# Tracking Students' Understanding of the Particle Nature of Matter

by

**Joi DeShawn Merritt**

**A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Education)
in The University of Michigan
2010**

**Doctoral Committee:**

**Professor Joseph S. Krajcik, Chair
Professor Brian P. Coppola
Associate Professor Elizabeth A. Davis
Assistant Professor Amelia W. Gotwals, Michigan State University**

## Dedication

To My Family

**ACKNOWLEDGMENTS**

My dissertation study would not have been possible without the support and guidance of so many. First, I would like to thank my advisor, Joseph Krajcik, whose support, guidance and feedback have been invaluable during my time as a graduate student. Joe has provided with many opportunities and experiences that have had an amazing impact on me both personally and professionally. I look forward to our future collaborations. I must also thank my other committee members, Betsy Davis, Brian Coppola and Amelia Gotwals. Betsy has always pushed me to think about the connections and what my work means. I thank her for always providing me with wonderful feedback that is insightful and thoughtful and most of all extremely helpful. Brian has pushed me to think about my work in terms of what does it all mean. I thank him for always leaving me with something to think about. Finally, I would like to thank Amelia, whose knowledge of science assessments has been absolutely invaluable. I want to thank her for her help and feedback. I thank you all for pushing me and helping me in ways that you could never know.

I would like to thank all of the IQWST sixth grade teachers that I have had the opportunity to work with. I would especially like to thank those teachers who provided me with the data to make this dissertation possible: Lisa Brody, Jeffrey Downes, Rachelle Fulkerson, Kristie Hannon, Kathleen Keenmon, Jamie Klausing, and Kalonda McDonald.

I must also thank the members of the IQWST project, both past and present that I have had the great fortune of working with. I would especially like to thank all the past and present members of the Chemistry team whose hard work and guidance has been an integral part of my journey.

I would like to send thanks to those who have helped me in completing my dissertation in various and different roles: Molly Yunker, James Hagerty, Yael Baumberger, Jenny Sealy Badee, Serena Salloum, Ashima Mathur, Kathryn Drago and Clara Cahill. Whether it has been to provide me feedback on different stages of my dissertation, help with coding or words of encouragement, you all have kept me going and I want to know how much I appreciate each of you.

Finally, I would like to thank my family for all their love and support. I would like to thank my parents for always being my biggest cheerleaders. Your belief in me has allowed me to make my dreams reality. To my brothers, Philip and Justin, I thank you for always encouraging me, making me laugh when I needed it most and for your support.

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF APPENDICES

# CHAPTER 1

## Introduction

The passage of the No Child Left Behind Act of 2001 (NCLB) has placed an increased emphasis on testing as a means for assessing student knowledge in core subject areas and for teacher accountability. The interpretation of NCLB for testing purpose has been left to each individual state and has resulted in different modes of testing for each individual state. If the ultimate goal of NCLB is to improve teaching and learning, it is not clear how these assessments are able to provide the information needed to facilitate improvements.

*Knowing What Students Know* (NRC, 2001) points out many of the limitations of current assessments. Many current assessments do not provide information to teachers about how to help students to improve, including their strengths and weaknesses or educational interventions that could improve their performance. In addition, these assessments are often not aligned with the curriculum that students experience in the classroom nor do they measure the complex knowledge and skills emphasized in standards. Moreover, they do not capture the growth in student understanding during instruction.

Learning progressions are depictions of students' increasingly sophisticated ideas about a specific knowledge domain over time (Duschl, Schweingruber, & Shouse, 2007; Smith et al., 2006). All learning progressions could be considered hypothetical because the path in which students learning a disciplinary idea is not developmentally inevitable

(Stevens, Delgado & Krajcik, 2010) and are not tied to a particular curriculum. Moreover, learning progressions provide an opportunity to examine students' increasingly sophisticated ideas over the long-term. But how does this translate to the small-scale timeframe of classroom instruction? The development and application of progress variables is one method that has been suggested as a means for addressing this question (Wilson, 2005; Wilson 2009).

Progress variables depict students' increasingly sophisticated conceptions over time, regardless of whether it is a matter of weeks or years. Progress variables are visualized through construct maps, which divides the complex levels of students' increasingly sophisticated understanding into distinguishable levels. Thus, a learning progression could be visualized as a single construct map, or composed of several related construct maps. In addition, progress variables mediate between big ideas and specific concepts and skills being learned and serve as a means for tracking student understanding during instruction (Wilson, 2005; Wilson, 2009). Thus, each unit of instruction contributes to students' progress and necessitates that assessment aligns with one or more progress variables. Once developed, progress variables can be used to provide information to both teachers and students about student progress during instruction (Wilson, 2005; Wilson, 2009; Kennedy, Brown, Draney & Wilson, 2006).

The particle nature of matter is a fundamental scientific concept – a big idea in science. As Smith, Wiser, Anderson and Krajcik (2006) point out, big ideas are powerful in that they are central to the disciplines of science and are the building blocks for learning within a discipline. The particle model of matter serves as the basis for understanding various phenomena, including states of matter, phase changes, and properties of substances. As a result, it has been an intense area of research with numerous studies documenting the difficulties middle, high school, and college students have in understanding of the particle nature of matter (Harrison & Treagust, 2002).

Traditional curriculum materials present the particle nature of matter as a topic, focusing on the history of the atom (Harrison & Treagust, 2002). At the middle school level in the United States, students are often taught the structure of the atom and that the different states of matter are related to the movement and arrangement of atoms (AAAS, 1993). This direct instruction assumes that once presented with the particle model, students will accept it as the correct model. At the high school level for example, a textbook presents the history the atom beginning with the Greek philosophers and ending with the current quantum model of the atom (Holt, Rinehart & Winston, 2006).

Moreover, students find it difficult to learn the particle model using traditional curriculum materials because they present particle concepts to students without helping them to develop these concepts or take into account their prior knowledge. Typically, the particle model of matter is introduced in either a short paragraph, or as a chapter on the atom and the history of the atom (Harrison & Treagust, 2002). Often students do not develop appropriate ideas because they never apply and reapply these ideas to explain phenomena. Several interview studies have suggested the need for the development of learning progressions for the particle nature of matter (Renstrom, Andersson, & Marton, 1990; Johnson, 1998; Nakhleh, Samarapungavan & Saglam, 2005; Liu & Lesniak, 2006; Margel, Eylon & Scherz, 2007). Smith, Wiser Anderson, & Krajcik (2006) have proposed a hypothetical learning progression for matter and atomic-molecular theory that spans kindergarten through eighth grade. The development of progress variables is one method for helping both students and teachers to track student progress in understanding this big idea of science (Wilson, 2009).

As mentioned earlier, assessments and instruction must align with the progress variable. Determining what students know is an inherent challenge faced in developing any assessment. It becomes even more complex when considering assessment in the context of classroom instruction. Often, assessments are developed separate from the curriculum materials that are used during instruction. As *Knowing What Students Know* (NRC, 2001) describes, there is a need to develop a conceptually rich system that links curriculum, instruction and assessment. The latter is the approach used in *Investigating and Questioning our World through Science and Technology (IQWST)* curriculum, where curriculum materials and assessment items were developed simultaneously (Krajcik, McNeill and Reiser, 2008).

