size connection, were used. As I show below, these relationships may be more useful for ordering and grouping than for absolute size.

Theoretical Task Analysis

As mentioned in chapter 2, relative scale is the characterization of the size of one object in terms of a reference object, using "iteration" – placing a unit end to end with no gaps or overlaps to cover a length (Wiedtke, 1990; Lehrer, 2003; Clements & Stephan, 2004). Figure 31 shows that relative scale tasks can be approached in several ways. Unlike absolute size, relative scale involves two objects. Vergnaud (1988) reported that not all students at the end of elementary school understand expressions like "three times more", or "three times less", often not realizing that these are of a multiplicative nature (p. 156). Thus, we cannot take for granted that students will understand the very idea of relative scale.

Strategy 1: relative scale by recall.

This strategy, highlighted in yellow at the upper left in Figure 31, consists of simply recalling relative scale, for example, that the thickness of DNA is approximately equivalent to 20 carbon atoms or that the diameter of the Sun is roughly 100 times that of the Earth. This strategy is similar to the second strategy for estimating absolute size, "absolute size through known relative scale", but lacks the second step of multiplying by the known absolute size of the referenced object. This strategy does not depend on any connection to other aspects.

Strategy 2: relative scale through absolute size.

This strategy culminates in the blue highlighted oval at the bottom left of Figure 31, and relies on calculating relative scale by dividing one previously known or measured absolute size by another. This strategy may involve conversions between units and between formats of numbers, as described above for ordering. For example, knowing that a red blood cell is around 7 μm in diameter and a pinhead around 1 mm (1000 μm) allows us to calculate that the red blood cell is around 140 times smaller than the head of a pin (1000 μm /7 μm = 143). This strategy depends on the connection between relative scale and absolute size. A "small minority" of middle and high school students in one study attempted to use known absolute sizes to calculate relative scales (e.g., the distance
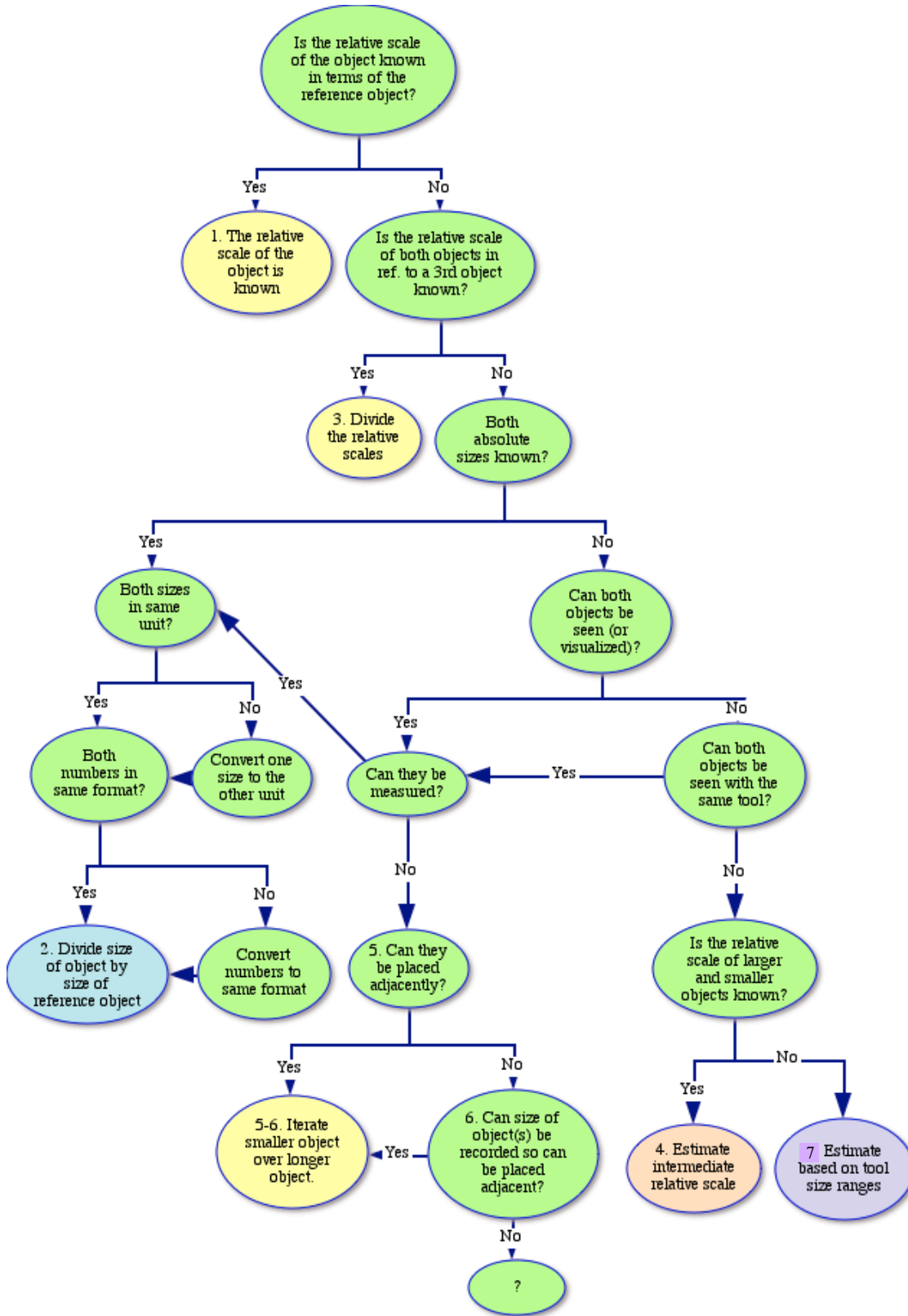
Legend: Ordering Pink, Grouping Purple, Relative Scale Yellow, Absolute Size Blue

Figure 31: Flowchart for strategies for relative scale

170

from California to North Carolina in body lengths – Tretter, Jones, & Minogue, 2006).

Strategy 3: relative scale using the known relation to a third object.

This strategy is depicted in the central part of Figure 31, in yellow. The strategy depends on the relative scale of both objects being known in terms of a third object. For instance, one could use hand spans to measure both the width of a chair and a conference table in order to calculate how many chairs can be placed at the conference table. One would simply divide the number of hand spans for the conference table by the number of hand spans per chair. A similar strategy could be employed when looking at two different slides through the microscope, one slide each for two objects of interest (e.g., white blood cell, Ebola virus) and getting an idea of their relative scale by comparing both to an object present in both slides (e.g., a red blood cell). This strategy does not depend on any connection to other aspects.

Strategy 4: relative scale using the known relative scale of larger and smaller objects.

This strategy culminates in the oval highlighted in pink at the bottom of Figure 31. It depends on knowing the relative scale of larger and smaller objects, and then estimating an intermediate value. For instance, if one knows that a buckyball diameter is equivalent to around seven carbon atoms, and the thickness of DNA to around 20 carbon atoms, and furthermore that a single-walled nanotube is intermediate in thickness between a buckyball and DNA, then one can estimate a value in between seven and 20 carbon atoms for the nanotube. The strategy would be very similar to the "squeezing" strategy for absolute size estimation (Joram et al., 1998). This strategy depends on the connection between ordering and relative scale.

Strategy 5: relative scale through direct iteration.

This strategy culminates in the oval highlighted in yellow in the central bottom part of Figure 31. The scale of one object relative to another can be obtained directly by repeatedly placing the smaller object with no gaps or overlaps (i.e., iterating) in a straight line to extend the same length as the longer object. This is identical to iteration in absolute size measurement (Wiedtke, 1990; Lehrer, 2003; Clements & Stephan, 2004)

171

but uses an object rather than a conventionally defined unit of measurement. This strategy does not depend on any connection to other aspects.

Strategy 6: relative scale through indirect iteration.

This strategy culminates in the same oval as for strategy 5, but follows a different pathway. If the two objects cannot be placed adjacently (for instance, if they are fixed in position), the length of one (or both) can be recorded onto a medium that can be placed adjacently, for instance, by marking the beginning and end point onto a string. This strategy depends on iteration (Wiedtke, 1990; Lehrer, 2003; Clements & Stephan, 2004) as well as conservation of length (Piaget & Inhelder, 1971; Piaget et al., 1960). This strategy does not depend on any connection to other aspects.

Strategy 7: relative scale through tool ranges.

This strategy ends in the purple oval at the right of Figure 31. If one object is known to be visible through an optical microscope (but not with the naked eye), and a second object is known to be visible through an electron microscope but not an optical microscope, then a very rough estimate of their relative scale might be obtained by comparing the ranges of the two instruments (e.g., dividing the geometric means of both ranges). The strategy depends on the connection between grouping and relative scale, as we are obtaining information about relative scale through the group membership of an object (e.g., forming part of the group of objects that can be seen through an electron microscope but not an optical microscope).

Summary of theoretical task analysis for relative scale.

The relative scale of an object can be previously known, measured at the time, obtained from absolute sizes, or obtained from other relative scales. Obtaining the relative scale of the object of interest from absolute sizes requires the connection between absolute size and relative scale, and may involve conversions between units of formats of numbers. Low precision estimates of absolute size can be obtained through tool ranges when both objects cannot be seen with the same instrument, or by squeezing (Joram et al., 1998). These two last strategies employ the connections between relative scale and grouping and between relative scale and ordering.

After this theoretical task analysis, I present my empirical findings for the specific relative scale task I employed in my studies.

Empirical Task Analysis

Before discussing students' strategies for relative scale, is important to note that several students seemed uncomfortable or unsure about relative scale as a construct. For instance, student #9002 commented: "I'm not sure how much times, like if you say, like 3 times the size of that, I'm not sure how much you're talking about…" Another student (#0094) entirely bypassed the issue of relative scale:

> I: How many times smaller is the atom than the pinhead?
> R: One size smaller…like, like, between, you see the size of this, maybe go down another size, might come smaller than that.

A third student (#0144) stated that asking how many times bigger the Earth is than the head of a pin was "Not fair…There's no way you could possibly figure it out!" Another student (#0097) was used to thinking about actual size, not how many times larger or smaller objects are relative to pin, and stated that it would be easier the other way around (to first define actual size, then figure out relative scale). These difficulties are consistent with Vergnaud's (1988) finding that students at the end of elementary may not understand expressions like "three times more". Students often redefined for themselves the relative scale task in terms of iteration, or how many objects "lined up" would fit across another object; the interviewers also resorted to this explanation when asked for clarification. Iteration seemed more understandable to students than the expression "how many times larger than…"

Using relative scale to introduce students to the size of objects has a superficial appeal, since research has shown that "relative" measures are easier for students than absolute measures (e.g., Vasilyeva & Huttenlocher, 2004) and relative scale does not require the use of units. However, these "relative" measures have usually involved ordering, not relative scale, and the prior research study that examined *quantitative* relative and

absolute measures did not find one type uniformly easier than the other (Tretter, Jones, & Minogue). The film "Powers of Ten" (Eames & Eames, 1977) is organized around relative scale (successive 10-fold changes in scale), and research has shown that viewing this film produced statistically significant increases in accuracy of ordering and ability to assign accurate sizes to objects (i.e., absolute size - Jones et al., 2007). Along with the unease these students expressed, it is worth examining in great detail whether relative scale or absolute size is a better way to introduce students to the size of objects. As mentioned in Chapter 5, the teaching experiment indicates that absolute size may be a promising way for students to learn about the submacroscopic world, if their instructional experiences also include learning about units.

Strategy 1: relative scale by recall.

This strategy was observed just once, as used by an expert (PhD in science) who employed a two-step strategy. First, he recalled that the distance from his nearest international airport to a major foreign city he visits frequently is roughly 1/3 around the world (relative scale by recall). Next, he used the known absolute size of that distance to calculate the circumference of the world and then the diameter, after which he calculated the relative scale of the Earth in relation to the pin (relative scale through absolute size, the next strategy discussed): "So, I think when I fly from X to Y, it's about 1/3 of the way around the planet, and I get 6000 frequent flyer miles. So I estimated the circumference of the earth and divided by pi, and then just converted it."

Strategy 2: relative scale through absolute size.

Several students and both of the experts used this strategy. A representative response is that of student #0099:

> I: Can you estimate how many times bigger is the height of the human than the diameter of the head of a pin?
> R: Hopefully a good estimate. So, 6 feet equals two meters, so if that's one millimeter…so I'm saying six feet equals about two meters, and so then, how many millimeters are in a meter. There's a hundred centimeters, er, thirty centimeters in a foot, so thirty times ten, so there's 300 mm in a foot, I think. So then 300 times six. So then 1800 millimeters in six feet…OK, so assuming all this math is right [so early] in the morning, I say it's like 1800 times…

As mentioned above, however, many students did *not* use this strategy as they thought relative scale and absolute size were not connected.

The role of mathematical content knowledge in this strategy was crucial. Student #1006 thought there was no number large enough for the relative scale of the Earth compared to the pinhead: "It'd be like infinity, 'cause there isn't any known number for the size of the pinhead to compare to the size of the Earth. Eventually there'll be a number." Similarly, student #1008 said of the relative scale of the Earth: "Too big of a number. Like, not a number I know." (The actual number is 13 billion.) Student #0081 thought that the atom might be more than trillions of times smaller than the pinhead, and she might not know a number large enough. Student #0080 apparently had trouble multiplying 100 by 500 – or else did not understand the multiplicative relationship between two scale factors, as can be seen in this excerpt:

> I: How many times bigger do you think a pinhead is than a red blood cell?
> R: About 500 times bigger
> I: Then the same thing with the atom
> R: I'd say probably like 100 times bigger.
> I: 100? 100 times?
> R: Oh, oh, than the, smaller…
> I: You said that the cell is 500x smaller
> R: Oh, I thought you said this [atom] to this [cell].
> I: Oh, sorry, sorry, so these two [atom to pinhead]
> R: Probably like 5000.

Strategy 3: relative scale using the known relation to a third object.

This strategy was observed in one student, #0080, who expressed the sizes of objects in atoms – far from accurately:

> R: Well, an atom would be one, and how many atoms make up each thing [inaudible]
> I: OK, so, like one atom makes up an atom, and then like how many atoms make up a cell…

175

R: Like 10 atoms make up a cell, and like 100 a pin head, and probably like 1000, or probably way more than that…
I: So the numbers below the line are the numbers of atoms…
R: That would make up that.
I: That would make up whatever.

Also, student #0052 compared the size of both mitochondrion and red blood cell to a white blood cell, but in qualitative terms, for ordering (see above).

Strategy 4: relative scale using the known relative scale of larger and smaller objects.

This strategy was not observed.

Strategy 5: relative scale through direct iteration.

The only object for which this strategy was possible was the height of the human relative to the pinhead. No student attempted to actually iterate the pinhead across even a portion of their body, although several students did so through visualization, usually with inaccurate results. Student #9027 stated: "I mean that's one [height of human] you can actually physically think of, 'cause you can picture how many pinheads make up 6 ft." Student #9036 claimed to have used the strategy for all of the objects:

I: And how did you come up with these numbers, with your estimates of how many times bigger these things are.
R: I just, I just like imagined. Like with the human, I just imagined like how many pins could fit on a human being. I'd just go by that.

Similarly, students #9034 and 9027 explicitly mentioned thinking of pinheads stacked on top of each other to constitute the height of the human. One student (#1006) employed this strategy iteratively in estimating relative scale of human compared to the pinhead:

Probably like a million times larger. 'Cause I was comparing the size of the pin to my thumb, and that'd be about like 100 pins, then to make up my whole thumb instead of just the edge would be maybe like 300, and then for my palm it'd be like 1000, then to make my whole body it'd be one million.

176

The student estimated reasonable factors for width of thumb to length of thumb ("edge" to "whole thumb") and length of thumb to palm, but greatly overestimated the number of pinheads that would fit across the thumb (10-20 rather than 100), or palms that would make up the body length (around 10, rather than 1000).

Strategy 6: relative scale through indirect iteration.

The strategy was not observed in the sample.

Strategy 7: relative scale through tool ranges.

The strategy was not observed.

Additional strategies.

As mentioned in strategy 4 above, one student (#9036) claimed to visualize the relative scale of Earth, red blood cell, and atom. Another student (#9037) used an analogical approach, explicitly stating the belief that the relative scale of atom to pinhead is equal to the relative scale of Earth to pinhead (both one billion).

Summary and discussion of empirical task analysis for relative scale.

A smaller number of strategies were actually used for relative scale than for absolute size. Direct iteration is not very feasible with submacroscopic objects, and recall strategies were much more common for absolute size than for relative scale. The definition of reference points stresses absolute size (Joram et al., 1998), not relative scale. In sum, the most common and practical strategy for relative scale of submacroscopic objects is to calculate it through absolute size – though few students used this strategy. These calculations may involve conversions of units or formats of numbers.

<div align="center">

Ordering

</div>

Theoretical Task Analysis

As mentioned in chapter 2, ordering "is the product of a set of asymmetrical transitive relations connected in series" (Inhelder & Piaget, 1969, pp. 5-6). This means that ordering is composed of successive comparisons of pairs of objects, establishing for instance that A is smaller than B, and (separately), that B is smaller than C. By

transitivity, A is also smaller than C: If A < B and B < C, then A < C. Through this serial process, the order A < B < C is established. Another way of thinking about these same inequality relationships is that if object B is larger than A but smaller than C, then B must be placed between A and C: if A < B and B < C, then A < B < C. Figure 32 shows that the asymmetrical relation between (i.e., comparison of) two objects can be approached in several ways, described below. For all of these strategies, however, students will need to compose the ordering through the use of transitivity.

Strategy 1: ordering by recall.

This strategy culminates in the pink oval at the top left corner, and is simply to recall that one object is larger or smaller than the other. The pink color indicates that this strategy involves the qualitative comparison of two objects that is the component process of ordering. Thus, it does not depend on the connection between aspects of size and scale.

Strategy 2: ordering by relative scale in relation to a third object.

Another strategy is to recall or measure the relative scale of both objects in relation to a third object, as we might do with paces or hand spans in determining whether a piece of furniture will fit through a doorway. This strategy ends in the yellow oval at the top left Figure 32. This strategy is identical in principle to strategy 3 employing absolute sizes, except that the size of both objects is expressed in terms of a reference object instead of a conventional unit of measurement. In the special case that one object is known to be smaller and the other larger than a third object, then the relative scales are unnecessary. This strategy depends on the connection between relative scale and ordering.

Strategy 3: ordering through absolute size.

This strategy for determining culminates in the blue oval at the left of Figure 32. The absolute sizes may be previously known (pathway on the left), or measured at the time (pathway on the right; the measurement can be carried out physically or mentally – see Joram et al. 1998). Comparing two absolute sizes may involve conversion of units for
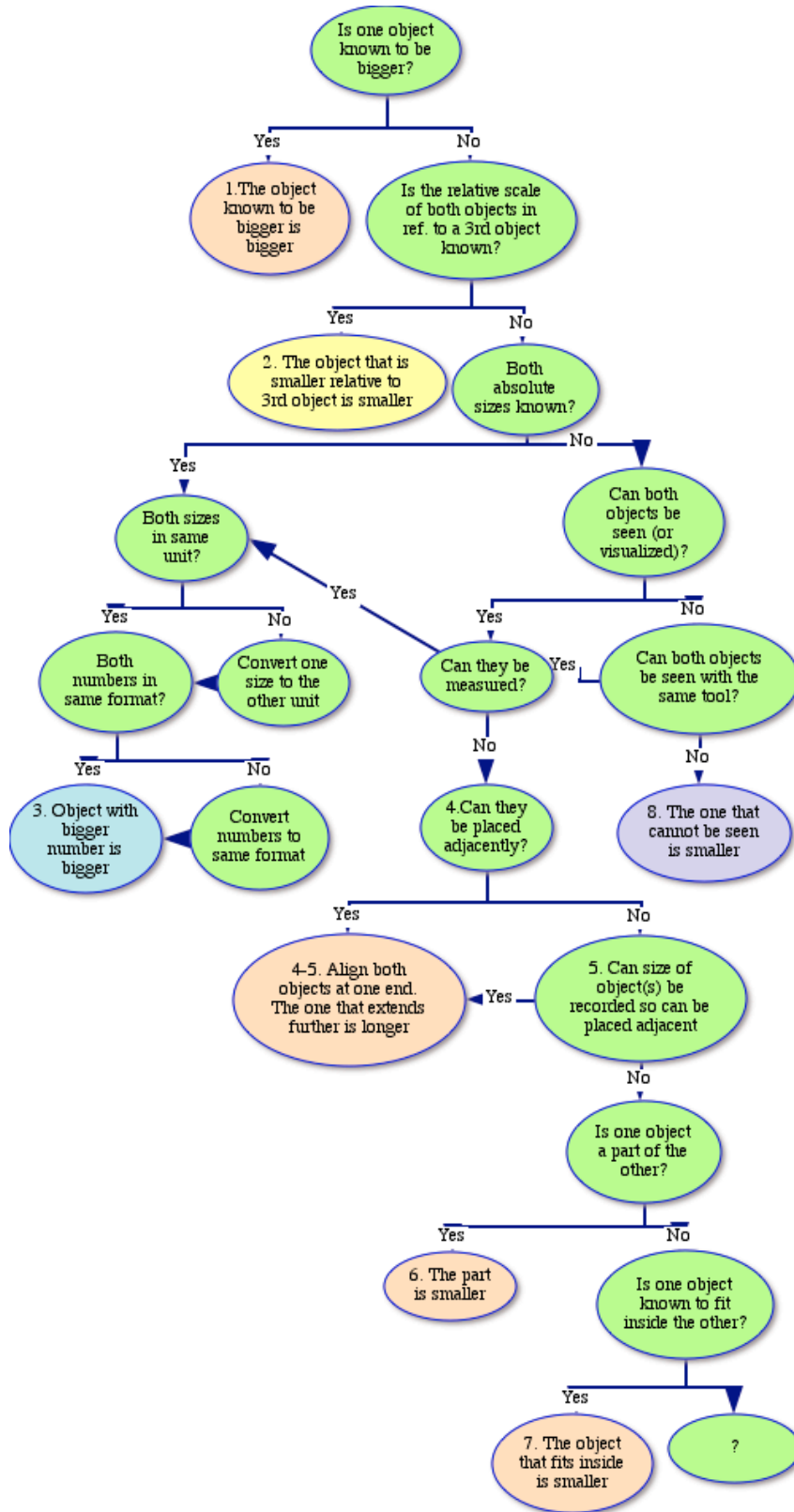
Legend: Ordering Pink, Grouping Purple, Relative Scale Yellow, Absolute Size Blue

Figure 32: Flowchart for strategies for comparison, component of ordering

ease of comparison; for instance, when comparing the height of a human (1.8 m) and the diameter of a pinhead (1 mm), the height of a human can be expressed as 1800 mm. Similarly, converting numbers to the same format can aid comparison, e.g., converting the 1 mm pinhead diameter to 10^{-3} m in order to compare to the 10^{-10} m diameter of an atom. Clearly, a student must be able to detect which of two numbers is larger, as well. This strategy depends on the connection between ordering and absolute size.

Strategy 4: ordering by direct comparison of two objects.

If both objects are present, a direct comparison of the two objects can establish which of the two objects is larger. This strategy culminates in the pink oval in the central part of Figure 32, and does not require connection to any other aspect of size and scale.

Strategy 5: ordering by indirect comparison of two objects.

An indirect comparison of two objects using the representation of one or both (e.g., two marks on a length of string denoting the length of an object) can establish which is larger. This strategy culminates in the same pink oval in the central part as strategy 4, but through a different pathway. It does not require the connection to any other aspect of size and scale. Conservation of length (Piaget & Inhelder, 1971; Piaget et al., 1960) is required to successfully carry out this strategy.

Strategy 6: ordering through part-whole relationship.

Comparison of the two objects is possible if there is a part-whole relationship. This strategy involves only the qualitative comparison of two objects, not requiring the connection to other aspects, and is therefore shown in pink, at the bottom of Figure 32. Students need to know that the part is smaller than the whole. Piaget showed that young children initially do not understand that the part is smaller than the whole, but do so by the time "sub-logical operations" are in place (Piaget & Inhelder, 1971), prior to the middle school years. This seemingly obvious relationship should be in place in the students in my studies, but cannot be taken for granted.

Strategy 7: ordering through the relationship of enclosure.

Another strategy to determine which of two objects is larger is to use a known

relationship where one smaller object fits inside a larger object, or "enclosure" (Piaget & Inhelder, 1971). The strategy involves only the qualitative comparison of two objects, does not depend on the connection to any other aspect, and is shown in pink at the bottom of Figure 32. DeLoache and colleagues (DeLoache, Uttal, & Rosengren, 2004) have demonstrated that young children may treat model objects (e.g., toy cars) as the real, life-size objects they portray, and try to get inside a toy car or don a doll's shoe. These errors peaked by age 2 and had decreased by 2 ½ yrs of age. This seemingly obvious relationship should be understood by the students in my studies, but cannot be taken for granted.

Strategy 8: ordering by tool used to visualize two objects.

The final strategy generated in this task analysis and included in Figure 32 relies on the tools that we employ to visualize two different objects. For instance, if we know that a red blood cell can be seen with an optical microscope but a molecule of water cannot, then we can determine that the red blood cell is larger than the molecule. This strategy, employing the connection between grouping and ordering, ends in the purple oval at the right of Figure 32. The red blood cell is one of many objects in the group of objects that can be seen using an optical microscope, whereas the molecule is not in this group, showing that grouping is involved in this strategy. Obviously, if both objects can be seen with the same tool then this strategy will not be useful. Students additionally need to know the order of increasing magnification capable with different instruments: naked eye, magnifying lens, optical microscope, electron or atomic force microscopes, and so on.

Summary of theoretical task analysis for ordering.

Ordering is composed of successive comparisons of two objects. Comparison can be obtained by recall, by direct comparison, or through a third object, without resorting to other aspects. Comparison can also depend on the connection between ordering and absolute size, or the connection between ordering and grouping (ordering by tool). Relationships between objects can also allow for comparison, specifically the part-whole relationship and enclosure.

After this theoretical task analysis, I present my empirical findings for the specific ordering task I employed in my studies.

Empirical Task Analysis

I examined students' actual strategies in ordering in order to complement the theoretical task analysis.

Strategy 1: ordering by recall.

Recalling that one object is larger than another was routinely used by students with the macroscopic objects, e.g., they knew humans to be smaller than mountains or the Earth. For instance, student #0109 said he would be "freaked out" if he saw an ant the size of a human, or a human the size of a mountain. Some students also used the strategy for submacroscopic objects. Student #3002 recalled a diagram in a textbook showing a virus injecting DNA into a cell, where the mitochondrion was larger than the virus. This strategy will result in incorrect ordering if the recalled information is incorrect; for instance, student #1004 ordered virus larger than cell because "in a virus there's lots of cells".

Strategy 2: ordering by relative scale in relation to a third object.

This strategy was not very common, but student #9010 used it to order the virus and the mitochondrion, saying that both are smaller than the cell, but viruses are a lot smaller than cells (implying that the viruses are smaller by a larger relative scale than mitochondria, relative to cells). Another student (#0052) ordered the mitochondrion smaller than the red blood cell by comparing both to a white blood cell. I did not observe incorrect applications of this strategy.

Strategy 3: ordering through absolute size.

Determining which of two objects is larger explicitly through the use of absolute size was not observed in the sample.

Strategy 4: ordering by direct comparison of two objects.

Strategy 4 involves the direct comparison of two objects. The strategy was not appropriate in our task except in the case of the pinhead and the human, which were the

182

only two objects physically present and visible all at once. It was not clear whether students ordered human larger than pinhead by simply recalling (strategy 1) or by visually comparing (strategy 6).

Strategy 5: ordering by indirect comparison of two objects.

Strategy 5 involves the indirect comparison of two objects. The strategy was not appropriate in our task except in the case of the pinhead and the human, which were the only two objects physically present. Several students attempted to use their fingers to compare the sizes of the *images* before we reminded them to order by the size of the represented *object*.

Strategy 6: ordering through part-whole relationship.

This strategy was widespread in ordering atom relative to molecule and mitochondrion relative to the red blood cell. For instance, student #1005 said, "I think the mitochondria is a part of the virus and the red blood cell. To be a part of something, it has to be smaller…", and student 9022 explained that cells are larger than molecules because cells have more than one molecule in them. Students also used a part-whole relationship to justify ordering the mountain as smaller than the Earth. Several students explicitly stated the part-whole relationship correctly but still ordered incorrectly. For instance, student #1004, who ordered cell < atom < molecule < virus < mitochondrion, said: "Mitochondria is, like, in the cell thing, so it's like smaller…in a virus there's lots of cells…a molecule is part of a cell…then a red blood cell, I thought that was the smallest." This finding also suggests that the basis of ordering, which is the successive comparison of two objects at a time and the building up of an order through transitivity, may need to be included in middle school curricula for the occasional student who is at a point in development where he or she has not yet consolidated this idea.

Strategy 7: ordering through the relationship of enclosure.

This strategy was used by several students to order virus in comparison to cell, noting for instance that "Viruses have to be small enough to infect blood cells" (#1007), and by students who stated that red blood cells are little (smaller than humans, at least) because they "travel through your body" (#9021) or have to fit into small veins (#0082). This

strategy was almost invariably used correctly, unlike strategies 3 (part-whole) or 1 (recall). However, one student (#0094) appears to have used this strategy incorrectly. The student responded to the prompt for the smallest object of which she could think with the Periodic Table, and the interviewer clarified that we were looking for actual objects, and provided the example that some students mentioned ants.

> I: Can you think of anything too small to see with the naked eye…
> R: What you cannot see that is too small to see, is you know how the ants, they make their little sand thingy, it's almost like their little home, and it's built up out of sand, yeah, you can't see…

The student thus appears to say that the ant's house (inside of which an ant can fit) is smaller than an ant. The student presented unusual responses on several tasks. Other than this one exception, though, this strategy was frequently and correctly employed.

Strategy 8: ordering by tool used to visualize two objects.

This strategy was used extensively by students to create two groups: macroscopic and submacroscopic. After this, they ordered within each group. Less frequently, students used the tool strategy to locate a specific object. For instance, student #1014 distinguished macroscopic objects from the cell by stating that you need a magnifying lens to see a cell. This strategy depends on accurate content knowledge about tools, something that was not always true of students. For instance, student #0006 thought that a molecule could be seen without a microscope ("sometimes").

Additional strategies.

Student #9001 noted that blood is liquid so cells must be very small. This is the use of a physical characteristic that bears roughly on size. Student #9036 used a deductive strategy to order atoms smaller than molecules, saying that they must be smaller than molecules because an atom is the smallest thing. Student #9031 thought the cell was smaller than a virus, because a virus attacks a blood cell by eating it. These three strategies are not robust or generally applicable.

Summary and discussion of empirical task analysis for ordering.

The most common strategies for ordering involved recall and relationships (part-whole, enclosure). Using tool range as a first step in ordering was also common. The quantitative aspects of size and scale were not used commonly in ordering. However, the widely reported confusion about the relative scale of atoms and cells might be mitigated through this strategy. The status of atom and cell as building blocks in chemistry and biology, the existence of components of each (e.g., both have a nucleus), and the fact that both are submacroscopic may lead to this confusion. Students might be creating disconnected, local ordering for different contexts that then have to be properly connected, perhaps through absolute size. Another possible way to connect local orderings is through interdisciplinary units that directly compare atoms and cells, as we used in the teaching experiment.

Grouping

Theoretical Task Analysis

As mentioned in chapter 2, grouping involves placing objects that are similar (e.g., in size) into a group, and objects that are different into different groups (Inhelder & Piaget, 1969). There may be several satisfactory ways to group objects by size, using different numbers of groups or criteria for boundaries between groups. Figure 33 shows that grouping objects can be approached in several ways. Strategies 1-4 depend on knowing the relative scale or absolute sizes of all objects, whereas strategies 5-6 do not. The diagram for grouping could thus be very large and complex, if relative scale and absolute size were included in this diagram. Instead, for clarity, I begin with known relative scales or absolute sizes for strategies 1-4.

Strategy 1: grouping by absolute size ranges.

Strategy 1 culminates in the oval highlighted in blue at the upper left of Figure 33. If the absolute sizes of all objects are known, then each object can be assigned to a group once the size ranges for each group have been determined. My initial coding rubric for the smallest object employed this strategy to categorize student responses. This rubric contemplated five exhaustive, non—overlapping size categories of: > 100 μm (i.e.,
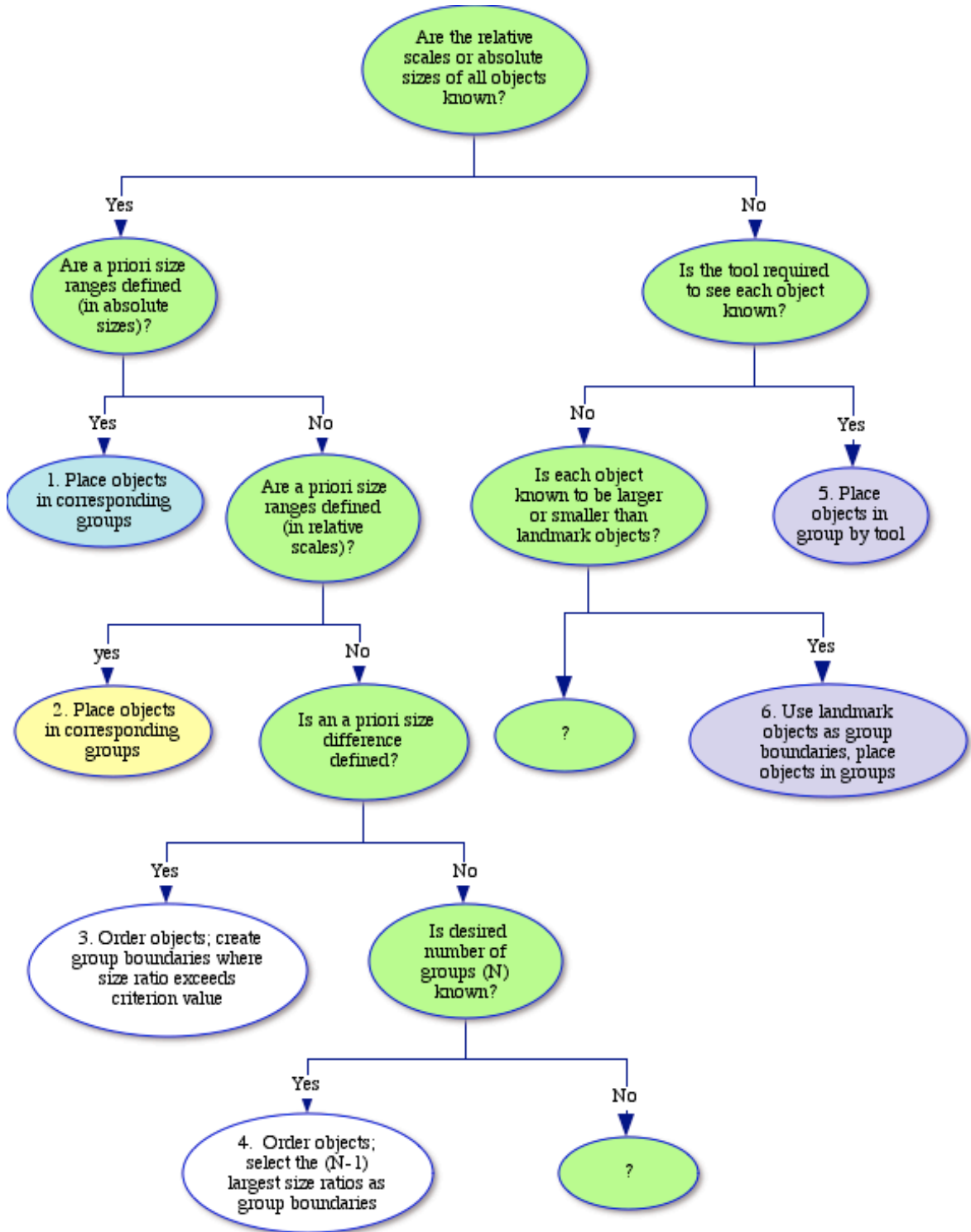
Legend: Ordering Pink, Grouping Purple, Relative Scale Yellow, Absolute Size Blue

Figure 33: Flowchart for strategies for grouping

macroscopic), 1-100 μm, 10-999 nm, 0.1-9.99 nm, and < 0.1 nm. The Scale of Objects Questionnaire Assessment used by Tretter and colleagues (Tretter, Jones, Andre, et al., 2006) requires respondents to employ this strategy. This strategy depends on the connection between grouping and absolute size.

Strategy 2: grouping by relative scale ranges.

This strategy, highlighted in yellow at the left of Figure 33, is similar to strategy 1 but uses relative scale. For instance, molecules could be classified (grouped) into the following categories: molecules that are between 1-10 diameters of a carbon atom in their largest dimension, those that are 10-100 diameters, 100-1000 diameters, and over 1000 diameters. This strategy depends on the connection between relative scale and grouping.

Strategy 3: grouping by criterion size ratio.

This strategy culminates in the white oval on the left of Figure 33. I use white to represent the fact that two aspects must be used in addition to grouping: ordering *and* relative scale or absolute size). First, the objects are ordered by absolute size or relative scale; then, we use absolute size or relative scale *ratios* between adjacent objects to determine group boundaries. For instance, consider eight objects with absolute sizes in millimeters (or relative scale, compared to a 1 mm pinhead) as follows:

2, 4, 15, 20, 220, 3400, 3800, and 40,000

If one used a criterion of a 10-fold size difference, then one would insert group boundaries as follows:

{2, 4, 15, 20}, {220}, {3400, 3800}, {40,000}.

If instead a criterion of a three-fold size difference were employed, the groups would be:

{2, 4}, {15, 20}, {220}, {3400, 3800}, {40,000}.

The "size" difference will be identical whether it is calculated using absolute size or relative scale, as it will result in the same unitless number in either case.

Strategy 4: grouping by number of groups desired.

This strategy is also highlighted in white, as it requires ordering *and* relative scale; it is shown on the bottom left of Figure 33. It relies on determining the desired number of groups (N). Then, the size difference between adjacent objects is calculated, and the N-1 largest relative scale differences are used as group boundaries. For instance, using the sizes above and three groups, one would create the following groups, with boundaries at the positions with greatest relative scale differences.

{2, 4, 25, 20}, {220}, {3400, 3800, 40,000}

See Table 22 for calculations of the relative scale differences: the two largest differences are 15.5 and 11. By splitting the ordered objects at these two positions, three groups are created.

Table 22: *Calculation of relative scale differences for hypothetical example*

Rel. Scale Diff.		4/2 =2		15/ 4= 3.75		20/ 15= 1.33		220/ 20= **11**		3400/ 220= **15.5**		3800/ 3400= 1.1		40,000/ 3800= 10.5	
Size (mm)	2		4		15		20		220		3400		3800		40,000

Strategy 5: grouping by tool.

If the relative or absolute sizes of the objects are not known, other strategies can still be employed to group objects by size. If the tool required to see each object is known, then a grouping can be made on this basis. This strategy is highlighted in purple, as it is a grouping strategy that relies only on grouping; it is shown on the right of Figure 33. Objects can be grouped by whether they can be seen with the naked eye, require an optical microscope to be seen, require an electron or atomic force microscope to be imaged, or are too small to be seen with any of these. For macroscopic objects, a

distinction may be drawn based on whether the objects can be seen all at once, or are too large. The grouping resulting from this strategy is approximate, because some objects can be visualized with more than one tool, or may be at the edge of the range of an instrument.

Strategy 6: grouping by comparison to landmark objects.

This strategy also depends on grouping only, and is shown in purple at the bottom right of Figure 33. The strategy depends on knowing whether each object is larger or smaller than one or more landmark objects (Tretter, Jones, Andre, et al., 2006) or reference points (Joram et al., 1998). A learner might thus create groups of objects that are smaller than an atom, objects intermediate in size between an atom and a cell, in between a cell and a pinhead, a pinhead and a human, a human and a planet, and larger than a planet.

Because there are many groupings possible for any given set of objects, I did not include a strategy that depended on recall of groups.

Summary of theoretical task analysis for grouping.

Grouping can be carried out by qualitative means: grouping by tool, and by the use of landmarks. Grouping by quantitative means involve a predetermined criterion: absolute size or relative scale ranges, number of groups, or size difference between adjacent objects to define group boundaries.

After this theoretical task analysis, I present my empirical findings for the specific absolute size task I employed in my studies.

Empirical Task Analysis

Strategies 1 through 4 depend on knowing the absolute sizes or relative scales of all the objects in order to group, and were not very common.

Strategy 1: grouping by absolute size ranges.

This strategy was used by student #3002 (an undergraduate), who grouped objects by units, employing nanometer, micrometer, millimeter, meter, kilometer, and thousands of

kilometers groups.

Strategy 2: grouping by relative scale ranges.

This strategy was not observed in the sample.

Strategy 3: grouping by criterion size ratio.

This strategy, establishing a relative scale difference as a criterion to determine group boundaries, was apparently used by one student (#9027) who said that the human was hard to place because the ratio of human to ant is smaller than the ratio of human to Earth. Another student (#9007) apparently used this strategy as well:

> Do the groups have to have more than one?... I think if I had my way I'd probably group them individually… it just seems that each one is so much bigger than the next, I mean, they're not close enough to be grouped.

Strategy 4: grouping by number of groups desired.

This strategy was not observed in the sample.

Strategy 5: grouping by tool.

This strategy was explicitly used by at least a third of the students. Many students, including some of the undergraduates, created just two groups, based on their visibility to the naked eye. Other students distinguished whether objects could be visualized with the unaided eye, with an optical microscope, or were too small to see even with an optical microscope. Some students made distinctions between macroscopic objects on the basis of whether they could be seen all at once (e.g., #9030). Around half of the students who used strategy 5 grouped incorrectly, in most cases due to creating groups among the submacroscopic objects that inappropriately excluded an object intermediate in size to two objects in the group. For example, student #0033 grouped atom and cell without also including the molecule, virus, and mitochondrion, which are intermediate in size. Some students omitted an object intentionally, as discussed below, in Additional Strategies.