Several studies have proposed the development of curriculum materials that focus on students' models of matter and the application of that model to explain macroscopic phenomena (Ben-Zvi, Eylon & Silberstein, 1986; Kozma, Chin & Marx, 2000; Justi & Gilbert, 2002; Harrison & Treagust, 2002; Snir, Smith & Raz, 2003). The sixth grade chemistry IQWST unit, entitled "How can I smell things from a distance?" takes this approach. Specifically, the development of a particle view of matter is the basis for understanding properties, states of matter and phase changes. In addition to the development of particle knowledge, students develop an understanding of the practice of modeling.

**Research Questions**

This research was conducted to track how students' understanding of the particle nature of matter changed as they participated in this contextualized and model-based chemistry unit. In this dissertation, I describe the process of developing and validating a progress variable for the particle model of matter and its use in explaining phase changes, states of matter and properties. Then, I examine the usefulness of this progress variable in tracking middle school students' understanding of the particle model of matter during instruction. Thus, the key research question informing this dissertation is:

- **How does middle school students' understanding of the particle nature of matter change during enactment of a model-based unit?**

The completion of this study provides insight into whether coherent assessment and curriculum fosters student development towards a particle model of matter.

Previously Merritt, Krajcik, & Shwartz (2008) examined pretest to posttest learning gains as well as the models students constructed at specific points throughout the unit to develop an initial progress variable for student understanding of the particle model. However, we had not empirically determined whether the pre/posttest items actually measure the particle model progress variable. The first study is guided by the following question:

- **Are the assessment items valid measures of students' knowledge of the particle model of matter progress variable?**

The calibration study of the progress variable was completed to determine whether assessment items were good measures of the progress variable as well as whether they were reliable and valid measures of the variable. The results of this study were then used to modify the progress variable.

The second study I conducted is the tracking study where the modified progress variable was calibrated so that I could follow students' understanding during instruction . The calibration of the variable was important because it "allows the creation of a calibrated scale to map the growth of students so teachers can track the progress of individual students as they undergo instruction" (Wilson, 2005, p.195). Two sub-questions helped me further understand how students' knowledge progressed throughout the unit. The progress variable allowed me to describe students' knowledge based on the different levels of the construct map. Examination of the relationship between students' initial knowledge of particle concepts and their knowledge at specific time points both during and after instruction allowed me to answer the following sub-question:

- **What knowledge of the particle nature of matter do students bring to the unit and how does this relate to students' progress towards a particle model?**

In other words, this study was conducted to determine whether students' initial understanding of matter influences their learning during instruction. This is important for understanding whether prior knowledge can be a predictor of student development of particle views of matter. In addition, students created models of phenomena throughout the unit. These models are composed of two parts, their drawing and their explanation. A comparison of the drawings and explanations allowed me to describe the relationship between the two parts of the model in terms of student's growth during instruction.

- **What is the relationship between students' drawings and explanations of phenomena?**

**Outline of dissertation**

Chapter 2 is a literature review that focuses on defining what a progress variable is. Then, key aspects related to student understanding of particle theory are discussed and how they relate to progress variables and the studies that were conducted. Research indicates that because the particle nature of matter is an abstract concept, students also need to develop an understanding of models and the practice of modeling (Harrison & Treagust, 1996; Harrison & Treagust 1998; Harrison & Treagust 2000; White & Frederikson, 1998; Schwarz & White, 2005). Moreover, teachers must understand models and their uses as well. Student conceptions are discussed as a starting point for instruction as well as important for the development of progress variables. The design of the sixth grade chemistry unit and its approach for helping students to develop a particle view of matter, including its educative features for teachers is discussed. Because the progress variable was developed to align with the chemistry unit, it also reflects the goals of instruction.

Chapter 3 describes how the key components of the BEAR Assessment System (BAS; Wilson & Sloane, 2000; Wilson, 2005) were utilized for both studies. The focus of this chapter is to describe the four building blocks of the BAS: the construct map, items design, the outcome space and the measurement model. Three of the building blocks are discussed within the context of conducting this study; including the development of the construct map, the item design process, and how the outcome space was developed for the sixth grade unit's assessments. The BAS is also discussed in relationship to the National Research Council's Assessment Triangle (NRC, 2001).

Chapter 4 describes and discusses the Calibration study. The context of the research is described and findings are presented about the importance of determining whether items are valid and reliable measures of the progress variable.

In Chapter 5, I describe and discuss the tracking study, which addresses the overall research question of the study. Findings are presented and discussed related to how an empirically validated progress variable can be used to track student understanding of the particle model during instruction. In addition, I explore the relationship of students' prior knowledge to their performance on subsequent assessments. I also compare students' development of the two components of the model (drawing + explanation) to see what it reveals about their understanding of the particle model of matter.

Chapter 6 presents a summary of the findings of both studies and their contributions to research literature. This chapter focuses on the importance of empirically validating progress variables so that they can be utilized to track students' progress during instruction. It also details how coherent curriculum and assessments can help to students to progress toward a particle view of matter. Suggestions for future research are also presented.

# CHAPTER 2

## Literature Review

### Introduction

In this chapter, I expand the argument for the importance of developing progress variables for tracking middle school students' understanding of the particle nature of matter. Since the implementation of the No Child Left Behind Act of 2001 (NCLB), there has been an increased emphasis on testing as a means for assessing student knowledge in core subject areas, including science. To meet the requirements of NCLB, many states re-evaluated their state science standards to align more closely with those of national standards. Teacher editions of many traditional science textbooks even include how state and national standards align. Both the Benchmarks for Scientific Literacy (AAAS, 1993) and National Science Education Standards (NRC, 1996) include standards related to the particle nature of matter for the middle grades, grades six through eight. Although the standards describe what students should know by the end of middle school, they do not describe in detail how these ideas build upon each other. Moreover, the current tests designed to assess student knowledge are not designed to capture student growth over time (NRC, 2001).

This chapter describes progress variables and their relationship to learning progressions as well as construct maps. I then describe the complexity of the particle nature of matter by examining four important areas related to students learning about the particle nature of matter: models and modeling in chemistry, student (mis)conceptions of matter, and the use of the particle model in teaching. Finally, I describe the curriculum that was specifically designed to help students in their development of a particle view of matter.

## Learning progressions and progress variables

> To be aware of children's existing ideas is important if we are to help children relate the ideas in their own minds to the learning experiences provided, so that sensible new ideas are constructed. We have to relate our teaching to their ideas, since we cannot control what they are thinking (Osborne & Freyberg, 1985, p. 53).

Learning progressions are depictions of students' increasingly sophisticated ideas about a specific domain over time. They are also a means for helping both students and teachers to track students' developing understanding over time (Duschl, Schweingruber, & Shouse, 2007; Smith et al., 2006). Moreover, learning progressions provide a means for thinking about how to present topics to students so that they build on each other through the years. Smith et al.'s progression is based on prior research related to matter and particle theory and focuses on students gaining more sophisticated understanding of matter and its properties as well as applying microscopic explanations to macroscopic phenomena. In addition, this progression identifies which topics are introduced each year and how knowledge is built in relationship to what students have previously learned. Developing a means for tracking students long-term progress for understanding the big ideas of science is important, but how do we track students' increasingly sophisticated understanding of concepts underlying these big ideas, especially within the timeframe of classroom instruction?