Strategy 6: grouping by comparison to landmark objects.

Many students used this strategy. For example, some students grouped based on the cell,

creating cell, sub-cell, and larger than cell groups (9006, 0095); student #9018 created a group he called "atomic level" (9018); and many students used the human as a reference, making groups called "human and up" (#0093), or things inside (and thus smaller than) the body.

Additional strategies.

When asked to explain their grouping strategies, many students respond in ways that seem to show that they had used a combination of strategies 5 and 6. For instance, student #9009 called the group with cell and virus "things inside the body" (Strategy 6) and the group with pinhead and ant "small things that can still be seen" (Strategy 5).
 One student (#9005) included atom, molecule, and mitochondrion in a group called "microscopic", and placed cell and virus in a group called "in your body". The student noted that the groups "contradict" each other, apparently noting the tension between the two strategies.

Some students may have used strategies 5 or 6 but then characterized their groups by other descriptors. These group labels for atom and molecule included non-living, "building blocks", "matter", and chemistry objects. Labels for groups with virus, cell, and mitochondrion included living things, "structure", "disease", and biology objects. Another common way of characterizing the groups were simply by describing their relative position, e.g., largest, medium, and smallest objects groups.

The descriptors for groups that did not correspond to tools or landmark objects mentioned above may be a reflection of a criterion change to non-size criteria (such as living vs. non-living or biology vs. chemistry). Inhelder and Piaget noticed that many young children spontaneously switch criteria when asked to classify (1969, p. 285). Other researchers have pointed out that children may group items for reasons other than belonging to the same conceptual category, including perceptual similarity (e.g., color and form; Melkman, Tversky, & Baratz, 1981). Some fifth grade students who were asked to group by size a number of cards depicting objects grouped by similarity in wording (e.g., placing all the cards that included the word "diameter" in one group –

Tretter, Jones, Andre, et al., 2006). As I mentioned in Chapter 3, one student (#1001) grouped human and molecule together, saying that they may not be the exact same size but kind of similar and still go together, hinting strongly at the existence of a second criterion.

Previous research found that students of all ages estimating the size of objects exhibit a secondary dimension (after the primary dimension of the size of the objects) according to whether they have had direct experience with objects or not (Tretter, Jones, Andre, et al., 2006). The secondary dimension of direct experience was evident in some student answers, for instance, distinctions between "normal" versus "not everyday" (#9021, 9001), "real" things (macroscopic objects) vs. others (#9015), or objects whose size can be comprehended and with which one can interact (pinhead, ant, human) versus others (#9027). However, various students appeared to establish distinctions even among objects with which they do not have direct experience, by degree of familiarity. Some students explicitly placed the submacroscopic objects with which they were not familiar in a single group (e.g., #1014, 9027, 0097). This resulted in order-group inconsistency.

Finally, one student (#0089) had a very unusual approach to the grouping task: he placed the largest object with the smallest, the second largest with the second smallest, and so on. He explained that he was grouping by size "for balance".

Summary and discussion of empirical task analysis for ordering.

The empirical data showed that few students grouped based on knowledge of the absolute size or relative scale of the objects, but those who did so grouped accurately. Most students grouped by tool or comparison to landmark objects, but with low accuracy. Some students appeared to inappropriately mix strategies or even change criteria from size to unrelated distinctions such as living/non-living or biology/chemistry. I also detected a secondary dimension by degree of direct experience *or* degree of familiarity with objects that cannot be experienced directly, consistent with but extending prior research (Tretter, Jones, Andre, et al., 2006). It may thus be useful to include discussions of the nature of grouping and how a single criterion should be used consistently, in LP2.

Theoretical Task Analysis

The process for selecting the smallest object is identical to the component process for ordering presented above: the successive comparison of two objects.

Empirical Task Analysis

While it might initially seem that this could be a recall-level response that involves simply remembering an object, closer analysis reveals that comparison is required. For instance, a PhD in theoretical physics (#4001) had to compare various possibilities:

> I: Can you think of some very small things that you know of, please?
> R: (One second pause) Quantum dots.
> I: Okay. What is the very smallest thing that you can think of?
> R: Hmmm (4 secs long)(2 sec pause) that's very tricky. Ummm, I would say a quark, but nobody really knows how big a free quark is (laughs), so the smallest thing I'm reasonably sure of might be a neutrino, but then neutrinos spread out in space… So now to be really on the safe side, you know, I'd probably say a mu meson (laughs).

This interview question was posed before submacroscopic objects were presented to students in the context of the subsequent tasks. Students often did not initially think about submacroscopic objects; for instance, student #0144 said, "I have to think molecular, don't I? First I thought [of] a flea…" and then responded with atom and the nucleus of an atom.

Smallest Unit

Theoretical Task Analysis

The process for selecting the smallest object is identical to the component process for ordering presented above: the successive comparison of two objects (units).

Empirical Task Analysis

The process for selecting the smallest unit known also involves the comparison of known units. While many students knew that the metric system uses prefixes, and that the base

unit for length is the meter, many did not know of a unit smaller than the millimeter or centimeter; some believed that no units smaller than the millimeter existed. One student (#9032) used multiple "milli-" prefixes, end up with a unit written with six m's – a "milli-milli-milli-milli-millimeter". Several students noted that any size can be written in meters, with appropriate exponents for powers of ten, and some students used scientific notation or powers of ten in their answers (e.g., #9026, 9037, 0087, 0109).

General Discussion of Task Analysis

Tasks for every aspect can in principle be addressed by recall, by strategies particular to that aspect (e.g., measurements using conventional units for absolute size), and by strategies that depend on the connection between aspects (e.g., calculating relative scale from known absolute sizes). In both the theoretical and empirical analyses, I found that consistency between aspects could be used as a tool to help students solve tasks.

Absolute size is the most powerful of the four aspects. Every task can be approached through absolute sizes. For instance, in the strategy for relative scale used by experts, the learner recalls or estimates the absolute size of the objects of interest and the reference object, and divides. The quantitative aspects contain more information than the qualitative ones. For instance, absolute size can determine order exactly, but order can only provide an approximate estimate of the absolute size of an object. Furthermore, the standardization that conventional units provides eases communication about sizes (Lehrer, 2003). Strategies that use the connection between absolute size and another aspect proved difficult for students, even at a conceptual level – many students do not realize that there is a (necessary) connection between relative scale and absolute size. These strategies may require accurate conversion of units, different formats of numbers, and calculating inverses. Thus, there is a developmental component to size and scale in that some logical ideas (e.g., the part-whole relationship, conservation of length, transitivity) and some mathematical ideas (e.g., inverses, decimals or fractions) must be in place before students can succeed. However, the fact that these ideas have not been developed in some students in the middle school does not mean that size and scale is beyond their reach; it only means that the curriculum for size and scale may have to

revisit and reinforce – or even introduce for the first time - these important mathematical ideas.

I found that students are more likely to try to be consistent across relative scale and absolute size for objects whose absolute size was known to them. Thus, accurate factual knowledge *may* aid students in developing more connected knowledge of size and scale.

I also found some strategies that I had not contemplated, though these tended to be inappropriate or of low power and generalizability.

The theoretical and task analyses reported above show characterized the relationship between factual knowledge and consistency of knowledge one task at a time and in a qualitative way. These analyses showed that a relationship does exist, but does not quantify its importance. Therefore, I next assess the weight of the relationship between consistency of knowledge and accuracy of factual knowledge, using a correlational analysis.

Correlational Analysis of Factual Knowledge and Consistency - Study One

As I mentioned above, I found that the connections across aspects of size and scale can theoretically be used in addressing factual knowledge tasks, and that students in fact do rely on the connections in many cases. I also found that accurate factual knowledge might be a factor in leading students to greater consistency. In order to assess the weight of the relationship between consistency of knowledge and accuracy of factual knowledge, I conducted a correlational analysis on the data from Study One. I found a strong, statistically significant correlation between overall score for factual knowledge and consistency level of 0.70 (p = 0.01). (The correlation is calculated considering only those 90%+ students who fit along the progression; however, using the sum of consistency codes for all students results in a very similar correlation of 0.68, p = 0.01.).

In order to contextualize the strength of this correlation, I compared it to the correlation between factual knowledge and grade level. See Table 23. I expected the level of factual

knowledge to be correlated positively to grade level, as formal schooling is likely to be the main source of information about the size of objects like the atom, the cell, and the earth - and it was weakly to moderately correlated: 0.34 (p = 0.01). Consistency was also correlated in a weak but statistically significant manner to grade level: 0.29 (p = 0.01). Using a t-test, I found that the overall factual knowledge score and consistency level are correlated to each other significantly more strongly than either is to grade in school (0.70 vs. 0.34 grade-content, 0.29 grade-consistency, t = 3.82 and 4.43 respectively, p < 0.01 in both cases). Since the consistency level is measured independently of accuracy of factual knowledge, the correlation suggests that there may be some mechanism operating between the two measures.

The cross-sectional nature of the data collected in Study One provides no information about the relationship in terms of possible mechanisms or patterns of development. I next use the data from Study Two in a regression analysis to gauge the relative importance of factual knowledge in helping the development of consistency, and in consistency in helping the development of factual knowledge.

Table 23: *Pearson correlations among consistency, science course, grade, and overall factual knowledge score*

	Overall factual knowledge score	Consistency level	Grade
Overall factual knowledge score	--		
Consistency level	.70**	--	
Grade	0.34**	0.29**	--

**p < 0.01

Regression Analysis of Factual Knowledge and Consistency of Knowledge – Study Two

In this section, I examine the relationship between factual knowledge and consistency of knowledge using the data from the teaching experiment. The teaching experiment's pre- and post-camp interviews afford an analysis of how levels of factual knowledge and consistency of knowledge coming into camp affect subsequent learning of factual knowledge and consistency of knowledge.

The model predicting post-camp overall factual knowledge with pre-camp factual knowledge and consistency of knowledge was statistically significant, $F(18) = 8.87$, $p = 0.003$, and accounted for almost half of the variance (adjusted $R^2 = 0.47$). Pre-camp level of consistency is a significant predictor ($\beta = 0.655$, $p = 0.012$) while the pre-camp overall factual knowledge is not significant. On the other hand, the model predicting post-camp consistency level was not statistically significant.

These findings support the idea that having a more connected understanding of size and scale facilitates students' increase of factual knowledge about the size of objects. This is consistent with the task analysis presented earlier that described ways in which consistency across aspects can allow students to use additional strategies to estimate relative scale and absolute size, order, or group by size, making possible greater accuracy of their factual knowledge than without consistency.

The measurement of consistency across ordering and the other three aspects, and between relative scale and absolute size, does not depend on accuracy of factual knowledge. The empirical task analysis *did* find a role for accuracy of factual knowledge of absolute size in making students more apt to realize that relative scale-absolute size consistency is possible and desirable. This regression analysis did not detect that factual knowledge has a significant effect on consistency.

I examined histograms of pre- and post-camp levels of factual knowledge and consistency of knowledge for normality of distribution (an assumption of regression).

The post-camp level of factual knowledge appeared skewed or bimodal, so I generated a Q-Q plot to check for normality; the data points lie near the line, showing that normality was adequate. See Appendix K for the histograms of all four variables and Q-Q plot.

Limitations of the Analysis of the Relationship Between Factual Knowledge and Consistency of Knowledge

The theoretical task analysis generated some but perhaps not all strategies for each aspect of size and scale. Similarly, the empirical task analysis may not have revealed all the strategies students use. Furthermore, the actual strategy students used was not clear in some cases, particularly in grouping. Thus, further work will be required to address these uncertainties.

The correlational study compared the strength of the correlation between factual knowledge and consistency with the correlation of each with grade in school. However, this study did not consider some pertinent differences within students in a particular grade, such as the science or math classes they had experienced, or their performance in the class. Nevertheless, the much higher correlation between factual knowledge and consistency of knowledge does suggest a mechanism operating between consistency and factual knowledge that the regression analysis seems to characterize as the facilitative role of consistency in support accuracy of factual knowledge.

The small number of respondents in the regression analysis (24) resulted in low power, which may have led to the lack of statistical significance in the model predicting post-camp consistency level. The low number of respondents also made the visual tests for homoscedasticity and linearity of the relationship between independent and dependent variables (plotting expected values vs. residuals) inconclusive. Thus, the results should be seen as tentative.

General Summary and Discussion

This chapter bolstered my claim for the importance of the connections across aspects of size and scale. This claim was originally driven by theory, from the general constructivist

position that the connections between pieces of knowledge are important. In this chapter I outlined specific ways in which these connections can facilitate accuracy of content knowledge, and found evidence that suggests that they do so, by enabling additional strategies. If a learner cannot recall the factual knowledge, and cannot use a strategy that only employs that same aspect, she can resort to the links between aspects to obtain information about the aspect of interest. Realizing that the four aspects of size and scale are logically linked may make it easier to retain information about the size of specific objects by helping students to contextualize and store it in several ways. A student's consistency level predicts future factual knowledge learning, possibly by providing more avenues for the student to remember the piece of information.

What determines the growth of consistency? The results are less conclusive. Consistency does not depend on accuracy of factual knowledge (Vosniadou, 2003), but students sometimes try to be consistent across relative scale and absolute size when they are more certain about the absolute size of the object. However, a regression model using initial levels of consistency and content knowledge was not significant. This suggests that other variables may be involved as well; from the task analyses, it seems that one of these variables is mathematical and logical knowledge. Logical structures including conservation of length, transitivity, the part-whole relationship and the relationship of enclosure may be required. Mathematical content knowledge required includes the ability to calculate the inverse of a number, convert between units and between formats of numbers (e.g., between scientific notation and decimals), and the ability to express small numbers between 0 and 1. Knowing units smaller than the millimeter may eliminate the need for some of these mathematical skills. Understanding the foundational ideas of measurement underlies many strategies.

Unless science educators take on the responsibility of helping students construct the required mathematical and logical knowledge alongside (or prior to) the scientific knowledge related to size and scale, there will be a developmental component to learning about size and scale, as logical and mathematical skills develop. Fortunately, Chapter 5 shows that learning about units smaller than the millimeter can help eliminate or reduce

the need for some of these. It remains to be seen if the introduction of these units can help students learn fractions and decimals, or instead interferes with this learning.

This analysis further defines how landmarks can be used in establishing a relational web of size and scale (Tretter, Jones, Andre, et al., 2006). The empirical task analysis, correlational analysis, and regression analysis also reveal that the links across aspects of scale are important and empirically useful; this is a dimension of size and scale only briefly noted or studied by prior research (Batt et al., 2008; Tretter, Jones, & Minogue, 2006). These analyses again support Study Two's findings that logical and mathematical knowledge supports size and scale tasks, and adds to prior research, which had identified proportional reasoning (Tretter, Jones, Andre, et al., 2006).

The theoretical and empirical importance of consistency across aspects of scale, along with the finding of a strong and robust pattern in the development of consistency, justifies my decision to organize the learning progression for size and scale based on levels of consistency. The second iteration of the learning progression is presented next.

CHAPTER 7: A LEARNING PROGRESSION FOR SIZE AND SCALE

In the following section, I revise and augment LP1 on the basis of the Study Two teaching experiment and the task analyses. LP2 can help define a student's current level using the interview protocol described in Chapter 3. Future work will be required to develop a more practical assessment instrument that can reliably and validly assess a student's level. I suggest instructional activities that address learners at each level, with the expectation that this can help meet the learner at his or her level and will lead to more effective learning about size and scale.

Instructional sequences not based on a learning progression are less likely to result in effective learning, as they may fall outside a learner's zone of proximal development (Vygotsky, 1978, 1985). I found many instances of activities that did not result in learning for some students. However, almost all students advanced. This suggests that students learned what was closest to their current level of understanding but not what lay too far beyond. For instance, some students did not use the millimeter to express the size of tiny macroscopic objects and submacroscopic objects even after this unit was mentioned to them. Similarly, students at the lower levels on the learning progression did not use the newly introduced micrometer and nanometer to estimate the size of cells and atoms, while students at higher levels adopted the new units enthusiastically. No student advanced more than three levels during camp. Instruction thus needs to be closely fitted to students in order to achieve results. Fortunately, some activities we designed for the teaching experiment simultaneously address various levels (see Table 11, Chapter 5). Next, I provide a prose description of the levels in the learning progression. A more schematic and abbreviated form is included as Figure 34.
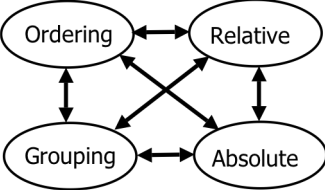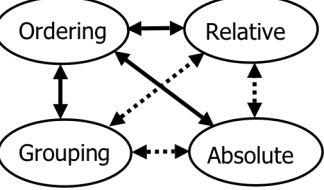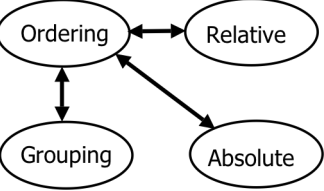
Level and Grade	Landmarks	Smallest object, unit	Logical/ mathematical skills	Strategies	Instruction
Level 5: End of Grade 12 *(Ordering, Relative, Grouping, Absolute diagram)*	Mental measurement line with atom, cell, small macroscopic object, possibly DNA, virus, bacteria, presumably Earth,	Sub-atomic particles. Micrometer, nanometer.	Proportional reasoning in all cases. Unit and number conversions.	Can group, order, estimate relative scale and absolute size for landmark objects, and objects with known relationship to these. Can group accurately and justify criteria.	Linear and logarithmic scales?
Level 4: End of Grade10 *(Ordering, Relative, Grouping, Absolute diagram)*	Can construct a mental measurement line with atom, cell, small macroscopic object (e.g., hair), presumably Earth	Atom. Micrometer, nanometer.	Proportional reasoning to connect relative scale and absolute size in some cases	Can construct a scale by use of landmarks and part-whole, enclosure, tools strategies.	Fractions, decimals, negative powers of ten
Level 3: End of Grade 8 *(Ordering, Relative, Grouping, Absolute diagram)*	Atom, cell, small macroscopic, human (absolute). Cell and human (relative). Clear separation of macroscopic and submacroscopic worlds	Atom. Micrometer, nanometer.	Begin to estimate size of objects intermediate between landmarks	Can order by known absolute sizes, or use "squeezing". Can estimate consistent sizes for 2 objects < 1 mm.	Proportional reasoning, how to construct a scale (interpolation using relative scale). Unit and/or number conversions

Figure 34: Learning progression for size and scale

Level and Grade	Landmarks	Smallest object, unit	Logical/ mathematical skills	Strategies	Instruction
Level 2: During Middle School Ordering ↔ Relative ↕ Grouping Absolute	Human (metric), cell (relative), atom < cell	Atom. Micrometer, nanometer	Unitizing	Can order A and B by reference of both to C. Begin to coordinate landmarks.	Units smaller than the millimeter; expressing the sizes of successively smaller objects.
Level 1: End of Grade 6 Ordering Relative ↕ Grouping Absolute	Human (English, possibly metric).	Germs, cells. Millimeter	Transitivity	Order objects by tool, part-whole, or enclosure. Begin to create worlds, incl. submacroscopic	Quantitative version of transitivity, inverse relationship between unit size and unit number, nature of relative scale, nature of grouping, units < mm
Level 0: Start of Grade 6 Ordering Relative Grouping Absolute	Human (English units)	Small macroscopic, inch or centimeter	Iteration, part-whole, enclosure	Can compare objects by part-whole or enclosure	Focus on making the submacroscopic world real; transitivity; the mm and relation to inch, cm

Figure 34 (cont'd): Learning progression for size and scale

Level 0: the Lower Anchor

Level 0 is what we can expect from incoming sixth grade students, although many will be more advanced. The findings from the task analyses and the Study Two teaching experiment now allow additional characterization of students at this level beyond that presented in Chapter 4 (summarized below for convenience).

> Students at this level do not connect any aspects of size and scale, as shown by their inconsistency across aspects. Their only size landmark is the height of the human in English units, and they do not know of submacroscopic objects. They can usually order familiar macroscopic objects correctly, although they overestimate vertical distances relative to horizontal ones (Tretter, Jones, Andre, et al., 2006). They are unfamiliar with the millimeter and smaller units.

From the empirical task analysis (Ch. 6) I found that students understand iteration, even if they have trouble with expressions like "five times longer than" that express the same concept. Students at this level also understand the part-whole and enclosure relationships, though some may neglect to use the information these relationships provide when ordering various objects (e.g., student #1004 - see Ch. 6). This may show that their transitivity is not well established, so they cannot coordinate several comparisons to generate an ordering.

Instruction for students at this level should focus on making the submacroscopic world real. Students can view familiar, small macroscopic objects (e.g., grain of sand, hair) next to submacroscopic objects (cheek cells) using the low power of the optical microscope as we did in camp. This activity built the hair as a landmark object at the boundary of the visible and submacroscopic worlds. Including a ruler in this lesson builds familiarity with the millimeter and its relation to the centimeter and inch.