Jim Minstrell (2001) proposed "facets of students' thinking" as a means for helping teachers to make instructional decisions. Facets are descriptions of different levels of students' knowledge and/or reasoning strategies as they are learning and are based on prior research. They serve as links between the standards (what students should know and be able to do) and what they actually "seem to know and do" (Minstrell, 2001, p. 426). The different levels of these facets represent qualitatively different levels of student knowledge. Progress variables are a means by which one can combine research and empirical results to define and develop levels of students understanding.

Progress variables are similar to facets in that they represent a range of student thinking about a particular knowledge domain, or construct. A construct "can be part of a theoretical model of a person's cognition…their understanding of a certain set of concepts" (Wilson, 2005, p. 6). Just as in learning progressions, constructs are assumed to range from low to high knowledge of a domain, with increasing complexity in between. Thus, one or more progress variables could be used to track student understanding of a particular construct over time frames as short as a curriculum unit to a learning progression that covers multiple years (Wilson, 2009).

As mentioned in chapter one, progress variables mediate between big ideas and specific concepts and skills being learned and serve as a means for tracking student understanding during instruction (Wilson, 2005). In addition, progress variables allow one to focus on student growth over time in their understanding of a construct (Wilson, 2009). This means that instruction contributes to student progress. Therefore, what students are learning must be clearly defined as well as a theoretical framework for students' progress are necessary to establish the construct validity of an assessment system (Wilson, 2009). Assessments conducted within the context of the classroom serve to make students' thinking visible (NRC, 2001); thus, embedded assessments must also be aligned with the progress variable.

Construct maps are a visual depiction of these variables that divide the levels of complexity into distinguishable levels. A more detailed discussion of construct maps can be found in the next chapter. When a construct map is developed in relationship to innovative curriculum, the construct map also represents the goals of teaching (Wilson, 2009). Furthermore, progress variables serve as a framework for assessment development. This study focuses on the development of the particle model of matter (PMM) progress variable, which was developed in relationship to a particular curriculum intervention. Thus, the framework for the development of the instructional materials must match that of the development of the assessment items (Wilson, 2009).

In sum, learning progressions are a means for determining how to support student learning of the big ideas of science. They are hypothetical in that they are hypotheses for how student understanding changes over time. They are big picture in that they cover learning over large time frames. Moreover, they are research-based in that they also take into account prior research related to student understanding of a particular domain. Progress variables are one method for development of assessments for tracking student growth over time. They are versatile as they can serve as a means for tracking students' progress during instruction, or for longer time frames like those of learning progressions.

The development of the construct map is important because they are the visual depictions of the progress variable and assessments for tracking students' understanding must align with construct map. A construct map could represent a single learning progression, or one or more construct maps could be used to represent the levels of a single learning progression (Wilson, 2009), though more complex relationships among construct maps for a learning progressions could also exist. The importance of these different relationships is that they influence the way in which the assessment is structured (Wilson, 2009).

The PMM progress variable has been developed to determine how student understanding of the particle nature of matter changes during instruction. It was developed in relation to a particular curriculum, which focuses on student development of a particle view of matter using models of matter that they construct. The next section focuses on the practice of modeling and its importance for student development of a particle model of matter.

## Models and Modeling in Chemistry

> The great game of science is modeling the real world, and each scientific theory lays down a system of rules for playing the game. The object of the game is to construct valid models of real objects and processes. Such models comprise the core of scientific knowledge. To understand science is to know how scientific models are constructed and validated. The main objective of science instruction should therefore be to teach the modeling game (Hestenes, 1992, p. 732).

An important tool for scientists is the scientific model. Scientists use scientific models to think about, explain, and predict phenomena in the world. For this research, a scientific model is defined as a representation of objects, theories, relationships, or dynamic events used to predict, test, and explain phenomena. By defining a scientific model as such, key functions that a model possesses are highlighted. Without these functions, any representation could be defined as a model. For example, a teacher could have an air ramp and cart in the classroom. The air ramp and cart by themselves are not models. They become a model when used as a demonstration to explain a scientific concept, such as friction on an inclined plane.

First, models are not a reflection of the real world, but a way to explain an aspect, or aspects of phenomena (Schwarz et al., 2009). This is why the simplicity of a model is important – it focuses on the important entities necessary to explain a particular aspect of the phenomenon. For example, a student may use F=ma to calculate ideal force on an object. Inherent in the use of this law is that it applies to an ideal situation, which does not account for the real world effects of friction.

Second, models are limited in scope. A model may be used to explain one aspect of a phenomenon, but is limited in its ability to explain other aspects of that same phenomenon. These limitations are important in understanding a model as a set of assumptions "that are designed to help them [scientists] think about how to explain some aspect of reality" (Snir et al., 2003, p. 798). Students may use ball and stick models to demonstrate the bond angles in a compound, but they do not accurately demonstrate the constant motion of atoms and molecules.

In addition, different models can be used to explain different aspects of the same phenomena (Snir et al., 2003). Because models have limitations, more than one model can be used to explain the same phenomenon. This is not to say that one model is correct and the other is not. Instead, this emphasizes that each model highlights a different aspect of the same phenomenon. This also indicates that there can be several different types of models. For example, a ball and stick model, a simulation and a two-dimensional model could all be used to discuss the arrangement of atoms into molecules.

Finally, models are evaluated on their ability to predict as well as explain phenomena (Snir et al., 2003; Schwarz et al., 2009). The accuracy and plausibility of a model are important in evaluating a model because they inform the limitations of the model as well as revisions that need to be made to a model as new observations and analyses of phenomena provide greater insights into the elements, and the relationships of those elements, of phenomena. The assumption that matter is made of particles is not enough to explain why the odor of a perfume sprayed in the front of a room can be smelled in the back of that same room. An additional assumption must be made in the model in that these particles are in motion.

Many studies emphasize the importance of students understanding models and the process of modeling in order to better understand scientific phenomena (Harrison & Treagust, 1996; Harrison & Treagust 1998; Harrison & Treagust 2000; White & Frederikson, 1998; Schwarz & White, 2005; Schwarz et al., 2009). These studies emphasize the importance of students not only understanding the different types of models that can be developed for a single phenomenon, but also the nature of models and the practice of modeling. In particular, Schwarz et al. (2009) found that elementary and middle school students who are engaged in model-based curricula have the ability to construct models of abstract phenomena that could be used to explain and predict phenomena. In addition, students were able to revise their models as they learned more about phenomena. However, challenges emerged in that students still saw modeling activities as a normal part of schooling. Furthermore, they saw it as a means of providing answers to the teacher and not a communication of their own ideas.

Others have shown or promoted using, creating, and understanding the nature of models as a means to help students understand physical phenomena (Grosslight et al, 1991; Hestenes, 1992; Vosniadou, 1994; Harrison & Treagust, 1998; Justi & Gilbert, 2002; MacKinnon, 2003; Saari & Viiri, 2003; Mikelsis-Seifert & Leisner, 2005; Schwarz and White, 2005). Students are introduced to abstract topics like particle theory through the use of multiple models. Teachers introduce different models (i.e. physical models, simulations and 2-D models) based on the model's ability to explain different aspects of the same phenomenon. The various models utilized to represent specific phenomena confuse many students. This is especially true for the teaching of abstract concepts in which analogies and models can be confused with reality. Moreover, teachers should help students to shared and unshared attributes of models and assist students in determining where a model breaks down (Harrison & Treagust, 1996).