Students at this level are often intransitive – they rank A < B in ordering but place the objects in groups that imply A > B. In order to build transitivity, the activity viewing the hair and cheek cell under the microscope (see Table 11, Ch. 5) could be done in two

steps; first, viewing the hair and ruler, and second, the hair and cheek cells *without* the ruler. Discussion of what these two relations (hair < millimeter, and cell < hair) imply for millimeter and cell (cell < mm) in qualitative terms should be useful in building understanding of transitivity. This discussion should be followed by actually viewing all three objects simultaneously.

Broader learning about cells in tandem with or prior to these activities is essential. Without understanding what cells are, why they are important, how they function, and possibly their components, viewing cheek cells under the microscope and learning about their size is not likely to be very useful for science learning. Since school explorations of cells often involve microscope activities, the learning activities for size and scale can be incorporated. This will help interweave the relatively content-independent learning progression for size and scale into more content-rich learning progressions.

Level 1: Order-Group Consistency

Based on the teaching experiment, we can expect sixth grade students to reach Level 1 with adequate instructional experiences. The findings from the task analyses and the Study Two teaching experiment now allow additional characterization of students at this level beyond that presented in Chapter 4 (summarized below for convenience).

> Students at this level connect ordering and grouping, as shown by their consistency across tasks. This may open the *possibility* of creating "worlds" of different size ranges, since objects whose place in ordering is known cannot go into just any group. Their only size landmark is the human, in English units and possibly also metric. They may know of an object too small to see, often "germs" or cells, but may require prompting to recall the object. They do not know of units of measurement below the millimeter and have little idea of the relative scale or absolute size of submacroscopic objects. Their ordering and grouping reflects their knowledge that a cell is smaller than macroscopic objects, but they tend to place less-familiar submacroscopic objects as larger than small macroscopic objects like the head of a pin or ant.

The teaching experiment does not provide much information to modify this description, as only one student ended below Level 2 (at Level 0). However, the task analyses show that the connection between ordering and grouping opens up a powerful strategy:

ordering two objects by tool when only one of the objects can be visualized (which involves grouping by tool). This helps the student realize that an unseen world exists, by allowing her to categorize objects by whether they can be seen with the unaided eye. Comparisons of objects to their landmarks (currently, only the human) can also establish group boundaries. Thus, the construction of new landmarks will now be productive. The differentiation of size regions into visible and not visible is an important conceptual advance. Students' increased transitivity helps students to order, by allowing the pairwise comparison of objects to be assembled into ordering (A < B and B < C means A < C). Transitivity also helps students to group, by indicating that an object intermediate in size relative to two other objects that are in a group also belongs in that group (A < B and C > B means A < B < C). These abilities can be built through activities involving size and scale, but would be useful in comprehending or building a wide variety of organizational schemes that conceptually organize ranges or continua, for example, classifying organisms or types of interaction between atoms and molecules. More research is required to see whether classification abilities do in fact transfer broadly.

Instruction for students at this level should focus on building a more quantitative view of transitivity: if A is X times smaller than B, and B is smaller than C, then A is *more than X* times smaller than C. This will allow students to become consistent across relative scale and ordering, that is, to reach level 2. Instructional activities involve measuring a reference object with two non-standard units, and exploring the inverse relationship between unit size and unit number (the proportionality idea of measurement: Lehrer, 2003). For example, the camp activity in which students iterated Staph A bacteria and skin cells across the hair, all at the same magnification, demonstrates the inverse relationship. This activity also builds understanding of the nature of relative scale and expressions like "X times bigger than", which students may not understand (Vergnaud, 1988). This activity can take place in the context of activities designed to help students build their understanding about (eukaryotic) cells and bacteria, so that increased knowledge about the size of the objects is accompanied by increased knowledge of their functions and characteristics. Introduction of units smaller than a millimeter can accompany these activities.

Another focus of instructional activities concerns the nature of grouping by a single criterion. A suggested activity is to engage students in ordering by size an assortment of objects that would result in intransitive ordering (by size) if ordered by another characteristic, for instance, whether they are living or non-living. This activity could be incorporated into any sixth grade topic that involves grouping, permitting additional interweaving of the learning progression for size and scale with other learning progressions. As students begin to realize just how small cells and atoms are, they can also begin to appreciate the complexity of matter and living systems, allowing additional connections to content-area learning progressions.

Level 2: Order-Relative Scale Consistency

Based on the teaching experiment, we can expect most sixth grade students to reach Level 2 with adequate instructional experiences, as all but one camper reached this level. The findings from the teaching experiment show some changes in factual knowledge due to the effect of the curricular intervention, compared to LP1. In comparison to LP1, students at Level 2 who participated in the teaching experiment also knew of micrometers, nanometers, or both, after camp; they also in many cases had an accurate idea of the relative scale of the red blood cell (within a factor of ten). Furthermore, they now could characterize the size of the human in metric units, and knew that atoms are smaller than cells. Other parts of the description of Level 2 remain the same, as summarized below:

> In addition to being consistent across ordering and grouping, students at this level are also consistent across ordering and relative scale. This means that their estimates are larger for objects more distant in size from the reference object. This may open up the *possibility* of establishing coordinated landmarks along a mental measurement line. They still order or group submacroscopic objects other than cell and atom inaccurately.

In addition, from the task analysis I found that Level 2 students can now establish the order of two objects if the relative scale to a third object is known (e.g., a cell is made up of several organelles but of a very large number of atoms; therefore, an organelle is larger

than an atom). Students can now use unitizing as a strategy, characterizing the size of two objects in terms of a reference object. Students at this level can now coordinate the local and limited ordering of objects they might have learned in relatively disconnected physical science and life science courses (e.g., atom < macroscopic, and organelles < cell < macroscopic), thus establishing a broader, interdisciplinary ordering (e.g., atom < organelle < cell < macroscopic) that helps build a mental measurement line. This understanding in turn can be leveraged to help students understand the enormous degree of complexity of even the tiniest cell.

Students at this level can benefit from activities that encourage them to think about how to express sizes of objects smaller than 1 mm, and that show that objects that are smaller have smaller absolute sizes. Camp activities aimed at this goal included the heavily scaffolded calculation of the absolute size of the hair by comparison to the millimeter markings on a ruler, and the calculation of the absolute size of the cheek cells by comparison to the hair. This allowed them the opportunity to compare the absolute sizes for an ordered set of objects. Additional meta-level discussion of the connection between absolute size and ordering would be desirable for this activity. These instructional activities provide opportunities to introduce units smaller than the millimeter and to talk about fractions and decimals; units and/or fractions and decimals are essential to get to Level 4. Calculating the relative scale of very large and very small objects affords opportunities to reflect upon the difference between large but finite numbers, and infinity – an opportunity to interweave with math learning progressions. Learning about cells and atoms as fundamental units of biology and chemistry is essential in order to build the learners' conceptions of these objects beyond their size.

Level 3: Order-Absolute Size Consistency

Level 3 might be a reasonable goal for the end of middle school, since Level 2 is achievable for sixth grade students and students who began camp at Level 2 all reached Level 3. In addition to being consistent across ordering and grouping, and ordering and relative scale, students at this level are also consistent across order and absolute size. Thus, students at this level are capable of generating ordered estimates for absolute size

for two objects smaller than 1 mm. This requires the use of fractions, decimals, or scientific notation in tandem with a millimeter; or units smaller than a millimeter. The findings from the teaching experiment show large changes in factual knowledge due to the effect of the curricular intervention, compared to LP1. Students at this level know of micrometers and nanometers, and rank all the submacroscopic objects smaller than all the macroscopic objects. Thus, they have a clear separation of the macroscopic and submacroscopic worlds, each of which is characterized by some representative objects and units. Their grouping is mainly correct, and their groups do not intersperse submacroscopic and macroscopic objects. Their size landmarks now include human (in both units), atom, and cell, and students have a sense of the relative scale of the cell and the human compared to small macroscopic objects (e.g., pinhead).

The task analysis shows that students at this level can order two objects if their absolute sizes are known. This requires the ability to convert units and/or formats of numbers (decimals, fractions, scientific notation or powers of ten). Students can also estimate the size of an object that is intermediate between two objects of known size, what Joram and colleagues call the "squeezing" strategy for estimation. The larger importance of this new ability is that students can now estimate the absolute size of objects in between two, opening the *possibility* for the construction of new landmarks and a finer and finer mental measurement line *if* knowledge of how to construct a scale is present (Joram, 2003). Scale-building knowledge includes the connection between relative scale and absolute size, which defines Levels 4 and 5. Thus, instruction at this level should focus on building an understanding of the connection between relative scale and absolute size, a form of proportional reasoning. Activities used in the teaching experiment to address this learning included calculating the absolute size of one object from relative scale and the known absolute size of another object (e.g., of the skin cell given that it is around 7 times smaller than the thickness of a 0.1 mm hair). These activities also can address unit and number conversions – in the camp curriculum we did not address either in depth, relying instead on units to avoid number conversions, and scaffolding students in unit conversions.

Given that students are constructing the atom as landmark at this level, simultaneous or prior instruction about the properties and characteristics of atoms is important. This will allow the interweaving of the size and scale learning progression with the learning progression(s) for matter (e.g., Stevens, Delgado, & Krajcik, in press). The students' ability to represent the size of submacroscopic objects using units can provide opportunities to talk about measurement, precision, accuracy, and error, which could be its own learning progression or another dimension of this one.

Level 4: Partial Relative Scale-Absolute Size Consistency

Level 4 might be a reasonable goal for the end of 10th grade, since Level 3 is feasible by the end of middle school and Level 5 by the end of high school. Around 1/5 of the students in Study Two's teaching experiment ended camp at Levels 4 or 5. Students at this level begin to connect relative scale and absolute size, generating consistent estimates for some objects, in addition to the types of consistency evinced by Level 3 students. At this level, students understand that relative scale and absolute size are related, but do not always see the need to be consistent across them. They see the need more often with objects with a well-known absolute size, that is, with landmarks. The findings from the teaching experiment show large changes in factual knowledge due to the effect of the curricular intervention, compared to LP1, though only improved ordering beyond Level 3 on LP2 (described immediately above). Ordering now includes being able to locate molecules, viruses, and organelles intermediate between the landmarks of atom and cell, though not necessarily the order of the intermediate objects.

The task analysis shows that students at this level can now (if they choose to do so) calculate an absolute size from relative scale and the known absolute size of a reference object. In principle, this means that students have "knowledge of how a scale is constructed" by iterating a landmark to measure an object that is not a landmark itself (Joram et al., 1998, p. 427). This completes the construction of a mental measurement line. In practice, the landmarks would have to be close enough that one is not iterating a landmark 100 times to measure an object. DNA, small and large viruses, organelles, and bacteria are potential landmark objects that students can build in a more sophisticated

mental measurement line. The task analysis also shows that students at this level *could* estimate an absolute size by estimating apparent size and dividing by the magnification of the tool. This is likely a very useful technique for experts who routinely use a tool such as the optical microscope, although the known absolute size of pointers could also be used. However, the ability to estimate absolute size would have to be built up before students could use this strategy. Activities aimed at this goal could become an additional strand in this learning progression, or be included in one on tools.

Calculating an absolute size of an object smaller than 1 mm given the relative scale requires calculating the inverse of a number (e.g., the size of a cell 200 times smaller than a 1-mm pinhead is 1/200 mm), so students at Levels 3 and 4 benefit from the instruction of fractions, decimals, and powers of ten or scientific notation they are likely to receive in upper elementary and middle school. This instruction can be in the context of size and scale; Resnick's (1992) hierarchy of four types of mathematical thinking suggests that measuring physical substance to build knowledge of fractions, decimals, and powers of ten may be a more efficacious and natural strategy than teaching these as pure numbers. Learning activities for these skills can be included in activities in the teaching experiment where the size of successively smaller objects were calculated from relative scale and the known absolute size of other objects (e.g., skin cell from hair, Staph A from skin cell, etc.). In the camp we bypassed fractions, decimals, and powers of ten through the use of units, but these could be addressed, given more time. These activities allow for opportunities to interweave math and size and scale, and to teach these mathematical skills in the context of science and measurement.

Given that students often take Chemistry in the 10th grade, the smallest object they are familiar with can now be sub-atomic particles like the electron.

Level 5: Full Relative Scale-Absolute Size Consistency

My findings and those in the literature show that Level 5 is a feasible goal for 12th grade students. Based on the middle school students' ability to construct the atom and cell into landmarks, I extrapolate that high school students might additionally have other

landmarks including some of the important objects intermediate in size: DNA, viruses, bacteria, and thickness of a hair.

Students at this level connect relative scale and absolute size in all cases. This means that they use proportional reasoning cases to calculate relative scale from absolute size, or to generate a consistent absolute size from an estimated or known relative scale. Their consistency affords accuracy in relative scale estimation equal to absolute size estimation. They can calculate the size of objects that are not landmarks, if they know or can estimate how many times larger the objects are than the landmark. They may still have trouble accurately ordering and grouping due to uncertainty about the size of viruses compared to the cell, parts of cells, or both. Students at level 5 know of subatomic particles (or could, with improved curricula).

At this level, students can use the ratios of the size of two objects compared to a third (or unitizing, a Level 3 skill) to calculate the relative scale of the two objects. For instance, knowing that a Staph A bacterium is around 7 times smaller in diameter than a red blood cell, and that a white blood cell is around twice as large as a red blood cell, will allow students to calculate that the bacterium is around 14 times smaller than the white blood cell. This allows students to establish a multiplicity of relative scale relations between objects of interest, from other known relative scales or the absolute sizes of their various landmarks, creating a well-developed mental measurement line. While this study did not examine students' capacity to learn about the size of the Earth, I surmise that they could establish the size of the Earth as a landmark from their ability to learn the absolute size of the atom as well as prior research showing that students' accuracy of estimates of very large objects is greater than for very small objects (Tretter, Jones, & Minogue, 2006).

The task analysis shows that students at Level 5 are proficient with unit conversions and translating numbers from one format to another including decimals, fractions, and scientific notation. The students can group objects and justify their criteria for group boundaries and/or ranges. Students at this level are fully consistent across all aspects of size and scale, and can use a variety of strategies to solve related tasks. I surmise that

their proportional reasoning skills developed in the context of size and scale would transfer to other situations as well.

Discussion

In this chapter I presented LP2 for middle and high school one-dimensional size and scale, informed by cross-sectional data, a theoretical task analysis, and a teaching experiment. LP2 traces the gradual growth of connections across aspects of size and scale that students may experience over the middle and high school years, along with the mathematical and logical knowledge that is required. It also traces the growth of factual knowledge about the size and scale of important objects in science, including the atom and cell that is likely to occur if a student experiences suggested instructional activities. LP2 uses the same levels as LP1, but adds several dimensions. First, the typical content knowledge associated with each level is revised (upward) based on Study Two. Second, the logical and mathematical knowledge required for each level is described (based on the theoretical task analysis). Third, specific learning performances that students can accomplish at each level are described. Fourth, tested or suggested instructional activities targeted at students at each level are included. Fifth, possible connections to other content areas of science learning are suggested

Even though this learning progression covers six years, the actual amount of content considered is narrow. Thus, I consider how this learning progression might be expanded in future work.

Before Level 0

Before reaching even the very modest level of knowledge contemplated in the lower anchor, students have learned an enormous amount about the size of objects. For instance, young learners need to achieve conservation of length (Piaget, et al., 1960). Learners need to learn to differentiate size (three-dimensional) and weight (Piaget, 1974; Smith, Carey, & Wiser, 1985). They have to learn to understand the diverse meanings of words like "big" or "little" (Gelman & Ebeling, 1989). Other studies have documented young children's difficulty in seeing a scaled-down object as a symbol for the full-size object (DeLoache et al., 1999). Thus, the learner needs to begin by building the concept

213

of size itself. LP2 can thus be extended downwards into early childhood based on the literature and perhaps some new empirical studies. This is outside of the scope of this dissertation, however,

Beyond Level 5

Although this study did not examine student knowledge beyond Level 5, I can describe some potential learning that builds directly on this level. Once students have constructed a mental measurement line (Joram et al., 1998), they may also represent it externally. While linear scales and graphs are most familiar to students, these do not permit the representation of values that differ by more than 3 or 4 orders of magnitude on a computer screen or notebook-size sheet of paper (e.g., a single pixel is the lower limit of resolution on a screen, and the screen on a current Mac laptop is under 2000 pixels across). Logarithmic scales permit the representation of many more orders of magnitude. Recent studies have shown that undergraduates at a selective university may not choose or know how to use a logarithmic scale to represent widely varying sizes (atom to football field, 12 orders of magnitude). Those who do may create hybrid log-linear combinations (e.g., using linear spacing between powers of ten) or include other non-normative features (Light, Swarat, Park, Drane, Tevaarwerk, & Mason, 2007; Confrey, 1991). Thus, students' learning about linear and logarithmic scales is one direction in which this learning progression can continue.

Parallel Directions and Developments

Two- and three-dimensional size (area, volume) are intrinsically more complicated than one-dimensional size studied here. The scaling of these quantities is not linear, as the AAAS *Benchmarks* (1993) point out. As size drops towards the nanoscale, the surface area to volume ratio increases enormously, creating changes in properties responsible for unique phenomena such as the catalytic properties of gold nanoparticles (while larger chunks of gold are chemically unreactive and non-catalytic). The understanding of these phenomena will build upon a Level 5 understanding of 1-D size and scale.

Another manner in which the understanding of size and scale can advance, at all levels, is through the interweaving with learning progressions for other science content areas, such

as the nature of matter or forces and interactions. In all big ideas, size and scale is likely to be a factor – thus its status as a crosscutting common theme (AAAS, 1993). This affords the ability to learn size and scale in the context of important science areas, and also for size and scale to support and complement the learning of those areas. Furthermore, size and scale's broad applicability and developmental nature, depending as it does on broad logical and mathematical skills, may provide a framework with which many other learning progressions can be synchronized and interwoven.

LP2 is one of the first examples of the iterative construction, testing, and revision of a learning progression. The positive results observed provide support to the usefulness of the learning progression approach. By being a relatively content-independent learning progression and common theme, it could provide an organizational framework or scaffold for deeper, more integrated, interdisciplinary science learning.

CHAPTER 8: CONCLUSION AND IMPLICATIONS

Conclusion

The purpose of this dissertation was to develop a learning progression for size and scale, to be used to guide the development of more effective curriculum, instruction, and assessment for this important "common theme" (AAAS, 1993). Due to the limited research base on size and scale, however, I carried out research to fill significant gaps in the literature. Specifically, I addressed the following research questions:

1) What do middle school through undergraduate students know about one-dimensional size and scale, given their current curricular experiences?

2) What do middle school students know about one-dimensional size and scale, after a focused curricular experience?

3) What is the relationship between factual knowledge and consistency of knowledge of size and scale, if any?

The research I conducted guided by these research questions allowed me to iteratively build a learning progression for one-dimensional size and scale, for middle and high school. Several findings emerged from this work. Specifically, from my research guided by Research Question 1, I defined and empirically investigated a dimension of student conceptual understanding of size and scale - consistency of knowledge – and detected a progression in which students appear to build this understanding. This is important because constructivist theory points to the importance of connectedness of knowledge (e.g., Piaget, 1983; Linn, Davis, & Eylon, 2004; diSessa 1988), yet prior research had not fully examined this dimension in the case of size and scale (e.g., Batt et al., 2008). The findings of this dissertation produce a more complete picture of conceptual understanding

216

of size and scale.

For the teaching experiment conducted for Research Question 2, I developed an instructional unit for middle school students that was effective in moving students along the progression for size and scale, resulting in statistically and educationally significant learning gains. This is an important development in and of itself; however, the teaching experiment also provided evidence to validate and refine the learning progression generated by the cross-sectional study used to investigate the first research question. Importantly, the teaching experiment lends support to the learning progression hypothesis (Smith et al., 2006; Duschl, 2007; Wilson & Bertenthal, 2005): the idea that learning progressions can lead to the development of effective science curriculum, instruction, and assessment resulting in student learning. This finding is important because it lends support to a new approach being developed by the science education community that could be instrumental in building more coherent curricula that might result in less fragmented, more robust student knowledge. Since US standards documents in science education identify scale as a concept that pervades science, and that can be used to unify student learning across disciplines, topics, and grades (AAAS, 1993; NRC, 1996), a learning progression for size and scale may be a particularly powerful tool in fostering more connected science learning.

From Research Question 3, analyzing the relationship between factual knowledge and consistency of knowledge, I identified key logical and mathematical skills required for size and scale, building upon and extending prior researchers' observations that proportional reasoning is essential (e.g., Tretter, Jones, Andre, et al., 2006). This is important because it provides a road map for the development of instructional materials both for size and scale in science education, and for proportional reasoning in mathematics education. I identified the strategies learners can use in approaching size and scale, extending prior research into the estimation of absolute size synthesized by Joram and colleagues (1998) and allowing for the identification of common elements for strategies across aspects that enable a meta-level look at size and scale tasks.

217

From the learning progression itself, an account emerges of how students discover the unseen world, develop reference points to anchor and organize this world, and learn to characterize the size of objects in between the anchors. This work is to my knowledge the first to provide actual empirical evidence for the viability and effectiveness of the learning progression approach to the design of curricula, instruction, and assessment (proposed by Smith et al., 2006; Duschl et al., 2007; and Wilson & Bertenthal, 2005), which could be transformative in creating more coherent and cohesive science education instruction. This dissertation also provides one possible model of how to develop a learning progression for a core idea, using an iterative, design research approach, and thus helps define and operationalize a promising educational reform movement in science education.

Next, I expand upon my conclusions outlined above, and describe implications for each research question and for the learning progression itself, in greater detail.

Knowledge of Students Given Their Current Curricular Experiences

My dissertation further characterized the factual knowledge of students in the absence of a focused curricular intervention for size and scale, beyond the current literature base. For instance, the smallest object students recall were found to include "microscopic objects" by Tretter, Jones, Andre, and colleagues (2006); my study additionally showed that students tend to recall atoms and cells or microorganisms, showing that these are potential landmark objects. Previous studies had shown that middle school students tend not to know of units of measurement of length smaller than a millimeter, and many do not know how to define a millimeter (Jones et al., 2007). My study additionally showed that some students do not initially think of millimeters but seem to recognize them once these are mentioned to them; other students do not use the millimeter even after this measure is provided to them, and these students tend to perform lower on size and scale tasks. My dissertation research thus further characterizes the knowledge of lower-performing students beyond a simple lack of knowledge of millimeters or smaller units. My study showed that younger students perform much better at estimating the height of a human using English units compared to metric, but gradually increase their mean

accuracy with metric units as they progress through school; the undergraduates at a selective research university displayed no difference in performance at all. Going beyond previous research showing that students tend to overestimate the sizes of small objects and underestimate the sizes of large objects (e.g., Tretter, Jones, Andre, et al., 2006), my work showed that a group of students will tend to have a wide variety of estimates for a given object, with smaller ranges for objects nearer human size.

In addition to contributing to the sparse literature reporting on the knowledge of students about the size of objects, this research opened up a dimension of knowledge of size and scale that had barely been explored (e.g., Tretter, Jones, & Minogue, 2006; Batt et al., 2008). By studying whether and how students connect the different aspects of size and scale (ordering, grouping, relative scale, and absolute size), my work built upon previous research examining each of these in isolation and contributed to the characterization of conceptual understanding of size and scale. By exploring the growing connectedness of the aspects of size and scale, my work supported and added detail to Resnick and colleagues' (Resnick, 1992; Resnick & Singer, 1993) hypothesis that mathematical knowledge begins with protoquantitative and quantitative mathematical knowledge depending on interactions with amounts of physical substance.

This examination of students' factual knowledge and consistency of knowledge led to the development of the first iteration of a learning progression for size and scale. Jones and Taylor (in press) recently presented a trajectory for student learning of size and scale; my work contributes to the effort to characterize a progression for size and scale by presenting additional details for a research-based pathway in which students develop their factual and conceptual understanding of size and scale, including the dimension of consistency of knowledge.

Knowledge of Students Given a Focused Curricular Experience

Using the first iteration of the learning progression for size and scale as a guide, I developed a curriculum for size and scale aimed at middle school students. I built upon

existing but isolated instructional activities and simulations (e.g., Tretter, 2006; Tretter, 2005; Sullivan, n.d.) in order to create this 12-hour instructional unit. This instructional unit contributes to the literature base on several levels. First, it provides a coordinated set of activities proven to help students construct their knowledge of size and scale. Second, it provides a model for the development of such a unit, following a learning progressions approach. Third, student learning gains lend weight to the learning progression hypothesis (Smith et al., 2006; Duschl et al., 2007; Wilson & Bertenthal, 2005): the idea that learning progressions can lead to the development of effective science curriculum, instruction, and assessment.

As of this writing, there had been no empirical tests of the hypothesis that using learning progressions to design curriculum and instruction will lead to improved student learning. Plummer (2006) used a cross-sectional study coupled with a teaching experiment to develop a learning progression for elementary students' development of astronomy concepts related to the apparent motion of celestial objects. However, while the curriculum was effective in promoting learning gains, it was not explicitly developed based on a learning progression, as was the case in my study. Similarly, Onyancha and colleagues found higher learning gains in student learning related to the carbon cycle with designed instructional materials developed in the context of a learning progressions research project; however, the role of the learning progression in designing these materials is not described (Onyancha, Lee, Choi, Draney, & Anderson, 2009). Other research groups are developing and implementing science curricula based on learning progressions, but have not yet reported their results concerning student learning. In this dissertation, I followed a learning progression approach, using the learning progression explicitly to guide the development of effective curriculum and instruction, constituting an existence proof for the effectiveness of this approach.

The teaching experiment embodied and exemplified Vygotsky's ideas (1978, 1985) that students most readily learn concepts that are within a certain distance of their current knowledge: the Zone of Proximal Development. Students tended to advance around one to two levels on the six-level progression; no student advanced more than three levels.

This finding explicates the value of the learning progression approach, which carefully traces possible ways students construct their knowledge. This approach to the design of curriculum and instruction can thus ensure that students have activities that are within their ZPD. This work also extends the work of Pea (1988) by showing that measurement units such as micrometers and nanometers functioned as an inscriptional system that allowed students to off-load the need for fractions or decimals onto the environment. Units are powerful tools for student learning about the unseen world.

Relationship Between Factual Knowledge and Consistency of Knowledge

The theoretical and empirical task analyses I conducted resulted in the identification of multiple strategies students can use in approaching size and scale tasks. This extends the work of Joram and colleagues (1998) synthesizing strategies for absolute size estimation. I found that for each aspect, learners can use strategies involving recall, using that single aspect, or employing more than one aspect. The strategies that involve multiple aspects rely on the connections across aspects defined by this dissertation and that were instrumental in defining levels of the learning progression.

Previous research identified proportional reasoning as a key skill for size and scale (e.g., Tretter, Jones, Andre, et al., 2006). However, the mathematics education community has amply established that proportional reasoning is a difficult milestone for middle school students. Building upon Hiebert's (1981) findings that conservation and transitivity are required to understand the inverse relationship between unit size and unit number (which is a fundamental principle of measurement – Lehrer, 2003), I determined through theoretical and empirical task analyses what specific strategies for tasks related to the different aspects of size and scale require these logical and mathematical skills. The task analyses highlighted the importance of part-whole and enclosure relationships, and mathematical content knowledge such as the ability to calculate the inverse of a number.

By identifying precursor skills for relative scale-absolute size consistency a type of consistency that constitutes a form of proportional reasoning), my work suggests a

possible sequence or hierarchy for the development of mathematics curriculum and instruction for proportional reasoning that builds upon Resnick's work (1992).

Learning Progression for Size and Scale

This dissertation also provides one possible model of how to develop a learning progression for a core idea, using an iterative, design research approach. The methodology I used, involving cross-sectional data guiding curriculum development, and a teaching experiment employing this curriculum to generate a second iteration of a learning progression, provides a model that other researchers can adopt or adapt. The field of science education is gradually developing a consensus definition of learning progressions and the methodology for their development, through the *Learning Progressions in Science* conference (June 2009) and upcoming special issue of the *Journal of Research in Science Teaching*. This dissertation provides one possible model of development, an exemplar of a learning progression, and lends support to the learning progression approach. It thus helps define and operationalize a promising educational reform movement in science education.

The learning progression I generated builds upon Tretter and colleagues' concept of landmarks, and upon Joram and colleagues' idea of a mental measurement line. The learning progression describes how students gradually establish landmarks (Tretter, Jones, Andre, et al., 2006) that help them construct a mental measurement line (Joram et al., 1998), allowing them to increase the accuracy of their factual knowledge in tandem with an increased connectedness of knowledge. Students' conceptual understanding grows as they learn to coordinate and connect different aspects of size and scale; the growing connectedness of their knowledge facilitates the growth of factual knowledge as well.

Implications

This dissertation lends support to hypotheses concerning cognitive development – the

interaction of content knowledge and meta-knowledge. Examining how content knowledge interacts with the development of meta-knowledge was identified as a promising strategy in studying cognitive development (Chi & Ceci, 1987). This interaction had not previously been examined in the context of size and scale, but it proved very fruitful. I found from the theoretical task analysis that the (meta-level) connections across aspects enable strategies that can help students improve the accuracy of their factual knowledge. I found from empirical analyses that factual knowledge and consistency of knowledge are strongly correlated, and that consistency of knowledge predicts future learning of factual knowledge. I also found from the empirical task analysis that accurate factual knowledge of the size of an object motivates some students to try to be consistent across tasks. Thus, this dissertation supports Chi and Ceci's call for researchers to examine how factual or content knowledge and conceptual knowledge interact.

This study suggests some possible ways to interweave learning progressions for science and mathematics with the learning progression for size and scale. The learning progression for size and scale is relatively domain-general; this means that it can easily be coordinated with content-heavy learning progressions for areas of traditional science content. Learning progressions for common themes might serve as guideposts or frameworks for the alignment of science content across disciplines and grades – helping students connect their learning as is proposed in the *Benchmarks for Science Literacy* (AAAS, 1993).

Previous research had pointed out that students need proportional reasoning to understand size and scale (Tretter, Jones, Andre, et al., 2006). This dissertation identifies the specific mathematical knowledge and logical skills students require for size and scale, prior to proportional reasoning. The learning progression for size and scale contains an account of development that may be useful in helping students construct their proportional reasoning, by scaffolding their consistency across increasingly more complex and quantitative aspects of size and scale. This hypothesis, that the learning progression for size and scale can lead to improved instruction for proportional reasoning, can be

empirically tested. This study provides guidance on how to implement and test this hypothesis – beginning with activities that require the coordination of qualitative aspects of size and scale (i.e., ordering and grouping), progressing to qualitative/quantitative (e.g., ordering and relative scale), and culminating in the coordination of quantitative aspects (relative scale and absolute size).

Ultimately, this study concerns connected knowledge, at multiple levels. At the individual level, the learning progression describes how individual students tend to gradually construct a more connected and factually more accurate conception of size and scale. The development of a learning progression allows for the development of a more systematic curriculum for size and scale that addresses the zone of proximal development (Vygotsky, 1978, 1985) for each learner, improving the vertical connectedness of instruction. Finally, by suggesting how this learning progression can be interwoven with instruction of existing topics in science and mathematics at various nexuses, this dissertation shows how horizontal connectedness of instruction could be increased via size and scale. Together, these dimensions of connected knowledge could begin to realize the promise of the "common theme" (AAAS, 1993) and of the learning progression hypothesis, to help students connect and structure their knowledge across topics, disciplines, and grades resulting in more robust, flexible, and useful scientific knowledge.

Appendix A: Interview Protocol

SIZE AND SCALE INTERVIEW

-What is the very smallest thing you can think of?

If response macro,

-Can you think of something too small to see with the naked eye?

If ambiguous ("nucleus/particle")

-What do you mean by that? Could you be clearer?

-What else do you know of that is too small to see with naked eye?

-What type of measurement units would you use to express the size of that object?

(If necessary, prompt by saying that the width of the table could be expressed in centimeters or inches)

-What is the smallest unit you know of?

(Lay out cards)

OK. Take a look at these cards. I'd like you to put them in order by the size of the objects, from largest to smallest. Are you familiar with the objects? (If necessary, clarify: atom of carbon, molecule of water, HIV virus, state that mitochondrion is a part of a cell). (Demonstrate the size of the head of a pin at this point).

(Select the micro- and nano- cards in pairs.)

-Could you please tell me why you ordered these cards the way you did?

Could you please place the cards into groups of objects of similar size? Make as many groups as you think makes sense.

-Can you tell me how you decided to group these cards together?

(Repeat for tape recorder how many groups and what cards in each).

-What do they have in common?

-What would you call this group?

(Repeat for each group)

Interviewer selects five cards from task 3. These will be atom, cell, pinhead, human, and Earth if they are ordered correctly. If atom and cell are out of order, select one of those, and choose another card in the correct order.

OK. Here are five of the cards you ordered. I want you to think about the length of these objects. For the pinhead, think how wide it is, that is, the diameter *(Trace width with finger)*. For the person, the height. (*Trace*). For the Earth, atom, and cell, the diameter. (*Trace*). (*Record all answers on worksheet*)

-How many times larger is the human than the pinhead*?*

-How many times larger is Earth than the pinhead?

-How many times smaller is the (red blood cell or substitute item) than the pinhead?

-How many times smaller is the (atom or substitute item) than the pinhead?

OK, a pinhead is about 1 mm wide. That's a little less than 1/16th of an inch

-Would you write down the size of the other objects?

(Pass the student pen and worksheet, and offer scratch paper. Remind student to specify units, if necessary. Ask them to also use metric system, if they use English units.)

(Indicate the relative scales recorded on sheet)

-Did you use these numbers here to think about the sizes of the objects?

If yes,

how?

If no,

How do you think the two sets of numbers are related, if at all?

If you can think of other ways to express the sizes using different units or different ways of writing the numbers would you please write them below too?

Appendix B: Claim, Evidence, and Tasks for Interview

Ordering

A)

Claim	The student is able to order a range of macroscopic and submacroscopic objects by size
Evidence	Given a series of objects, the student's work includes ordering that shows atom < virus < cell < small macroscopic objects < human < large macroscopic objects.
Task	Order the following objects by size: diameter of an atom, length of a small molecule (e.g., water or carbon dioxide), diameter of a rhinovirus, diameter of a mitochondrion, diameter of a red blood cell, pinhead, ant, height of an average adult human, height of a mountain, diameter of the earth

B)

Claim	The student can recall objects at the micro-, nano-, and/or sub-nanoscales.
Evidence	The student mentions a submacroscopic object when asked for his/her smallest known object
Task	What is the smallest object you can think of? Can you think of anything too small to see?

C)

Claim	The student can recall units of measurement for 1-D length that are smaller than the millimeter, including the micrometer, nanometer, picometer, and/or Angstrom
Evidence	The student mentions a submacroscopic unit when asked for his/her smallest known object
Task	What unit of length can you use to conveniently express the size of the smallest object you can think of? What is the smallest unit you can think of?

Grouping

Claim	The student can group by size using reasonable criteria
Evidence	The student makes groups that accurately reflect the size order of the objects, and can explain the criteria for groups or group boundaries.
Task	Group these objects (atom, molecule, virus, mitochondrion, red blood cell, pinhead, ant, human, mountain, Earth) by size. Make as many groups as you think makes sense. Explain what the objects in each group have in common, or what you would call each group is.

Relative Scale

Claim	The student is able to estimate the size of a range of objects in terms of a convenient and familiar reference object.
Evidence	Given a series of objects and a reference of known size, the student's work includes estimates of sizes that are accurate to within one order of magnitude.
Task	How many times bigger or smaller than the head of a pin (1 mm in diameter) do you think the following objects are: diameter of an atom, diameter of a red blood cell, height of an average adult human, diameter of the earth?

Absolute Size

Claim	The student is able to estimate the absolute size of a range of objects in terms of standard units.
Evidence	Given a series of objects and a reference of known size, the student's work includes estimates of absolute sizes that are accurate to within one order of magnitude, expressed in standard units.
Task	Estimate the absolute size of the following objects: diameter of an atom, diameter of a red blood cell, height of an average adult human, diameter of the earth?

Appendix C: Answer Recording Sheet

18

ATOM   CELL   PIN HEAD   HUMAN   EARTH

3000

100

1000

5 000 000

ATOM   CELL   PIN HEAD   HUMAN   EARTH

1 mm   1.2 mm   1 mm   5.8   700 cm

ATOM/V/CELL/MOL/m.T/PIN/ANT/HUMAN/ANTN/EARTH

GROUPS: AT/V PIN/CELL MITO/ANT HUMAN/MOL EARTH/MITUS.

Appendix D: Coding Rubric for Factual Knowledge About the Size of Objects

I. Smallest object respondent knows of

Object	Size	Code
non-matter/nonsense/do not know		0
macroscopic (e.g., grain of salt)	> 100 μm	0
cell/microorganism/germ	1-100 μm	1
part of cell/large protein	10-1000 nm	1
atom/small molecule (in 1+ D, e.g., DNA)	0.1-10 nm	2
sub-atomic (e.g., electron)	< 0.1 nm	3

IIA. Unit to express the size of that object

Unit	Code
do not know/do not exist/non-length	0
macroscopic-sized, English (e.g., inch)	1
macroscopic-sized, metric (e.g., mm)	1
submacroscopic fraction of unit, English or metric (e.g., 1/1000 of inch)**	2
submacroscopic unit (e.g., nm)	2

Note: students who answer I above with a macroscopic object are reported separately from this coding scheme.
**Accept descriptions of micrometer symbol

IIB. Smallest unit known: code as above. Code 9 if not asked.

III. Ordering 10 cards by size of the object depicted

Ordering	Code
Macroscopic objects ordered incorrectly or interspersed with submacroscopic objects (i.e., NOT pin/ant/human/mountain/earth as largest)	0
macro OK; atom ordered as larger than cell	1
macro OK; atom < cell, but atom not smallest of all	2
macro OK; atom smallest, but cell not ranked larger than molecule, virus, mitoch	3
macro OK; atom smallest, cell ranked larger than molecule, virus, mitochondrion; but molecule/virus/mitochondrion out of order; or correct order without rationale	4
all objects in the correct order AND student gives rationale for ranking.	5

IV. Grouping

Grouping	Code
groups are incorrect (e.g., {pin, human} excludes ant), or 10 groups, or 1 group	0
groups are correct but mix macroscopic and submacroscopic objects	1
2 correct groups	2
3 correct groups	3
4 correct groups	4
5-9 correct groups	5

V, VI. Size relative to pinhead and absolute size for atom, cell, human, earth

Object	Relative: Range for code 1 (else 0)	Absolute: Range for code 1 (else 0)
atom	500,000-100 million	0.01 nm-2 nm
red blood cell	10-2000	0.5-100 μm
Human	150-20,000	0.15 m to 20 m*
Earth	1 billion-150 billion	1,000-150,000 km

*Code for any units (metric or English) and for *metric only*.
Note: For students who rank cell < atom < pinhead, substitute cell or atom to get another series of two submacroscopic objects ranked smaller than pinhead. Acceptable ranges: virus 1000-200,000; mitochondrion 50-10,000; molecule: cannot determine due to range of sizes of molecules. If student did not rank two objects < pinhead, code as 0 for atom and cell if cannot code following this scheme.

Appendix E: Coding Rubric for Consistency Across Aspects of Size and Scale

I) Ordering-Grouping Consistency

Description	Code
cannot group, or makes 10 groups of 1, or 1 group of 10	0
groups inconsistently with order (e.g., A < B < C < D < E; {A, C}{B, D, E})	0
groups consistently with order, independently of accuracy of order (e.g., A < D < C < B < E; {A, D}{C, B, E})	1

II) Order-Relative scale Consistency

Description	Code
did not estimate relative scales	0
relative scale of any pair of objects inconsistent with order (larger or equal factor assigned to intermediate object than largest or smallest object)	0
relative scales are consistent with order	1

III) Absolute Size-Order Consistency

Description	Code
did not estimate absolute sizes	0
absolute size of any pair of objects inconsistent with order (a size $\geq$ assigned to an object ranked smaller in the ordering task)	0
relative scales are consistent with order	1

IV) Relative scale-absolute size consistency

Description	Code
did not estimate relative and/or absolute size	0
relative scale and ratio of absolute sizes consistent for 0 or 1 object	0
relative scale and ratio of absolute sizes consistent for 2 or 3 objects	1
relative scale and ratio of absolute sizes consistent for all 4 objects	2

Appendix F: Sample Interview Summary

0099 S&S Summary

6/2/06 - CD

S&S starts at 18:00

Smallest: electron.

Unit: nm. "Nanomicrometer"

Atom/mitochondria/virus/molecule/cell/pinhead/ant/human/mountain/earth

(24:00) Groups: {atom/mitochondrion}, {virus, molecule, cell}, {pinhead, ant}, {human, mountain, Earth}. The two largest groups: visible to naked eye. Thought about placing them all together, but pin, ant, small so in their own group. Smallest group: building blocks. Second smallest group: made up of red group, but not visible to naked eye. Red blood cells are made of molecules, viruses are made of atoms, mitochondria are in the nucleus of a cell (or atom?). Viruses have mitochondria. Cells are made up of atoms.

(25:00) I: How many times bigger is human than head of a pin?

R: Oh, seeing as that's like a millimeter, and so, like 6 feet tall, is how many meters? Is two meters, so a millimeter is a thousandth - is it OK if I write something?

I: Sure - could you do it on the, here [back of the paper].

R: Hopefully a good estimate. So, 6 feet equals two meters, so if that's one millimeter (asked to speak into recorder). (26:00) So I'm saying 6 feet equals about two meters, and so then, how many millimeters are in a meter. There's a hundred centimeters, er, thirty centimeters in a foot, so thirty times ten, so there's 300 mm in a foot, I think. So then 300 times 6. So then 1800 millimeters in six feet...OK, so assuming all this math is right at [early] in the morning, I say it's like 1800 times. [Asked about Earth to pin] Oh my God (laughs) How many miles is the diameter of the Earth? Are you allowed to tell me how many miles the diameter of the Earth is?