In chemistry, students must learn and make meaning of new terms, symbols, graphs, tables and several other representations (Justi & Gilbert, 2002; Kozma et al., 2000). Wu and Shah (2004) found that visualization is key to understanding chemical representations and conducting research. Visuospatial thinking is the ability of learners to construct and make sense of both visual and spatial information. Moreover, it has been found that visuospatial abilities partially explain achievement in chemistry (Baker & Talley, 1972; Wu & Shah, 2004). Thus, helping students in understanding visual representations and the scientific concepts related to these representations can help students in understanding chemistry and chemical concepts (Barak & Dori, 2001; Ealy, 1999). This is where scientific models can be used to help students understand chemistry and chemical concepts.

In the field of chemistry, more than one model is used to explain different aspects of the same phenomena. A water molecule can be described using a ball-and-stick model, structural formula or line-angle drawing. Each of these different models demonstrates different aspects of the particle model. For example, the ball-and-stick model shows the space that molecules take up as well as the angles at which the atoms in the molecules bond, while structural formulas show the different atoms and how many of these atoms make-up a molecule. However, students have difficulty understanding how to interconvert between ball-and-stick models, structural formulas and line-angle drawings (Ferguson & Bodner, 2006; Kozma et al., 2000; Baker & Talley, 1972). These difficulties stem from students not understanding how chemical concepts can be explained through the use of these different models in addition to their visuospatial capabilities. Thus, for students to understand the significance of these different models they must be provided with the skills to translate between different representations and how phenomena inform the creation of these models.

Therefore, students need help in understanding models used to explain particle theory. In addition, students need to have instruction that helps them to understand why the particle model helps them in understanding the particle nature of matter. As a result, the sixth grade unit includes opportunities for students to both understand models and the particle model. But there also needs to be a way to identify where students' understanding is at during instruction. In order to accomplish this, we must be able to track students' knowledge. The unit includes activities at specific points in which students create models of the same phenomena, which illustrate their understanding of the phenomena and how their understanding of the phenomena has changed during instruction. In other words, students' models of matter created at specific points during instruction provide the means for tracking students' understanding of matter.

As discussed earlier, progress variables are the means for tracking student understanding, which are represented through construct maps. The construct map represents students understanding from naïve to more sophisticated. Therefore, it is important to understand common student errors, or misconceptions to help define and distinguish the levels of the construct map (Wilson, 2009). In the next section, I explore student conceptions of matter that research has illuminated.

## (Mis)Conceptions of The Particle Nature of Matter

The particle nature of matter is a fundamental concept for learning and understanding many physical and chemical processes. Novick and Nussbaum (1978) studied students' ideas about the particle nature of matter as it relates to gases. They found that students did not internalize ideas related to the vacuum concept (empty space), the intrinsic motion of particles or the interaction between particles during a chemical change. Other studies have shown that students assign macroscopic properties of substances to the atoms/molecules that compose the substance (Ben-Zvi, Eylon & Silberstein, 1986; Nakhleh, 1992; Lee et al, 1993). Moreover, learners and many adults hold non-normative science ideas regarding the structure of matter. Misconceptions are non-normative science ideas about a phenomenon.

Many of the misconceptions students possess have been documented (Driver et al., 1985; Driver et al., 1994). For example, students misconstrue mass and size of an object. For instance, students hold the idea that a balled up piece of aluminum has more mass than a flat piece of aluminum foil. In addition, there are areas in which students hold on to their non-normative models of matter despite instructional strategies used (Driver et al., 1994). If the goal is to track students' understanding as they learn the particle model of matter, how does one address these misconceptions in teaching?

Research cites conceptual change as the means for dealing with misconceptions, which is the replacement of misconceptions with expert ideas. Conceptual change seems to treat misconceptions as ideas that interfere with students learning (Smith, diSessa, & Roschelle, 1993). Research indicates that students must undergo a conceptual change in order for students to move from a continuous view of matter to a particle view (Nussbaum & Novick, 1982; Lee et al, 1993; Vosniadou, 1994; Harrison & Treagust, 2002; Niaz et al, 2002). Moreover, students' conceptions are constantly changing due to both their experiences and instruction (Strike & Posner, 1992).

Researchers have also suggested that some misconceptions related to the particle model are developed during instruction. In some instances, instruction can be enveloped in prior malformed misconceptions or learned as a misconception due to the student's method of learning. As Harrison & Treagust (2002) note, "this practice of providing token evidence and making the assumption that students will accept the new ideas as fact is not an uncommon phenomenon in teaching and learning chemistry" (p. 191). Ben-Zvi, Eylon and Silberstein (1986) designed a comparison study aimed at investigating students' views of matter. They found that although classroom discussions involved the correct terminology (i.e., atoms, molecules), one-third of students still attributed properties of a substance to its atoms. For example, this type of view would mean students would come to the conclusion that gold atoms are yellow in color because a gold brick is yellow in color.

Lee et al. (1993) also completed a comparison study, which found that students were applying observable properties to molecules. This study also found that students had no concept of empty space between molecules, molecules being the same size as tiny objects (i.e., dust, bacteria, cells) and that molecules are not constantly moving. ). These studies, as well as other studies focused on students' understanding of the particle nature of matter often mention the mismatch between the language students use for describing phenomena/matter and students' views of matter (Ben-Zvi, Eylon & Silberstein, 1986; Lee et al, 1993; Driver et al, 1994; deVos & Verdonk, 1996; Johnson, 1998; Renstrom et al.; 1990; Taber, 2003).

Strike and Posner (1992) also determined that students bring their own mental models of phenomena to the classroom, which many not be fully developed or articulated. When a student's conception is met with teacher demonstrations, students will reconcile their own conception with accepted scientific content to produce an alternative conception (Harrison & Treagust, 1996).

Smith, diSessa, & Roschelle (1993) argue that instead of looking at misconceptions as ideas that must be changed, that they be viewed as a starting point for students' development of expertise. Therefore, the goal of instruction would not be to replace misconceptions, but to "provide the experiential basis for complex and gradual processes of conceptual change" (Smith, diSessa, & Roschelle, 1993). This applies to both students learning about big ideas such as the particle nature of matter, but also to students applying these ideas to phenomena such as phase changes. Thus, conceptual change becomes not a means for replacing ideas, but a means of building knowledge. Research focused on students' understanding of the particle nature of matter have found students' misconceptions to be the starting point for learning (Nussbaum & Novick, 1982; Nussbaum, 1985; Vosniadou, 1994; Nakleh, Samarpungavan & Saglam, 2005; Claesgens, Scalise, Wilson & Stacy, 2010). For as Minstrell (2001) notes:

> Peoples' explanations generally progress from a description of the phenomenon or description of procedures for creating the effect, through identification of relevant concepts, to understanding particular mechanisms of causality, to a more model-like weaving of concepts, mechanisms, and relations among factors. (p. 424)

Therefore, there is a need to track students' understanding of the particle model, as well as how they develop more sophisticated explanations of phenomena using this model.

In sum, students' understanding of matter originates both from everyday experiences and classroom instruction. Therefore, students' conceptions should not be looked upon as misconceptions, but as resources for developing greater knowledge. In addition, student misconceptions have provided insight into the development of the PMM construct map, which provides the opportunity to: 1) track student understanding during instruction, 2) to determine students' prior knowledge, and 3) to gain an understanding of how this knowledge changes through this study. Moreover, the ability to track student progress also serves as insight into how instruction impacts these changes.