I: No.

R: OK. Well, what I would do was, I'd think of how many mm in a foot, then I'd just find out how many miles in the diameter, find it out. I'm a math person. You just want an estimate...

I: Do you have any idea what the diameter of the earth might be?

R: We always have the conversion factors in our textbook and we just open it to do all our math problems, so I kind of mindless calculating. (28:00) So let's see, from here to Boston is around 3000 miles [NOTE: actual distance is about 1/4 that]. Let's say it's like 10,000 miles, but I have no idea...so it's got to be like millions of times...

I: If you had a calculator, could you do it?

R: The diameter of the earth? Oh yeah, like the diameter of the Earth? Yeah.

I: Well, I can be your calculator if you need a little bit of help! You've got 10,000 miles. What do you know about miles?

R: There's 5280, or 60? 5260, 80...(NOTE: correct number is 5280)

I: So how many feet is 10,000 miles then? What operation would you do?

R: OK, so this is how many feet per mile, so then times 10,000.

I: OK, and then how would you get to, you've got feet now.

R: OK, feet, and you want it in mm...you multiply by 300. Assuming that's right, but I don't know about that.

I: (mumbles numbers) About 15 billion then.

(Meanwhile, R is calculating with pencil and paper. She uses units at several steps, including mm/ft in a conversion factor)

Interviewer helps her notice that there is one zero missing. (31:10)

I: Did you use the numbers here to think of numbers here?

R: Yes. Because it's the conversion factor. I mean, it's the same thing...

I: So this number is dictated by the number here?

R: Yes. Correspondingly. Like compared to this, it's basically writing the same thing.

I: If a person (who did it wrong said they're different?)

R: Once you give it a number, it corresponds, it comes out to be the same thing...

I: Do you think there was a time, when you were smaller, in middle school or elementary, when you didn't know numbers like this had to be connected to numbers like that, or do you think you ALWAYS knew how to do this?

R: Well, probably, every skill is leaned, but I can't remember a specific time learning it.

I: It seems obvious now.

R: Yeah.

Appendix G: The 11th-Grade Low Performing Students

As noted in Chapter 4, the 11th-graders performed at a lower mean level than 10th-graders on most of the factual knowledge task. This is due to very low performance by a few 11th grade students rather than generally lower performance by all students in the 11th grade. Graphs of overall factual knowledge score and consistency score showing these low-performing students are presented below. All of these students were at the public school, and were interviewed by three different interviewers, indicating that the very low results are not due to a particular interviewer's technique. One of these students (#0073, marked case 13 in the graphs) ranked the ant as the smallest of the 10 objects, grouped Earth and atom together but could not explain why, and could not estimate absolute sizes, saying she was "not good at that". This student was in the non-college preparatory track for both science and math, and in fact failed her science course (Chemistry) the year she was interviewed. Thus, she was a particularly weak student in both math and science. Another low-performing 11th grade student (#0089) was a student recently arrived from a different country who grouped the cards for "balance", pairing the largest with the smallest, second largest with second smallest, and so on. The student ranked atom as the smallest object, but pin as the second smallest. His status as recent immigrant makes him atypical. The remaining low-performing 11th grade students were weak in science and math, as they were taking remedial courses, were 2-3 years behind in the course sequences, or both.

Another reason for the dip in 11th grade performance is that only one out of 17 11th-graders was from private school, whereas 11 out of 17 10th-graders were from the private school. Private school students on average performed a higher level than students in the public school. Thus, the dip in performance at the 11th grade should be seen as an artifact of sampling rather than a characteristic one would expect to see in the general population.
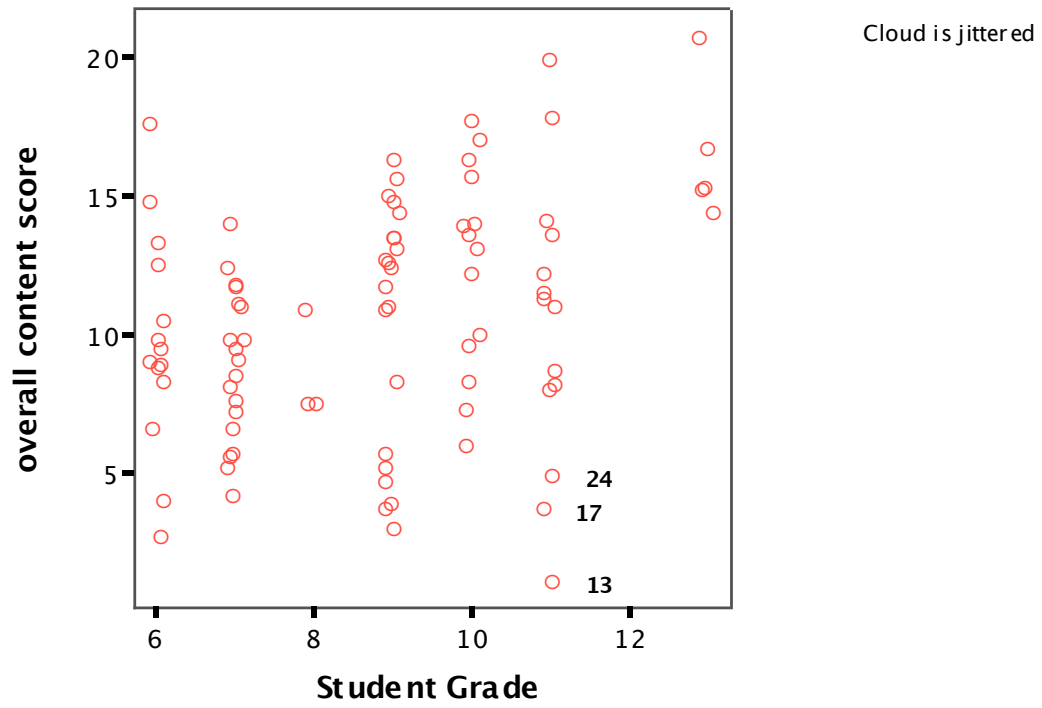
Figure 35: Scatter plot of overall factual knowledge score vs. student grade showing the 11th grade low-performing students
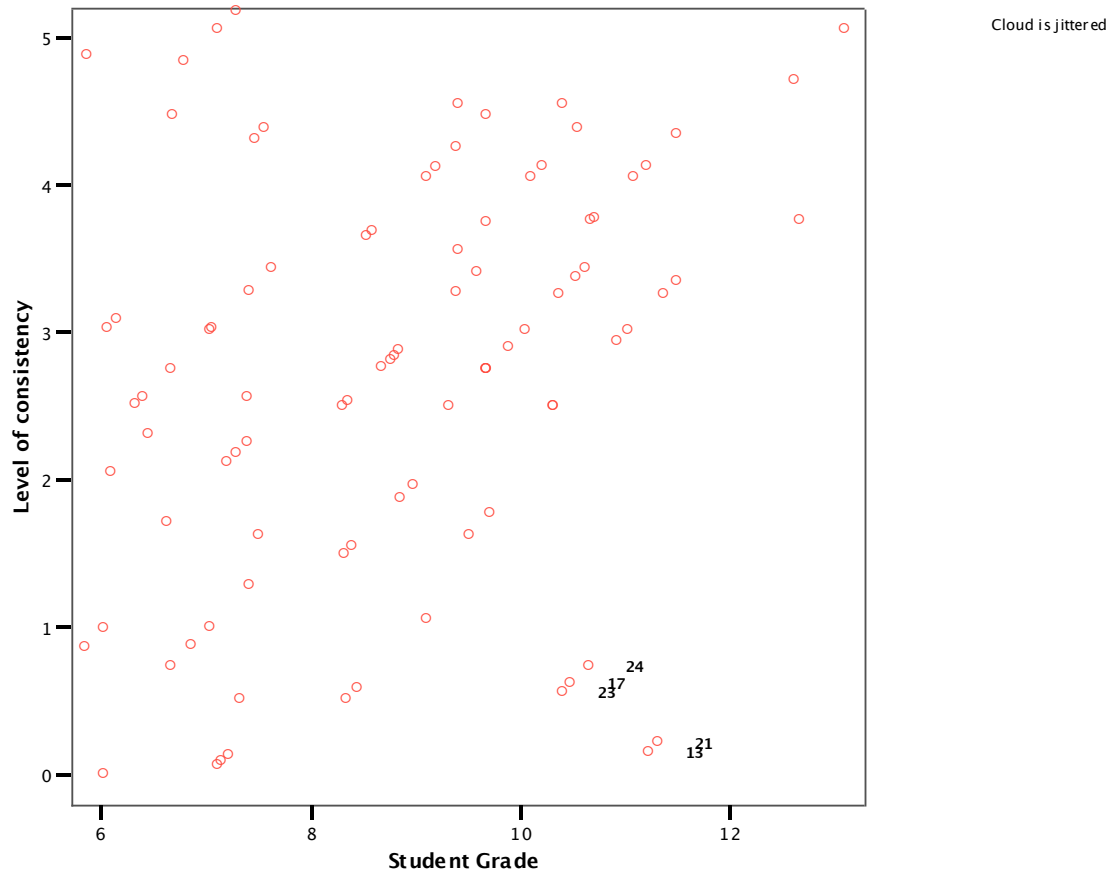
Figure 36: Scatter plot of consistency level vs. student grade showing the low-performing 11th grade students
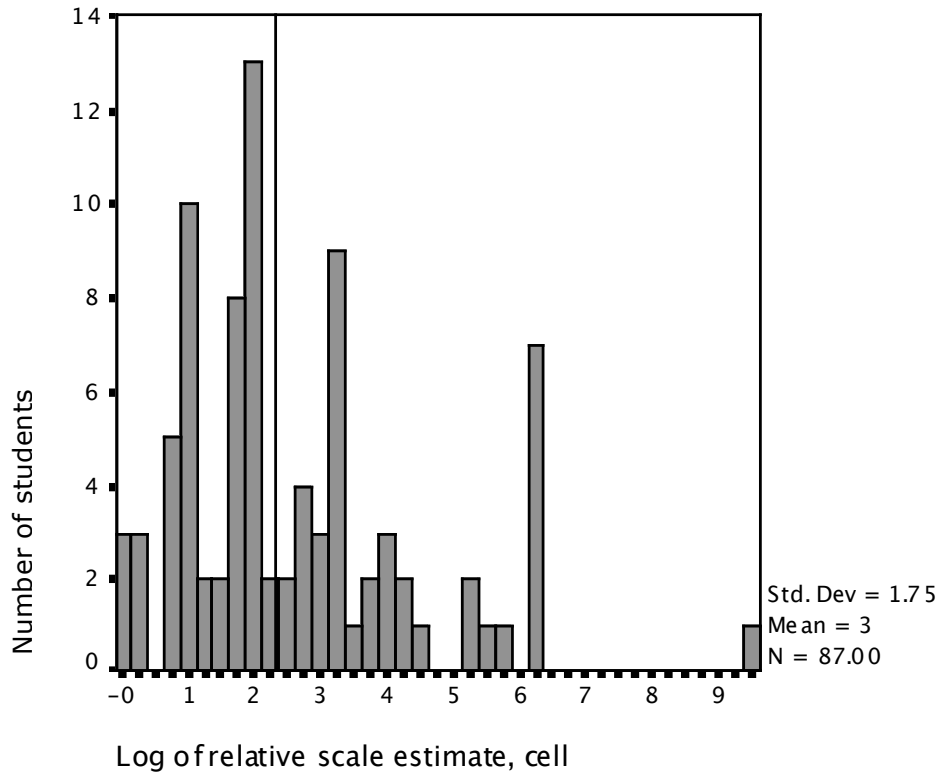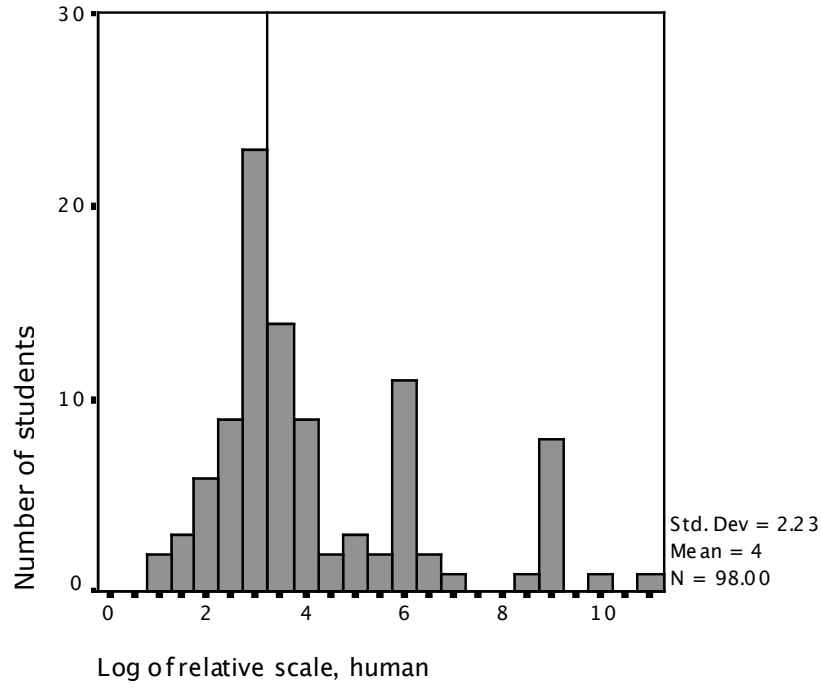
Figure 37: Histogram of relative scale estimates for cell. The line shows the normative value

Figure 38: Histogram of relative scale estimates for human. The line shows the normative value.
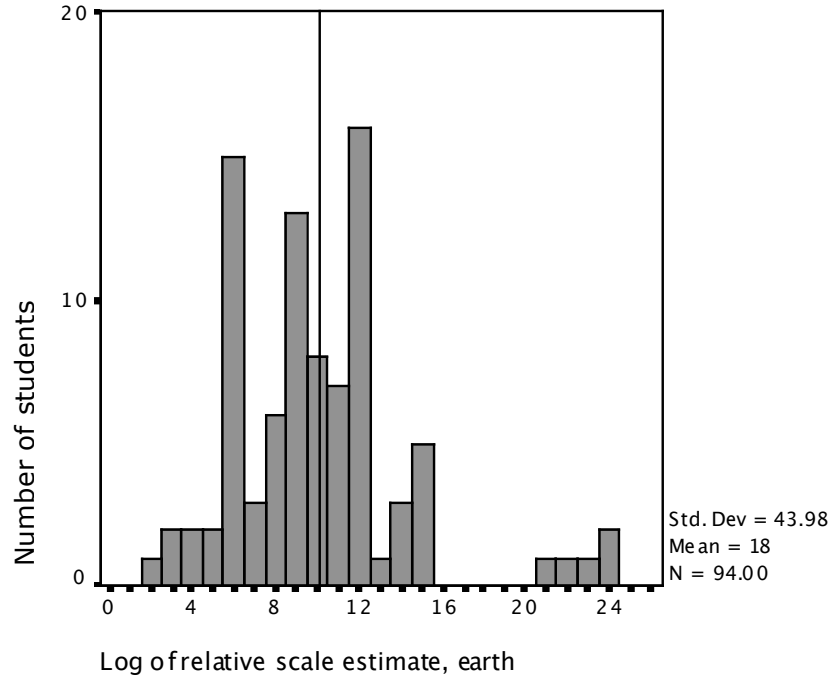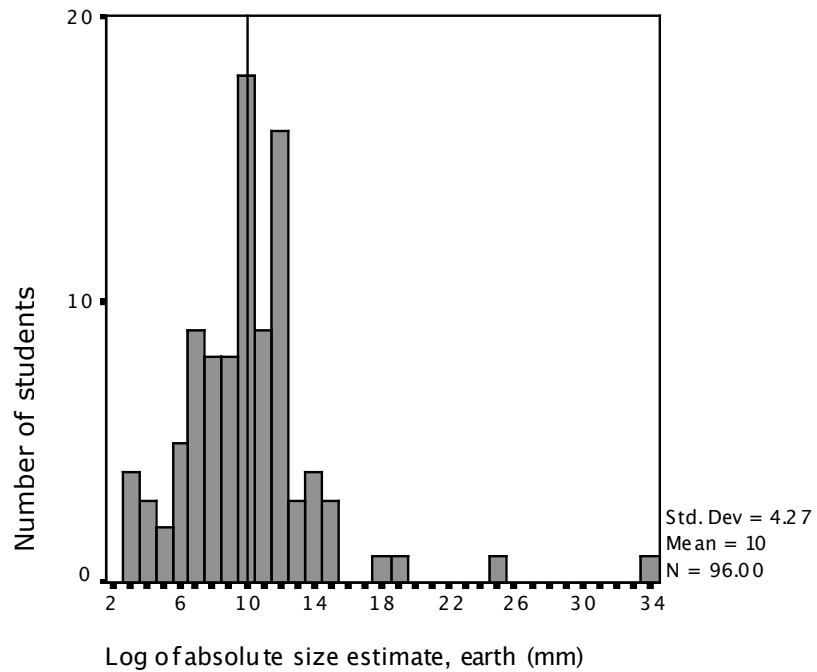

Figure 39: Histogram of relative scale estimates for Earth. The line shows the normative value.

239

Figure 40: Histogram of absolute size estimates for cell. The line shows the normative value.
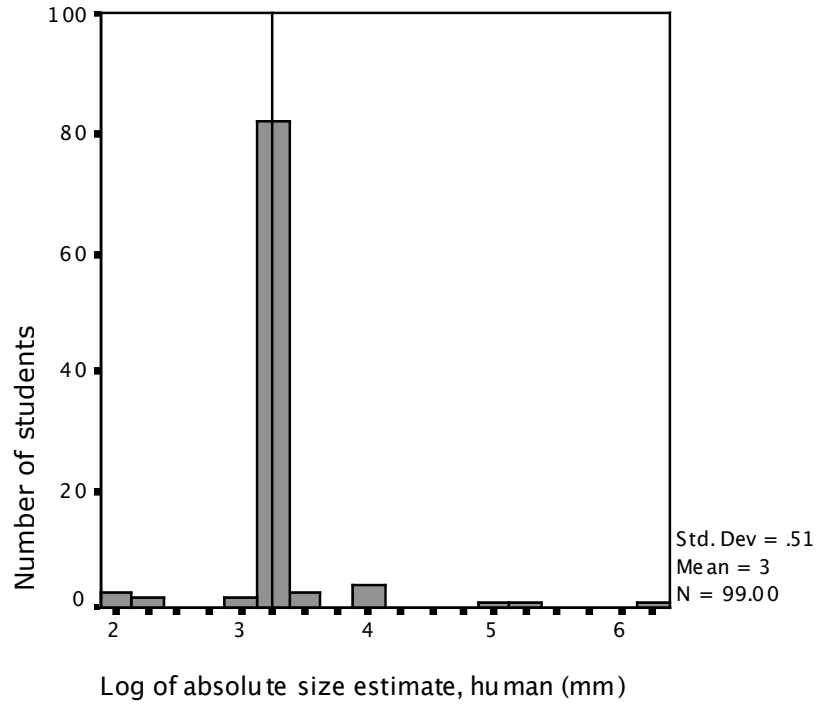
Figure 41: Histogram of absolute size estimates for human. The line shows the normative value.

Figure 42: Histogram of absolute size estimates for Earth. The line shows the normative value.