Yet, it is not enough to understand how students develop and use particle views of matter to explain different phenomena, it is important to develop curriculum materials that attends to student misconceptions. Tracking student conception is not only a resource for determining student progress; they could also serve as a resource for determining instructional practices to help students reach a more expert understanding. On the other hand, research indicates that instruction can also be a source of student misconceptions. Thus, it is important to examine the impact of instructional practices on student understanding of particle theory for both curriculum development and development of the PMM construct map.

**How is the particle model used in teaching?**

It is often taken for granted that students will just take up the particle model during instruction. Most curricula in the United States make no mention of alternative models students may hold. The only mention of alternative ideas relates to the delineation of the history of the atom found in many traditional textbooks (Harrison & Treagust, 2002). This is a very scientific view of how the particle model developed, focusing on the scientists and the experiments that led to the current quantum model of the atom.

Besides the lack of acknowledgement of alternative student conceptions, there are issues related to the language used in discussing the model (Ben-Zvi, Eylon & Silberstein, 1986; Lee et al, 1993; Driver et al, 1994; deVos & Verdonk, 1996; Johnson, 1998; Renstrom et al.; 1990; Taber, 2003). The particle model is important for explaining macroscopic phenomena using microscopic terms. For example, water boiling is explained as the rapid movement of water molecules from the liquid phase to the gaseous phase. In addition, the terms atom and molecule are often used interchangeably to describe materials on a microscopic level, which is often confusing for students and sometimes teachers (Taber, 2000). For example, students are taught that elements are made up of atoms. Oxygen is an element that is made up of oxygen atoms, but these atoms are always found as oxygen molecules (two oxygen atoms bonded together). This becomes confusing for many students because they conflate the definition of element with the term atom. As Harrison & Treagust (2002) note, the "…semantic differences between students' and teacher's meanings for commonly used terms in science are a source of alternative conceptions" (p. 525).

Textbooks also tend to introduce hybrid models, which hinder students developing understanding about the nature of model and their validity in respect to content (Justi & Gilbert, 2002; Taber, K., 2003). These hybrid models mix macroscopic descriptions of phenomena with particle and molecular ideas. For instance, they will show a diagram of water illustrating water molecules within a drawing of liquid water. This can result in students thinking of substances being made up of molecules/particles, but they cannot identify the molecules as being of that substance (Renstrom et al., 1990; Liu & Lesniak, 2006; Johnson & Papageorgiou).

Curricula can also introduce "teaching models" that do not contribute to student understanding (Justi & Gilbert, 2002; Taber, K., 2003). "Teaching models" are not based on scientific evidence, nor are they used for explaining scientific phenomena. Instead, they are analogies that teachers use in an attempt to help students understand scientific content. For example, a teacher will draw a cloud to represent the electron cloud surrounding the nucleus of an atom. However, representing the electron cloud as an actual cloud does not match up with what scientists know. Often, when teachers present students with models, they focus on the content of the model, not the nature of models and modeling and/or without emphasizing role of modeling in developing what is known about the chemical behavior of matter. Few efforts have been made to improve teachers' pedagogical content knowledge in this area (Justi & Gilbert, 2002).

Research demonstrates that teachers should explicitly present models to students as thinking tools (Grosslight et al, 1991; Hestenes, 1992; Vosniadou, 1994; Harrison & Treagust, 1998; Harrison & Treagust, 2000; Justi & Gilbert, 2002; MacKinnon, 2003; Saari & Viiri, 2003; Mikelsis-Seifert & Leisner, 2005; Schwarz and White, 2005;

Schwarz et al., 2009). Thus, teachers need to be aware of students' evolving conceptions through explanations of model meaning, model-based problem solving and students' constructing models, using models and exploring different models. This means tracking students' models of phenomena and allowing them to practice using multiple models to understand the strengths and weaknesses of different model types. In effect, teachers need

> [A] comprehensive view of: (i) the nature of a model in general; (ii) how their students construct their own mental models and how the resulting expressed models can be constructively used in class; (iii) how to introduce scientific consensus models in their classes; (iv) how to develop good teaching models – those that are created with the specific purpose of facilitating students' understanding of scientific consensus models; and, finally and most significantly, (v) how to conduct modeling activities in their classes [Gilbert 1997 ](Justi & Gilbert, 2002, p. 52).

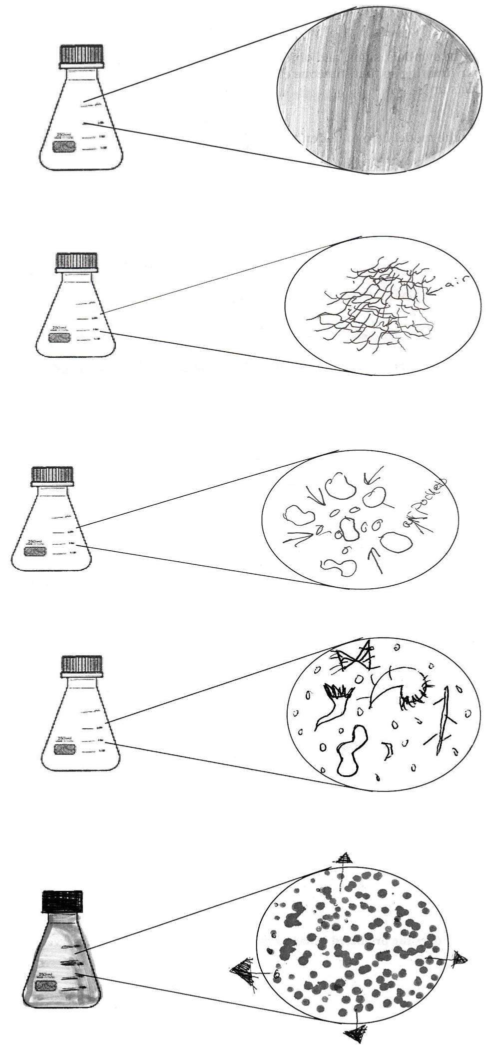
Thus, in order to help students understand the particle model, students as well as teachers need to understand the nature of models and participate in the practice of modeling. In addition, teachers need to be aware of the various paths students take to a particle model. In particular, teachers need to be aware of the hybrid models of matter that students develop during instruction.

**Hybrid Models of Matter**

Students' understanding of the particle model is extremely complex and varied. Several studies indicate that students' development of a particle model takes different paths and that as students' content knowledge grows, students' models can change - both towards a particle model and back to their initial understanding (Renstrom et al., 1990; Johnson, 1998; Nakleh et al., 2005; Liu & Lesniak, 2006; Margel, Eylon, & Scherz, 2008). For example, Renstrom et al. (1990) found that students represented six distinct conceptions of matter: matter as a) a homogeneous substance, b) substance units, c) substance units with "small atoms", d) aggregate of particles, e) particle units and f) system of particles. Studies by Johnson (1998), Nakleh et al. (2005) and Margel et al. (2008) found that students created some similar model types to those found by Renstrom et al. (1990). Our initial study showed students' developing five model types (see Figure 1) that were classified as (Merritt, Rogat, & George, 2006):

- Type 1: Continuous (no space)
- Type 2: Continuous with empty space
- Type 3: Mixed, particles and clouds or particles and lines
- Type 4: Particles, including everyday ideas (germs, water in air)
- Type 5: Particle

These different model types are similar to those found by Renstrom (1990), Johnson (1998), Nakleh et al. (2005), Margel et al. (2008) and Claesgens et al. (2010). Of particular interest are the Type 3 and 4 models, which represent a hybrid model. These hybrid models have been recognized by each of the aforementioned studies as an opportunity for helping students to develop understanding of the particle model of matter.