Appendix J: Graphs of Mean Factual Knowledge by Level on Learning Progression in the Absence of a Focused Curriculum on Size and Scale
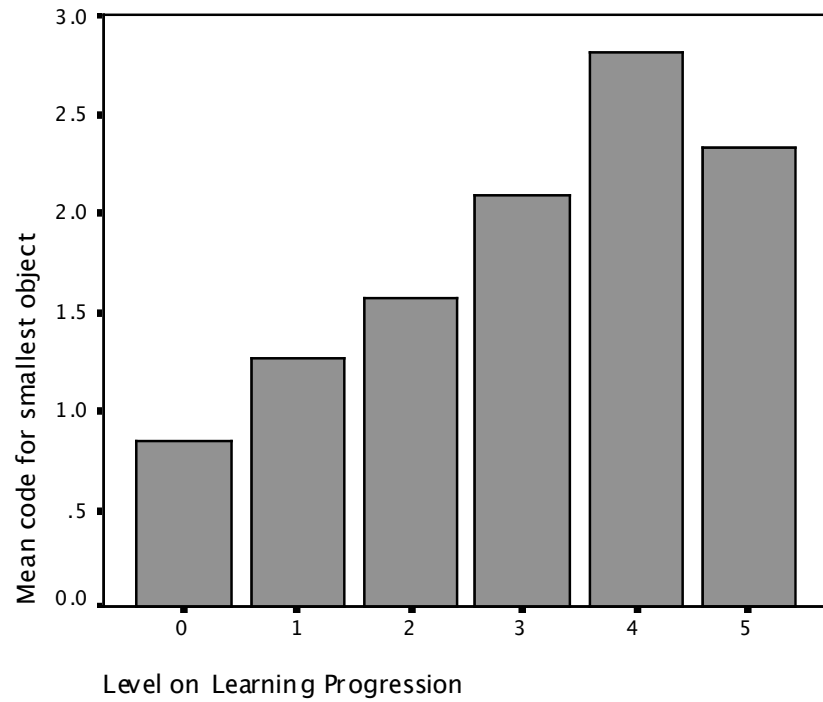
Figure 43: Graph of smallest object by level on LP1
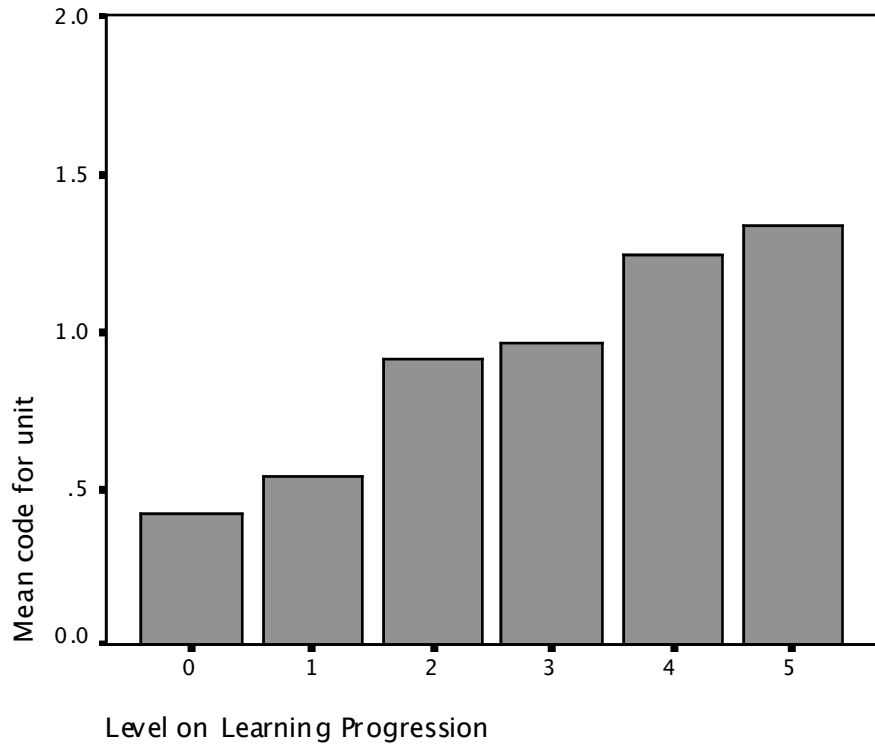
Figure 44: Graph of smallest unit by level on LP1
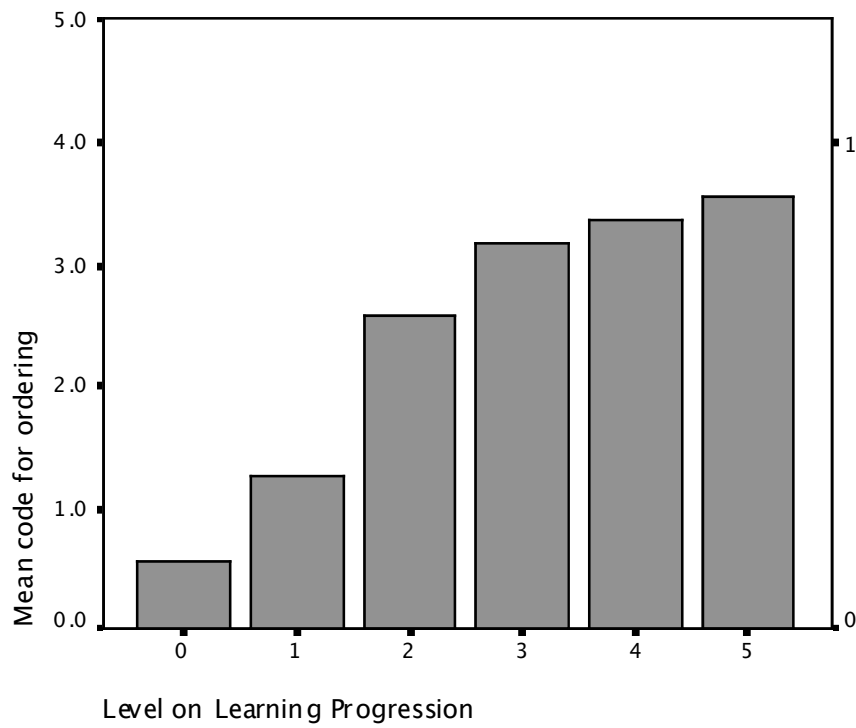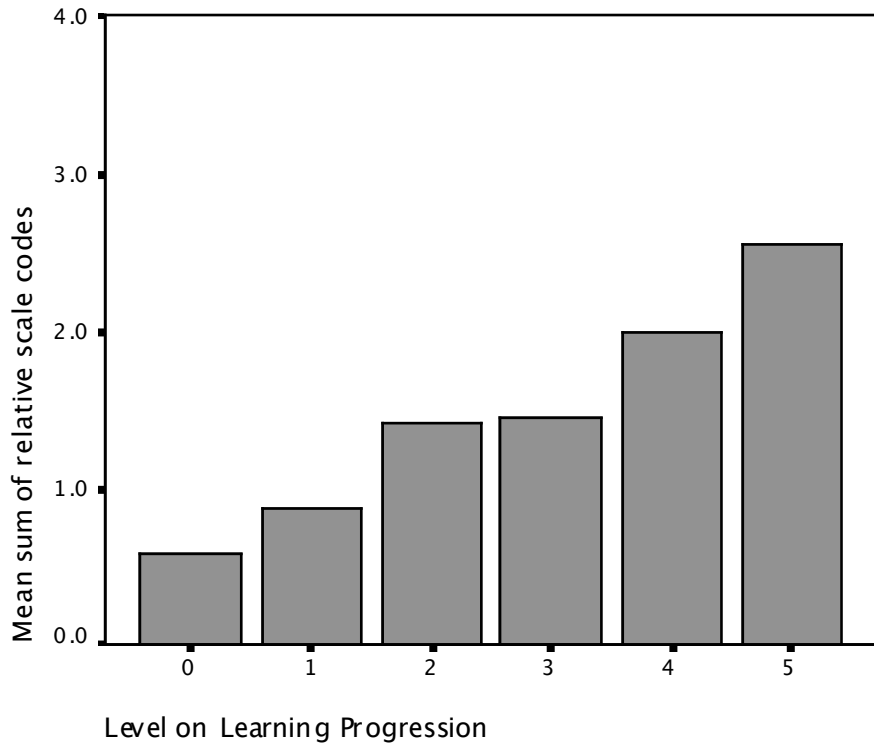

Figure 45: Graph of ordering by level on LP1

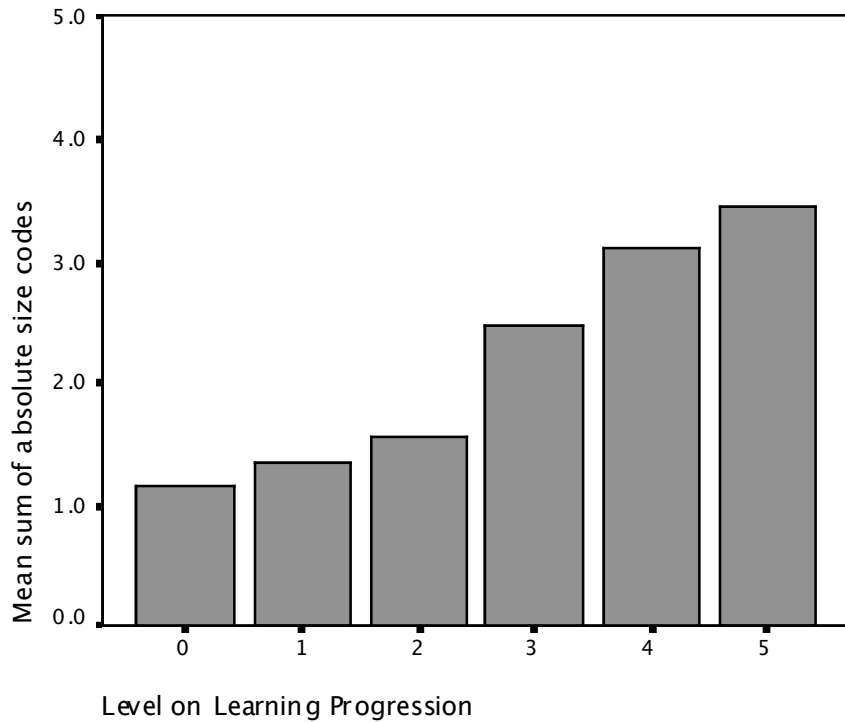*Figure 46:* Sum of relative scale by level on LP1

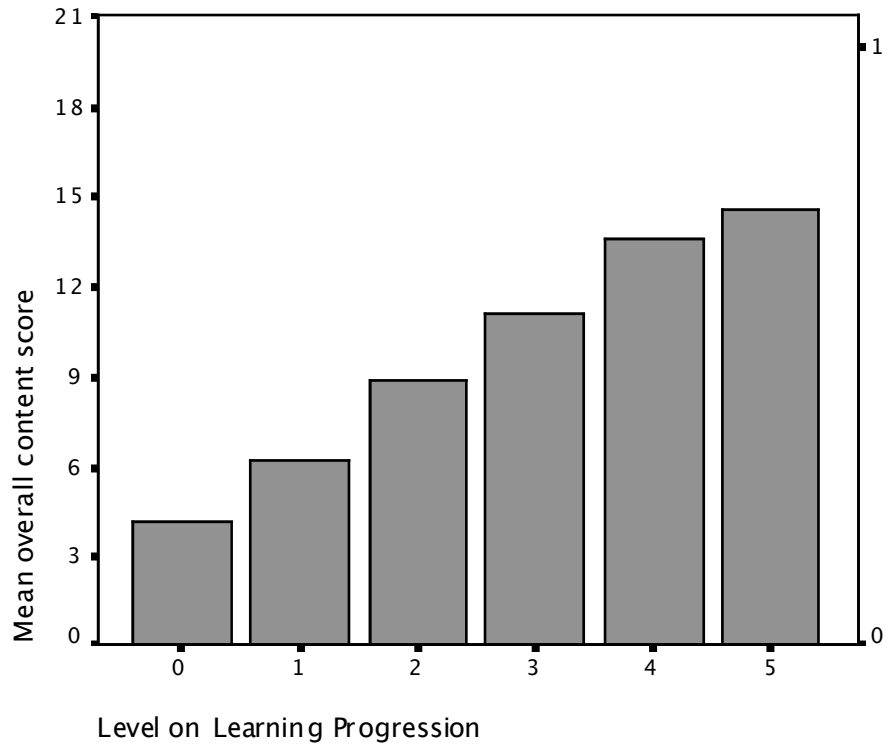*Figure 47:* Graph of sum of absolute size by level on LP1

Figure 48: Overall factual knowledge score by level on LP1
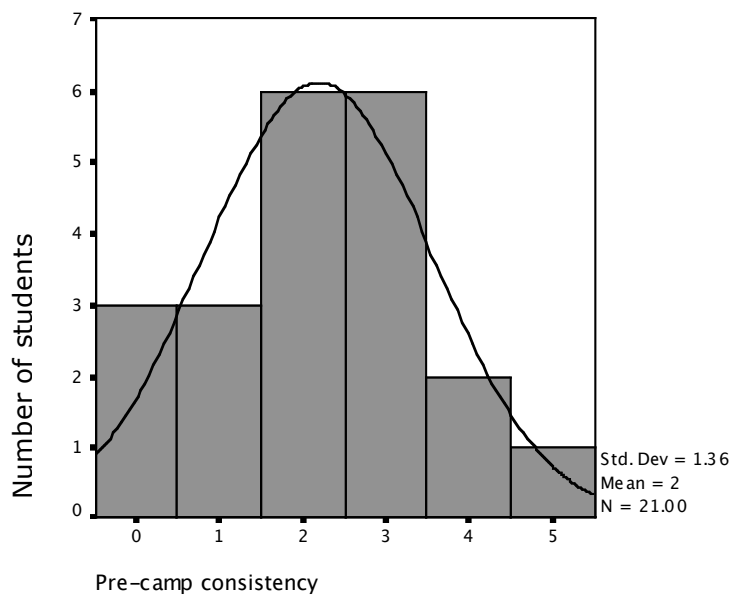
Appendix K: Regression Analysis Graphs
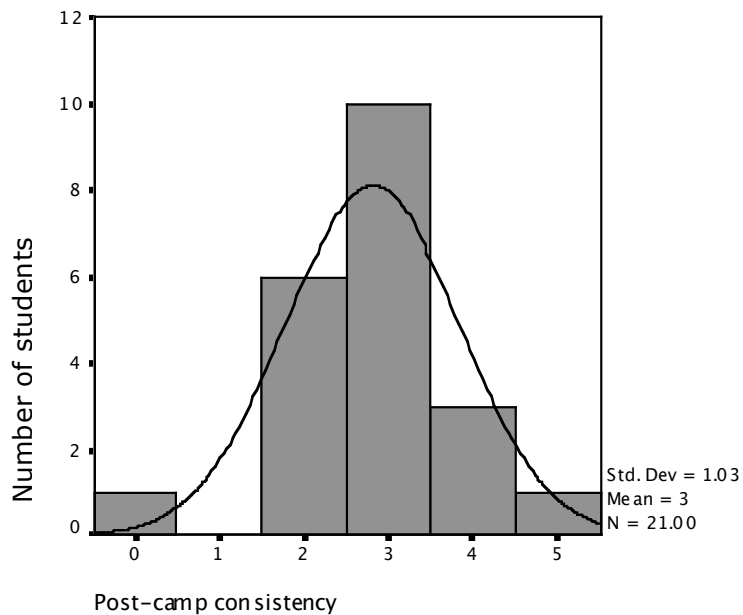

Figure 49: Histogram of pre-camp consistency

Figure 50: Histogram of pre-camp factual knowledge
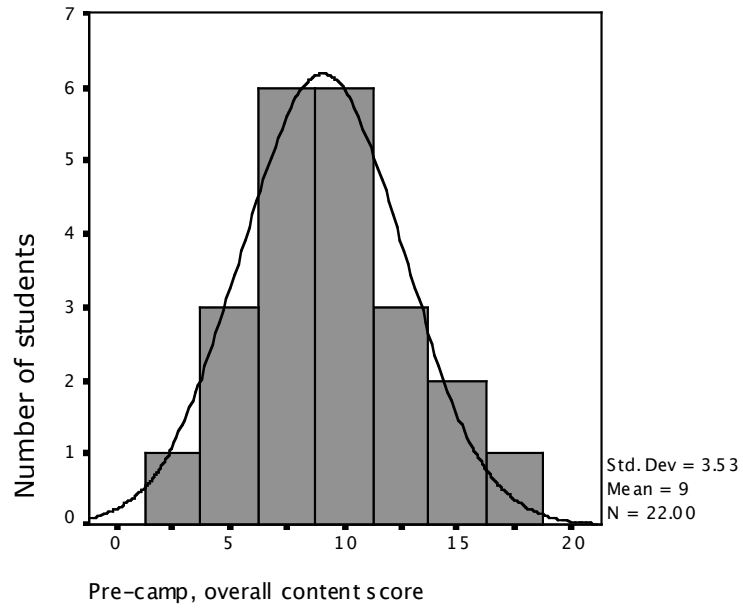
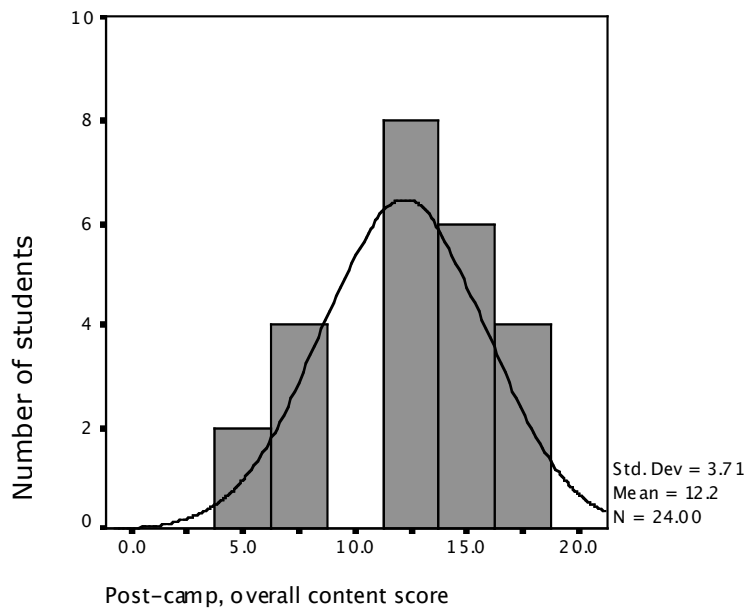*Figure 51:* Histogram of pre-camp factual knowledge

Figure 52: Histogram of post-camp factual knowledge
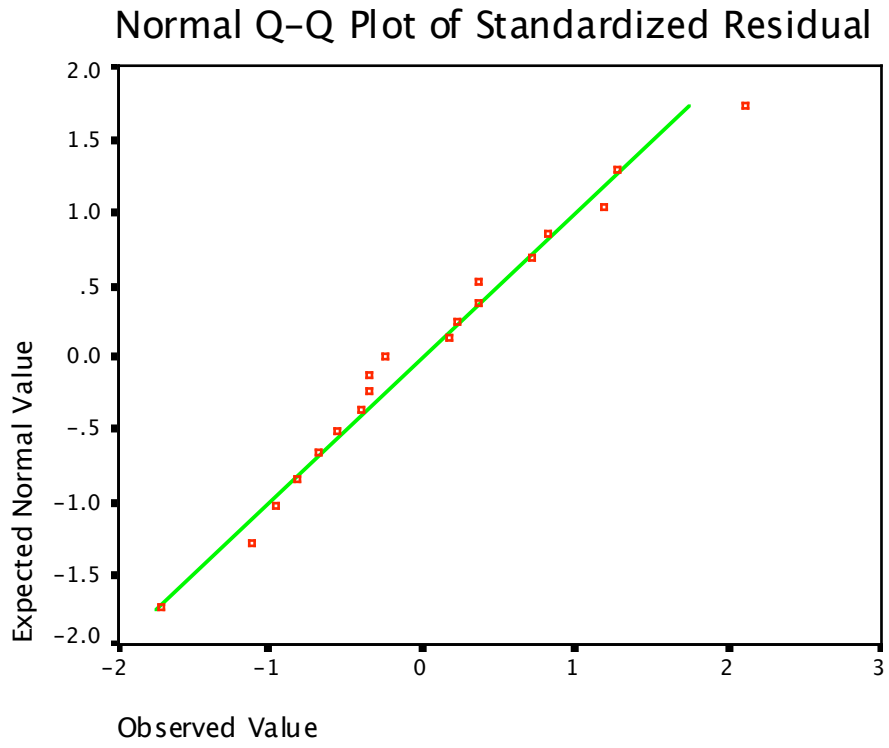
Normal Q–Q Plot of Standardized Residual

Figure 53: Quantile-quantile plot for regression for post-camp factual knowledge

BIBLIOGRAPHY

Alonzo, A.C., & Steedle, J.T. (2008). Developing and assessing a force and motion learning progression. *Science Education*, early view.

Ambert, A., Adler, P. A., Adler, P., & Detzner, D. F. (1995). Understanding and evaluating qualitative research. *Journal of Marriage and the Family, 57*, 879-893.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

American Association for the Advancement of Science. (2009). *Benchmarks for science literacy* (current version). Retrieved Feb. 11, 2009 from http://www.project2061.org/publications/bsl/online/index.php.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, C. W. (2008, February). *Conceptual and empirical validation of learning progressions*. Presented at the Meeting on Advancing Research on Adaptive Instruction And Formative Assessment, sponsored by the Center on Continuous Instructional Improvement (CCII). Philadelphia, PA.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart & Winston.

Barufaldi, J.P., & Dietz, M.A. (1974). The performance of children on visual observation and comparison tasks. *Science Education, 59* (2), 199-205.

Batt, C. A., Waldron, A. M., & Broadwater, N. (2008). Numbers, scale and symbols: the public understanding of nanotechnology. *Journal of Nanoparticle Research, 10*(7), 1141-1148.

Bazant, Z. P. (2002). *Scaling of structural strength*. New York: Taylor & Francis Books.

Bell, B., Osborne, R., & Tasker, R. (1985). Finding out what children think. In R. Osborne & P. Freyberg (Eds.), *Learning in science: Implications of children's science*. (pp. 151-165). Auckland, New Zealand: Heinemann Publishers.