Type 1: Continuous

Type 2: Continuous with empty space

Type 3: Mixed

Type 4: Particles, including everyday

Type 5: Particle

**Figure 2.1. Examples of the five types of drawings generated by students**

As mentioned previously, students understanding of the particle model involves not only an understanding of the particle nature of matter, but the use of this model to explain phenomena. Johnson (1998) found students' models correspond with their explanation of phenomena, such that a continuous model relates to macroscopic explanations of phenomena while a complete particle model relates to microscopic explanations of phenomena. Margel et al. (2008) found a similar pattern of students moving from a macroscopic to molecular model as well as macroscopic to molecular explanations within a 3-year curriculum in Israel. On the other hand, Nakleh et al. (2005)

found that students were able to give microscopic explanations for familiar substances, but their understanding was fragmented based on particular substance or phenomena. Tien, Teichert & Rickey (2007) as well as Taber (2008) found that students' molecular views of matter did not match their explanations of phenomena. Claesgens et al. (2010) found that students could have hybrid reasoning in which they apply macroscopic observations of phenomena to explanations on the molecular level. Thus, it is unclear whether students' explanations of phenomena become increasingly sophisticated as their mental model of matter because their written explanations do not always match the sophistication of their drawing. For example, a student may develop a particle view of matter, but explain certain phenomena using a macroscopic explanation.

In sum, students' understanding of the particle model of matter is two-fold. It includes both the development of the particle model of matter and how students apply their understanding of this model to explain phenomena. Moreover, it is important to track this knowledge to better understand the different paths students take in coming to a more sophisticated understanding of this complex scientific concept. Thus, one goal of this research is to develop a progress variable for students' understanding of the particle nature of matter, which incorporates what research has revealed about student conceptions. In addition, the curriculum that serves as the setting for this study was designed to address the aforementioned issues related to teaching particle theory, including teaching the practice of modeling, and using prior knowledge as a basis for instruction.

## Curriculum

The *Investigating and Questioning our World through Science and Technology* (IQWST) project (Krajcik, McNeill and Reiser, 2008) takes the approach of building student's ideas over time. Thus, in this unit students develop and use the particle model to explain phenomena, such as states of matter, phase changes, and properties. For example, the particle model can be used to explain a property like boiling point. The boiling point of a substance occurs at a fixed temperature and involves the rapid evaporation of anywhere in a bulk liquid. During heating, particles gain energy and move faster. The energy of these molecules is enough to overcome the attractive forces of the other liquid molecules so that it goes from the liquid to the gas phase.

The IQWST curriculum has also been designed to attend to curricular coherence (Shwartz, Weizman, Fortus, Krajcik and Reiser, 2008). Curriculum coherence is "presenting a complete set of interrelated ideas and making connections among them explicit" (Roseman, Linn, & Koppal, 2008). IQWST achieves curricular coherence within a unit by contextualizing inquiry within a driving question, sequencing learning goals and concurrently developing learning activities that build upon each other through the use of scientific practices (Shwartz, Weizman, Fortus, Krajcik and Reiser, 2008). Simultaneously, assessment items were developed based on the sequenced learning goals, activities and practices of the unit.

The unit is designed so that learning the particle model of matter is contextualized through the use of a driving question. The development of a driving question (Krajcik & Blumenfeld, 2006) serves to produce a context for students to learn about scientific phenomena. The development of the driving question also serves to anchor students learning within a context. In situated cognition, knowledge is a product of the situation and activities from which they originate and meaning is derived from the context of their use. Thus, context plays a vital role in situated cognition in that it “shows students the legitimacy of their implicit knowledge and its availability as scaffolding in apparently unfamiliar tasks” (Brown et al., 1989, p. 38). In our unit, students’ knowledge is the basis for instruction and discussion. Thus, student models provide a window into student thinking. The driving question “How can I smell things from a distance?” provides the anchoring context for all of the lessons and is revisited throughout the unit (Krajcik & Blumenfeld, 2006). Students’ create models throughout the curriculum so that they can apply both their real-world experiences and what they have learned through experiencing phenomena to their answering of the driving question. Moreover, the anchoring context is revisited at specific points of the curriculum as students gain greater knowledge and understanding of concepts related to the phenomena studied. The models students create related to the anchoring context also serve as the means for tracking student understanding during instruction.

Second, the unit involves the creation of student artifacts - the models that students create. Students experience various phenomena throughout this eight-week unit to help them to gain knowledge and understanding of the different aspects of the particle nature of matter. Peer-to-peer and whole class discussions are utilized to help students discuss and critique their models and understand scientific concepts, as well as serving as opportunities to address misconceptions students may have. The instructional materials include descriptions of the types of discussions to have with students, including suggested questions to ask questions. Topics of discussions range from discussing the models students to create for explaining a particular phenomenon to making sense of phenomena on a macroscopic and microscopic level. These discussions were strategically designed to provide the opportunity for both students and teachers to understand students' views of matter.

Research indicated that students should experience multiple models of phenomena to better understand the strengths and weaknesses of different models to explain phenomena. Our approach provides students with opportunities for using multiple models while students are developing their modeling skills. In this case, the use of multiple models refers to students creating and discussing a variety of models of matter (including their peers' models, simulations and physical models). In addition, teachers lead discussions of student models to help students understand both the particle nature of matter and the purpose of creating models.

A foundational piece for the development of this unit was the 1978 Novick and Nussbaum study. This study found that students least internalized aspects of the particle nature of matter that opposed their sensory perception of matter. The concepts they found relevant to developing a particle model of matter are: that matter exists as tiny particles, empty space (the vacuum concept) and intrinsic motion (particle kinetics). These aspects tend to lead students to forming a continuous-particle model, or mixed model in terms of the progression. In particular, students cannot conceive of empty space in ordinary matter, including gases.

Based on the findings of the Novick and Nussbaum (1978) study, the unit focuses on the development of a particle model of matter, including focusing on the following:

- Bulk properties of gases that may make it difficult for students to accept the idea of empty space (addition, subtraction, compression and expansion; air has mass and volume).
- Relationship between energy and speed of motion to get at the intrinsic motion of particles.
- Exposure to more phenomena that are dissonant with their sensory perception of matter that lead to greater accommodation of the particle conception of matter

The unit contains three learning sets. The first learning set (lessons 1-5) aims at helping students understand what matter is (anything that has mass and volume and exists in one of three states) and a consensus model of matter: matter is composed of particles, there is empty space between the particles and the particles are constantly moving. Learning Set 2 (lessons 6-9) helps students understand properties and that properties are a result of the arrangement of atoms in a substance. Learning Set 3 (lessons 10-15) involves students using their models of matter to explain phase changes.

The Smell unit is also designed to be educative for teacher. Educative curriculum materials are designed to promote teacher learning (Davis & Krajcik, 2005). As mentioned earlier, teachers need to understand the practice of modeling, the hybrid models of matter, and student misconceptions of matter. In this vein, the unit includes teacher boxes to help teachers in understanding models (and the particle model in particular), common student ideas (or misconceptions) and ways to help students with these ideas, and subject matter knowledge. In addition, the unit includes descriptions of the types of discussions they should use to help students in understanding the scientific content, phenomena they are experiencing, and about the models the students are creating throughout the unit. For each discussion, there is a purpose for having the discussion, suggested questions and a rational for way these questions help student understanding and what ideas the students should gain from the discussion.