Berryman, S. (2004). Democritus. *The Stanford Encyclopedia of Philosophy* (Fall 2004 Edition), E. N. Zalta (Ed.). Retrieved March 15, 2007, from http://plato.stanford.edu/archives/fall2004/entries/democritus

Beyer, C. J., Delgado, C., Davis, E. A., & Krajcik, J. S. (in press). Investigating teacher learning supports in high school biology curricular programs to inform the design of educative curriculum materials. *Journal of Research in Science Teaching*.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brook, A. Briggs, M., & Driver, R. (1984). *Aspects of secondary students' understanding of the particulate nature of matter*. Leeds: University of Leeds And Children's Learning in Science Project. Centre for Studies in Science and Mathematics Education.

Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Jr., Lindquist, M. M., & Reys, R. E. (1980). National assessment: A perspective of students' mastery of basic mathematics skills. In M. M. Lindquist (Ed.), *Selected issues in mathematics education* (pp. 215-227). Chicago: National Society for the Study of Education and Reston, VA: National Council of Teachers of Mathematics.

Case, R. (1996). Introduction: Reconceptualizing the nature of children's conceptual structures and their development in middle childhood. In R. Case, Y. Okamoto, et al. (Eds.), *The* role of central conceptual structures in the development of children thought. *Monographs of the Society for Research in Child Development, 61*(1-2). Chicago, IL: University of Chicago Press.

Case, R., Stephenson, K.M., Bleiker, C., & Okamoto, Y. (1996). Central spatial structures and their development. In R. Case, Y. Okamoto, et al. (Eds.), The role of central conceptual structures in the development of children thought. *Monographs of the Society for Research in Child Development, 61*(1-2). Chicago, IL: University of Chicago Press.

Castellini, O. M., Walejko, G. K., Holladay, C. E., Thiem, T. J., Zenner, G. M., & Crone, W. C. (2007). Nanotechnology and the public: Effectively communicating nanoscale science and engineering concepts. *Journal of Nanoparticle Research, 9*(2), 183-189.

Chalmers, A. (2005). Atomism from the 17th to the 20th century. *The Stanford Encyclopedia of Philosophy* (Fall 2005 Edition). E. N. Zalta (Ed.). Retrieved March 15, 2007, from http://plato.stanford.edu/archives/fall2005/entries/atomism-modern.

Chi, M. T. H., & Ceci, S. J. (1987). Content knowledge: Its role, representation and restructuring in memory development. In H. W. Reese (Ed.), *Advances in Child Development and Behavior* (Vol. 20, pp. 91-142). New York: Academic Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science 5*,121–152.

Clark, F., & Kamii, C. (1996). Identification of multiplicative thinking in children in grades 1-5. *Journal for Research in Mathematics Education, 27* (1), 41-51.

Clements, D. H. (2003). *Preface.* In D. H. Clements and G. Bright (Eds.), *Learning and teaching measurement: 2003 Yearbook* (pp. xi-xiii). Reston, VA: National Council of Teachers of Mathematics.

Clements, D. H., & Stephan, M. (2004). Measurement in pre-K to grade 2 mathematics. In D. H. Clements & Sarama, J. (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics* (pp. 299-317). Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, D., Ferrell, J., & Johnson, N. (2002). What very small numbers mean. *Journal of Experimental Psychology: General, 131*(3), 424-442.

Cohen, J., & Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Hillsdale, N.J.: Lawrence Erlbaum Associates. (Second edition)

Confrey, J. (1991). Learning to listen: A student's understanding of powers of ten. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 111–138). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Dahl, J., Anderson, S. W., & Libarkin, J. C. (2004). *Digging into Earth Science: Alternative conceptions held by K-12 teachers*. Retrieved February 7, 2007, from http://newton.bhsu.edu/eps/dahletal2004v2.pdf.

DeLoache, J. S., Peralta de Mendoz, O. A., & Anderson, K. N. (1999). Multiple Factors in Early Symbol Use: Instructions, Similarity, and Age in Understanding a Symbol-Referent Relation. *Cognitive Development, 14,* 299-312.

DeLoache, J. S., Uttal, D. H., & Rosengren, K. S. (2004). Scale errors offer evidence for a perception-action dissociation early in life. *Science, 304*(5673), 1027-1029.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the Computer Age* (pp. 49-70). Hillsdale, NJ: Lawrence Erlbaum Associates.

diSessa, A., & Cobb. P (2004). Ontological innovation and the role of theory in design experiments. *The Journal of the Learning Sciences, 13*(1), 77-103.

Duschl, A., Schweingruber, H., & Shouse A. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press. [Viewed electronically]

Eames, C., &Eames, R. (1977). *Powers of ten* [Motion picture]. United States: IBM.

Ericsson, K. A., Krampe, R. T., & Tesch-Roemer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363-406.

Feynman, R. P. (1963). *The Feynman lecture on Physics*. Reading, MA: Addison-Wesley.

Flores, F. (2003). Representation of the cell and its processes in high school students: an integrated view. *International Journal of Science Education, 25*(2), 269-286.

Gattis, M. (2001). Space as a basis for abstract thought. In M. Gattis (Ed.), *Spatial schemas and abstract thought.* Cambridge, MA: MIT Press. (pp. 1-12).

Gattis, M., & Dupeyrat, C. (1999). Spatial strategies in reasoning. In W. Schaeken, G. De Vooght, & A. Vandieredonck (Eds.), *Deductive reasoning and strategies.* (pp. 153-175)

Gelman, S. A., & Ebeling, K. S. (1989). Children's use of nonegocentric standards in judgments of functional size. *Child Development, 60* (4), 920-932.

Gibson, C., Ostrom, E., & Ahn, T.-K. (2000). The concept of scale and the human dimensions of global change: a survey. *Ecological Economics, 32* (2), 217-239.

Gilbert, J. K., De Jong, O., Justi, R., Treagust, D. F., & Van Driel, J. H. (2002). Research and development for the future of chemical education. In J. K. Gilbert, O. de Jong, R. Justi, D. F. Treagust, & J. H. Van Driel (Eds.), *Chemical education: Toward research-based practice* (pp. 391-408). Dordrecht, The Netherlands: Kluwer Academic Press.

Goodwin. L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation and Counseling and Development, 36*(3), 181-191.

Graham, F. K., Ernhart, C. B., Craft, M., & Berman, P. W. (1964). Learning of relative and absolute size concepts in preschool children. *Journal of Experimental Child Psychology, 1*, 26-36.

Griffiths, A.K., & Preston, K.R. (1992). Grade-12 students' misconceptions relating to fundamental characteristics of atoms and molecules. *Journal of Research in Science Teaching, 29*(6), 611-628.

Hiebert, J. (1981). Cognitive development and learning linear measurement. *Journal for Research in Mathematics Education, 12*(3), 197-211.

Hapkiewicz, A. (1999). Naïve Ideas in Earth Science. *MSTA Journal 44*(2), 26-30.

Harrison, A.G., & Treagust, D.F. (1996). Secondary students' mental models of atoms and molecules: Implications for teaching chemistry. *Science Education 80*(5), 509-534.

Hawkins, D. (1978). Critical Barriers to Science Learning. *Outlook 29*, 3-23.

Hiebert, J., & Carpenter, T. (1992). Learning and teaching with understanding. In Grouws, D. (Ed.), *Handbook of research on mathematics teaching and learning*. New York: Macmillan.

Hiebert, J., & Wearne, D. (1986). Procedures over concepts: The acquisition of decimal number knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 199-223). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hook, C. M., & Rosenshine, B. V. (1979). Accuracy of teacher reports of the classroom behavior. *Review of Educational Research, 49*(1), 1-12. 1

Inhelder, B., & Piaget, J. (1969). *The early growth of logic in the child: Classification and seriation* (E. A. Lunzer and D. Papert, Trans.) New York: W. W. Norton & Co. (Original work published 1959.)

Jones, M. G., & Taylor, A. R. (2009). Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching, 46*(4), 460-475.

Jones, M. G., Taylor, A., & Broadwell, B. (in press). Concepts of scale held by students with visual impairment. *Journal of Research in Science Teaching.*

Jones, M. G., Taylor, A., Minogue, J., Broadwell, B., Wiebe, E., & Carter, G. (2007). Understanding scale: Powers of ten. *Journal of Science Education and Technology, 16*(2), 191-202.

Jones, M.G., Tretter, T., Taylor, A., & Oppewal, T. (2008). Experienced and novice teachers' concepts of spatial scale. *International Journal of Science Education, 30*(3), 409 – 429.

Joram, E. (2003) Benchmarks as tools for developing measurement sense. In D. H. Clements and G. Bright (Eds.), *Learning and teaching measurement: 2003 yearbook* (pp. 57-67). Reston, VA: NCTM.

Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. *Review of Educational Research, 68*(4), 413-449.

Kennedy, C., Wilson, M., Draney, K., Tutunciyan, S., & Vorp, R. (n.d.). *ConstructMap* [computer software]. Berkeley, CA: Berkeley Evaluation and Assessment Research Center.

Kenney, P. A., & Kouba, V. L. (1997). What do students know about measurement? In P. A. Kenney & Sliver, E. A. (Eds.), *Results from the sixth mathematics assessment of the National Assessment of Educational Progress* (pp. 141-164). Reston, VA: National Council of Teachers of Mathematics.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academies Press.

Kind, V. (2004). *Beyond appearances: Students' misconceptions about basic chemical ideas,* 2nd ed. School of Education, Durham University, UK. Retrieved Jan. 6, 2009 from http://www.rsc.org/education/teachers/learnnet/pdf/LearnNet/rsc/miscon.pdf.

Klahr, D. & Siegler, R. S. (1978). The representation of children's knowledge. In H. Reese and L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 12) (pp. 62-116). New York: Academic Press.

Krajcik, J. S., McNeill, K. L., & Reiser, B. J. (2008). Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education, 92*(1), 1-32.

Lamon, S. (1994). Ratio and proportion: Cognitive foundations in unitizing and norming. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 89-122). Albany, NY: State University of New York Press.

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

Lesh, R., Post, R., & Behr, M. (1988). Proportional reasoning. In J. Hiebert and Behr, M. (Eds.), *Number concepts and operations in the middle grades* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.

Levin, S. A. (1992). The problem of pattern and scale in ecology. *Ecology, 73*(6): 1943-67.

Light, G., Swarat, S., Park, E. J., Drane, D., Tevaarwerk, E., & Mason, T. (2007). *Understanding undergraduate students' conceptions of a core nanoscience concept: Size and scale*. Proceedings of the First International Conference on Research in Engineering Education. June 2007, Honolulu, HI.

Lindquist, M. M., & Kouba, V. L. (1989). Geometry. In M. M. Lindquist (Ed.), *Results from the Fourth Mathematics Assessment of the National Assessment of Educational Progress* (pp. 44-54). Reston, VA: National Council of Teachers of Mathematics.

Linn, M. C., Davis, E. A., & Eylon, B.-S. (2004). The scaffolded knowledge integration framework for instruction. In M. C. Linn, E. A. Davis & P. Bell (Eds.), *Internet Environments for Science Education* (pp. 47-72). Mahwah, NJ: Lawrence Erlbaum Associates.

Lloyd, R. "Metric mishap caused loss of NASA orbiter." CNN.com (September 30, 1999). Retrieved February 23, 2009 from http://www.cnn.com/TECH/space/9909/30/mars.metric.02/

Lubienski, S. T. (2003). Is our teaching measuring up? Race-, SES-, and gender-related gaps in measurement achievement. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement: 2003 Yearbook* (pp. 282-292). Reston, VA: National Council of Teachers of Mathematics.

Markovits, Z., Hershkowitz, R., & Bruckheimer, M. (1989). Number sense and nonsense. *The Arithmetic Teacher, 36* (6), 53-55

Martin, T., & Schwartz, D. L.(2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science, 29* (4), 587-625 .

McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: Operational momentum in nonsymbolic arithmetic. *Perception & Psychophysics, 69*(8), 1324-1333.

McKusick V. A. (1969). On lumpers and splitters, or the nosology of genetic disease. *Perspectives in Biology and Medicine 12*(2), 298-312.

Melkman, R., Tversky, B., & Baratz, D. (1981). Developmental trends in the use of perceptual and conceptual attributes in grouping, clustering, and retrieval. *Journal of Experimental Child Psychology, 31*, 470-486.

Mislevy R. J., Steinberg, L. S., Almond R. G., Haertel, G. D., & Penuel, W. R. (2003). *Leverage   points for improving educational assessment*. PADI Technical Report. Menlo Park, CA: SRI International.

National Center for Education Statistics (1996). *Pursuing excellence: Initial findings from the Third of International Mathematics and Science Study*. Washington, DC: National Center for Education Statistics.

National Center for Education Statistics (2003). *TIMSS 2003 8th-grade mathematics concepts and mathematics items* (measurement). Retrieved Feb. 13, 2009 from http://nces.ed.gov/timss/pdf/TIMSS8_Math_ConceptsItems_4.pdf.

National Council of Teachers of Mathematics. (1989). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics. [Electronic resource]

National Research Council (1996*). National science education standards*. Washington, D.C.: National Academy Press. [Viewed electronically]

Nelson, K., & Benedict, H. (1974). The comprehension of relative, absolute and contrastive adjectives by young children. *Journal of Psycholinguistic Research, 3*, 333-341.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. New York, NY: Cambridge University Press.

Onyancha, K. M., Lee, Y., Choi, J., Draney, K. L., & Anderson, C. W. (2009, April). *Secondary students' accounts of carbon-transforming processes before and after instruction*. Paper presented at the Annual Conference of the National Association for Research and Science Teaching, Garden Grove California, April 18-21, 2009.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks: Sage.

Pea, R. D. (1988). Practices of distributed intelligence and designs for education. In Salomon, G. (Ed), *Distributed cognitions: Psychological and educational considerations.* Cambridge & New York: Cambridge University Press, pp. 47-87.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Pellegrino, J., Krajcik, J., Stevens, S., Swarat, S., Shin, N., Delgado, C., et al. (2008). Using Construct-Centered Design to align curriculum, instruction, and assessment development in emerging science. In G. Kanselaar, V. Jonker, P.A. Kirschner, & F. Prins, (Eds.). Proceedings from ICLS '08: *International perspectives in the Learning Sciences: Creating a learning world* (vol. 3, pp. 314-21). Utrecht, the Netherlands: International Society of the Learning Sciences.

Petrov, A. A., & Anderson, J.R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review, 112* (2), 383– 416.

Phillips, W.C. (1991) Earth science misconceptions, *Science Teacher, 58*(2), 21-23.

Piaget, J. (1974). *Understanding causality* (D. Miles & M. Miles, Trans.). New York: W. W. Norton.

Piaget, J. (1983). Piaget's theory. In P. Mussen (ed.). *Handbook of child psychology*. 4th edition. Vol. 1. New York: Wiley.

Piaget, J., & Inhelder, B. (1971). *The child's conception of space*. (F. J. Langdon & J. L. Lunzer, Trans.). London: Routledge & Kegan Paul Ltd.

Piaget, J., Inhelder, B., & Szeminska, A. (1960). *The child's conception of geometry*. (E.A. Lunzer, Trans.). New York: Basic Books.

Plummer, J. D. (2006). *Students' development of astronomy concepts across time.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Post, T. R., Cramer, K. A., Behr, M., Lesh, R., & Harel, G. (1993). Curriculum Implications of research on the learning, teaching, and assessing of rational number concepts. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.) *Rational numbers: An integration of research.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Rennie, L., & Johnston, D. (2004). The nature of learning and its implications for research on learning from museums. *Science Education 88* (Supplement 1): S4-S16.

Resnick, L.B. (1992). From protoquantities to operators: Building mathematical competence on a foundation of everyday knowledge. In G. Leinhardt, R. Putnam, & Hattrup, R. (Eds.), *Analysis of arithmetic for mathematics teaching.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Resnick, L.B., & Singer, J.A. (1993). Protoquantitative origins of ratio reasoning. In T.P. Carpenter, E. Fennema, & T.A. Romberg (Eds.), *Rational numbers: An integration of research.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Rutherford, F. J., & Ahlgren, A. (Eds.) (1990). *Science for all Americans.* New York: Oxford University Press. [Viewed electronically]

Schank, P., Wise, A., Stanford, T., & Rosenquist, A. (2009). *Can high school students learn nanoscience? An evaluation of the viability and impact of the Nanosense curriculum.* Technical report. Menlo Park: SRI International.

Schmidt, W. H., McKnight, C. C., & Raizen (1997). *A splintered vision: An investigation of U.S. science and mathematics education.* Boston: Kluwer.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition & Instruction, 16,* 475-522.

Scott, P., Mortimer, E. & Aguiar, O. (2006). The tension between authoritative and dialogic discourse: A fundamental characteristic of meaning making interactions in high school science lessons. *Science Education, 90,* 605-631.

Sera, M., & Smith, L.B. (1987). Big and little: "Nominal" and relative uses. *Cognitive Development, 2,* 89-111.

Skinner, B. F. (1971). *Beyond freedom and dignity.* New York: Knopf

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

Smith, C. 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 the atomic molecular theory. Focus Article. *Measurement: Interdisciplinary Research and Perspectives, 14,* 1-98.

Snir, J., Smith, C., & Grosslight, L. (1993). Conceptually enhanced simulations: A computer tool for science teaching. *Journal of Science Education and Technology, 2*(2), 373-388.

Song, M., & Quintana, C. (2009). WIIS: Multimodal simulation for exploring the world beyond visual sense. *ACM Conference on Human Factors in Computing Systems (CHI 2009) Extended Abstracts* (April 2009, Boston, MA), 4699-4704.

Sowder, J. (1992). Estimation and number sense. In Grouws, D. (Ed.), *Handbook of research on mathematics teaching and learning*. New York: Macmillan.

Stephan, M., & Clements, D. H. (2003). Linear and area measurement in prekindergarten to grade 2. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement: 2003 Yearbook* (pp. 3-16). Reston, VA: National Council of Teachers of Mathematics.

Stevens, S. Y., Delgado, C., & Krajcik, J. (in press). Developing a hypothetical multidimensional learning progression for the nature of matter. *Journal of Research in Science Teaching*.

Stevens, S. Y., Sutherland, L., & Krajcik, J. (in press). *The big ideas of nanoscale science and engineering*. Arlington, VA: NSTA Press.

Strutchens, M. E., Martin, W. G., & Kenney, P. A. (2003). What students know about measurement: Perspectives from the National Assessment of Educational Progress. In D. H. Clements and G. Bright (Eds.), *Learning and teaching measurement: 2003 Yearbook* (pp. 195- 207). Reston, VA: National Council of Teachers of Mathematics.

Sullivan, J. (n.d.). *Cells alive: How big is a...?* [computer simulation]. http://www.cellsalive.com/howbig.htm.

Tang, K-S. (2009, April). *Multimodality and inscriptions: Essential components of a multi-representational framework to study student discourse*. Paper presented at the Annual Conference of the National Association for Research and Science Teaching, Garden Grove California, April 18-21, 2009.

Tang, K-S., & Delgado, C. (in preparation). *Multimodality and inscriptions: Components of a multi-representational framework to study student discourse*.

Trend, R. D. (2001). Deep time framework: A preliminary study of U. K. primary teachers' conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching, 38* (2), 191-221.

Tretter, T. R. (2005). Godzilla versus scaling laws of physics. *The Physics Teacher, 43*, 530-532.

Tretter, T. R. (2006). Conceptualizing nanoscale. *The Science Teacher, 73*(9), 50-53.

Tretter, T. R., Jones, M. G., Andre, T., Negishi, A., & Minogue, J. (2006). Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching, 43*(3), 282-319.

Tretter, T.R., Jones, M.G., & Minogue, J. (2006). Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching, 43*(10), 1061-1085.

Tversky, A. (1969). Intransitivity of preference. *Psychological Review, 76*, 31–48.

Vasilyeva, M., & Huttenlocher, J. (2004). Early development of scaling ability. *Developmental Psychology, 40,* (5), 682-690.

Vergnaud, G. (1988). Multiplicative structures. In J. Hiebert & M. Behr (Eds.), *Number concepts and operation in the middle grades* (pp. 141-161).

Vosniadou, S. (2003). Exploring the relationships between conceptual change and intentional learning. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 377-406). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Vygotsky, L. S. (1985). *Thought and language*. Cambridge, MA: M.I.T. Press.

Waldron, A. M., Sheppard, K., Spencer, D., & Batt, C. A. (2005). Too small to see: Educating the next generation in nanoscale science and engineering. In C. S. S. R. Kumar, J. Hormes, & C. Leushner (Eds.), *Nanofabrication towards biomedical applications* (pp. 375-389). New York: Wiley.

Waldron, A. M., Spencer, D., & Batt., C. A. (2006). The current state of public understanding of nanotechnology. *Journal of Nanoparticle Research, 8*, 569-575.

White, R. T., & Gunstone, R. F. (1992). *Probing understanding*. London: Falmer Press.

Wiedtke, W. (1990). Measurement. In J. N. Payne (Ed.), *Mathematics for the young child* (pp. 229-249). Reston, VA: National Council of Teachers of Mathematics.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, M., & Bertenthal, M.(Eds.) (2005). *Systems for state science assessment*. Washington, DC: National Academies Press. [Viewed electronically]