Formative assessments are also an important feature of the curriculum. Formative assessments are activities undertaken by students and teachers that provide information to be used as feedback for modifying instruction (Black & William, 1998). These formative assessments take place throughout the curriculum in the form of activity sheets as well as particular types of IQWST discussions. Some of these formative assessments are also referred to as embedded assessments.

The purpose of embedded assessments is to track student progress throughout the unit. Embedded assessments are valuable because they are indistinguishable from normal instruction, generate feedback for students and teachers, and can be used to detail progress for stakeholders (teachers, parents, and administrators) (Kennedy et al., 2006). In the curriculum, the embedded assessments take the form of students constructing models of phenomena. Student models are defined as their drawing and the explanation. The drawing and explanation portions of the model represent students mental models expressed visually (Justi & Gilbert, 2002).

An early pilot of the unit identified the anchoring activity, the modeling activity of lesson 1 (see Appendix E) in which students explain how smell travels across a room, as an activity that could be repeated throughout the unit to assess students' understanding. The drawing for explaining smell must include a source and a detector. We then identified points along the curriculum in which we thought students were likely to have learned enough to cause them to revise or create new models to explain how smell travels.

As Kennedy et al. (2006) note, it is important to incorporate embedded assessments at “critical junctions where we wanted to make sure students were adequately prepared to learn the next segment of the curriculum” (p.4). Therefore, the smell modeling activity was added to lessons 5 and 15 (see Appendix E) for the purpose of monitoring student learning. As mentioned earlier, lesson 5 is the last lesson of the first learning set where students have learned the basic parts of a particle model. Lesson 15 is the last lesson of the entire unit and occurs after students learn about properties and phase changes on a molecular level.

The modeling activity for lesson 5 includes more scaffolding for the model than those for lessons 1 and 15. However, the main model questions (drawing and explanation) remain the same. These models are referenced according to the activity sheet lesson on which they appear. For example the lesson one model appears on activity sheet 1.1, so it is referenced as AS1.1. These models are used to assess students’ views of matter during instruction.

In sum, the IQWST sixth grade chemistry unit is a research-based unit that has been purposefully designed to help students develop particle views of matter, including the creation and use of their models of matter to explain phenomena. Teacher boxes have been included in the instructional materials to help teachers in understanding the practice of modeling, particle theory and how to address student misconceptions of matter. Embedded assessments of the unit provide the opportunity to track students’ views of matter. Three of these embedded assessments that occur at the beginning, middle and end of the unit are used in this study to track student understanding of matter during instruction. Since the construct map was developed for this unit, it should reflect the learning goals of the unit.

## Chapter Summary

In recent years, science education has focused on the development of learning progressions for the big ideas of science. Progress variables are a means for tracking student understanding of big ideas. Construct maps are a visual description of progress variables that illustrate students' increasingly sophisticated understanding. Thus, a learning progression could be assessed using a single construct map, or be composed of several construct maps.

This study focuses on the development of a research-based progress variable that represents the increasingly complex level of student understanding of the particle nature of matter. This variable could represent a single level of a larger learning progression. The construct map that has been developed for this study incorporates both student conceptions of matter as well as reflecting the goals of the curriculum. The construct map can then be used to track student's progress during instruction.

This chapter synthesized literature related to what research informs us about students' development of a particle model of matter. This information was used both in the development (and revision) of the curriculum as well as the development of its associated construct map.

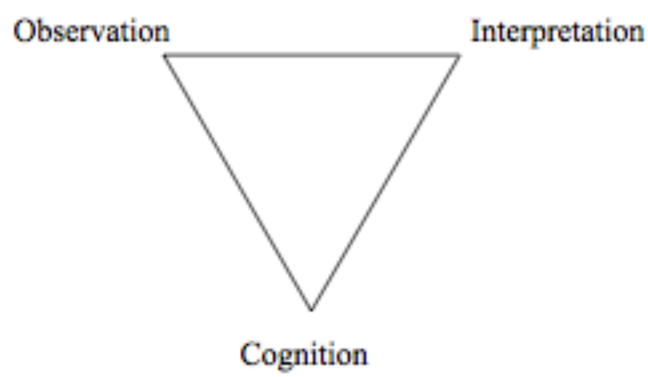
The next chapter describes the framework used to calibrate the PMM progress variable. This discussion further explains the construct map, how the assessment items were developed and associated with the construct map and how the assessment system can be used to iteratively calibrate a progress variable.

# CHAPTER 3

## The BEAR Assessment system

### Introduction

As discussed in Chapter 1, *Knowing What Students Know* (National Research Council, 2001) points out the need for assessments to be based on what we now know about cognition and psychometrics. The Assessment Triangle (see Figure 3.1) is a model of how the three key elements must work together in order to develop effective and efficient assessment tasks. The three corners of the triangle represent the key elements underlying any assessment: cognition, observation, and interpretation. In addition, these three elements must make sense on their own but also in connection with the two other elements (NRC et al., 2001).




(NRC, 2001, p. 44)

**Figure 3.1. The Assessment Triangle**

The cognition corner of the triangle represents the learning theories and beliefs about how students learn in a knowledge domain. The learning theories help to identify the knowledge and skills that are important for gaining competence in a subject domain as well as to help identify tasks to measure this knowledge and skills.

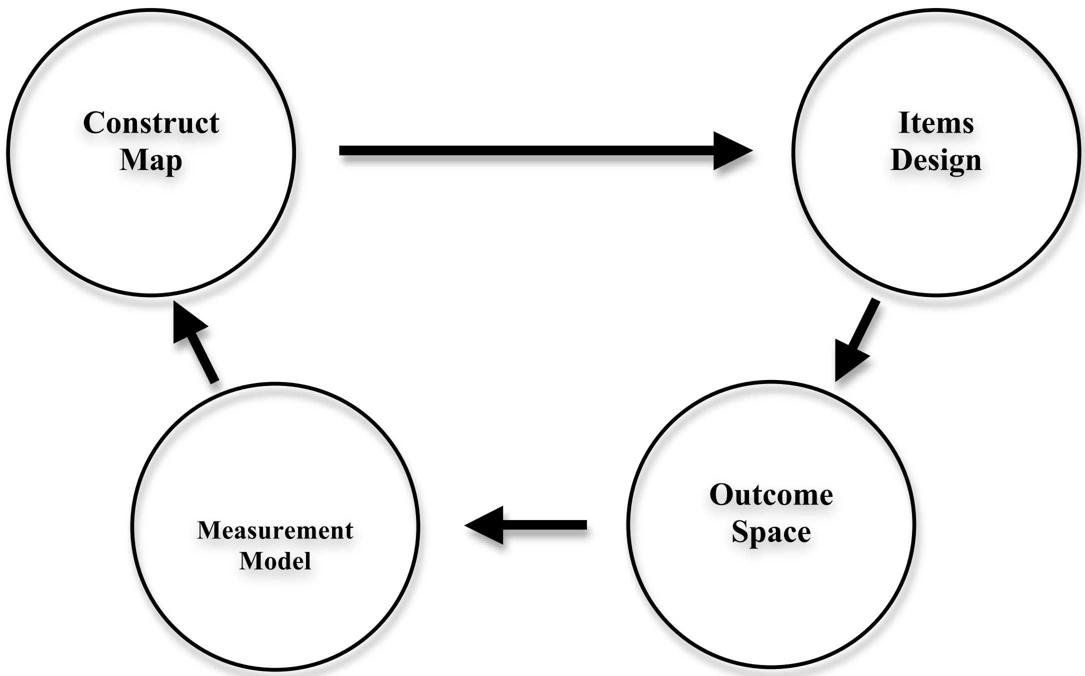
The observation corner of the triangle represents the assessment tasks. These tasks should be designed to elicit responses from students to provide evidence for demonstrating knowledge and skills of a domain (NRC, 2001). These tasks are intimately linked to the cognition corner of the triangle in that the learning theories and beliefs inform the measurer about what tasks will elicit evidence of competence in a domain.

But how do you translate the data that results from the assessment tasks into evidence of knowledge of the domain? The interpretation corner of the triangle represents the methods and tools used to make sense of these observations. The interpretation corner is connected to the cognition corner through identifying measurement models that help to interpret student performance as assessment results. In addition, the ability to reason from or interpret evidence from effective and efficient assessment tasks links the observation and interpretation corners of the triangle.

## The BEAR Assessment System

The Berkeley Evaluation and Assessment Research (BEAR) Assessment System (BAS; Wilson & Sloane, 2000; Wilson, 2005) is guided by four building blocks (see Figure 3.2) for instrument design (Wilson, 2005). Here, instrument refers to the methods used to relate what we observe (manifest/observed) to what we are measuring (latent/unobserved) (Wilson, 2005).

Progress variables are a means by which one can combine research and empirical results to define and develop levels of students understanding. They represent a range of student thinking about a particular knowledge domain, or construct. A construct "can be part of a theoretical model of a person's cognition…their understanding of a certain set of concepts" (Wilson, 2005, p. 6). Constructs are assumed to range from low to high knowledge of a domain, with increasing complexity in between.




(Wilson, 2005, p.17)
**Figure 3.2. The Four Building Blocks.**

**Designing a Construct Map For the Particle Model of Matter**

Progress variables mediate between big ideas and specific concepts and skills being learned and serve as a means for tracking student understanding during instruction (Wilson, 2005). Thus, each unit of instruction contributes to students' progress and necessitates that assessment aligns with one or more construct maps. Alignment of assessment with instruction "allows the creation of a calibrated scale to map the growth of students so teachers can track the progress of individual students as they undergo instruction" (Wilson, 2005, p.195). Therefore, assessments must reflect the variety of instructional practices of the curriculum. In addition, the variables serve as a means for relating curriculum to standards as well as to assessment that are not related to the curriculum. Construct maps are a visual depiction of these variables that divide these levels of complexity into distinguishable levels. The development of construct maps are the first building block of the BEAR Assessment System (BAS).

Thus, we developed a construct map (see Table 3.1) for students' development of an integrated understanding of the particle model of matter. This map serves as the basis for tracking students developing understanding during the IQWST sixth grade chemistry unit. We developed this construct map by an iterative process of considering the logic of the discipline, what was known about how students ideas regarding the particle model (see Chapter 2), and empirical work based on the curricular intervention.

This map illustrates how students' understanding of the particle model builds over time. It also takes into account the instructional sequence. For example, the unit focuses on the particle model before applying the model to explain properties and then phase changes. The "Particle Model" construct map encompasses both the varying starting points students had before the curriculum began and their varying endpoints. This map reflects students' increasingly sophisticated understanding of the particle model as it relates to properties and phase change, starting from the simplest understanding, the "descriptive model," to the most sophisticated understanding, "complete particle model."

## Table 3.1. Particle Model of Matter Construct Map.

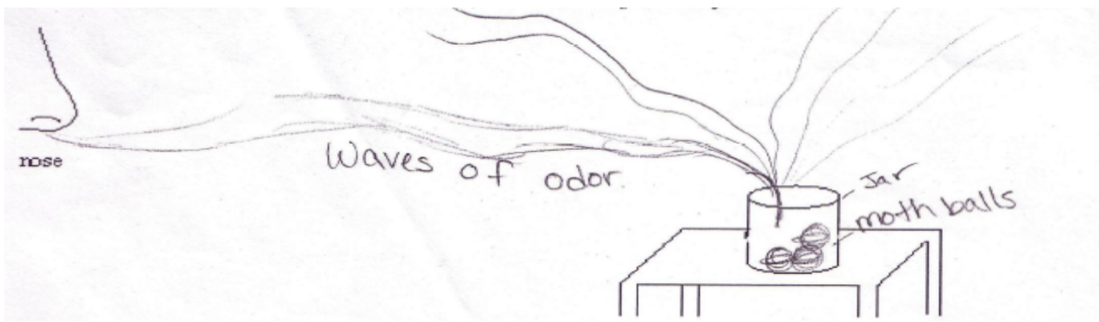
| Category | Description | Example | Progressing to Next Step |
|---|---|---|---|
| Complete Particle Model | Student uses a particle view to describe phenomena. Particles are identified as atoms/molecules of that substance. There is empty space between the particles. The particles are in motion relevant to the particular state they are in. Different substances have different properties because they are made of different atoms OR have different arrangements of same atoms. | Water vapor, liquid water, and ice are all made up of molecules of $H_20$. The molecules in water vapor are far apart and move around freely. In a liquid, they are close together, but move around each other. In a solid, they are close together and vibrate.<br><br>Sugar and water are not the same because they are made up of different molecules. | |
| Basic Particle Model | Students use atoms and molecules to explain phenomena. There is empty space between the particles. Particles are in motion in all states, but may be incorrect, especially for substances that are in the solid or liquid state(s). Different substances have different properties because they are made of different atoms OR have different arrangements of same atoms. | Water vapor is made up of molecules of $H_20$ that are spaced far apart and move freely everywhere. Liquid water is made up of molecules of $H_20$ that are moving, but are closer together than in water vapor. In ice, the molecules are also closer together.<br><br>Sugar and water are not the same because they are made up of different molecules. | Students need to discuss the difference in movement of substances in different phases. For example, a simulation of the same substance as a solid, liquid and a gas should include the same representation for water molecules, but with different spacing and movement, including how movement changes as temperature changes. |
| Incomplete Particle Model | Students use a particle view to describe substances. Particles may be identified as atoms/molecules, but it is not always clear if the atoms/molecules are of the substance. There is empty space between the particles. The student may describe the motion of the particles on the particle level, but it is relative to the other states of matter. Students can describe motion on a macroscopic level. Different substances have different properties because they are made of different atoms. Solids, liquids and gases are made up of particles that have different spacing between them. | Water in its liquid form is made up of particles of $H_20$ that are close together. Water in its solid form (ice) is also made up of particles that are close together. There is empty space between the particles. There are also molecules.<br><br>Sugar and water are not the same because they are made up of different particles. | The idea that water is made up of the same atoms/molecules no matter the state should help students to realize that a substance's atoms/molecules do not change. In addition, creating models should help students to further understand this idea. For example, a model of ice, water and water vapor should include the same representation for water molecules, but with different spacing and movement. |
| Mixed Model | Students use both particle and descriptive views when explaining everyday phenomena. When asked to describe what makes up a substance, students at this level often describe particles within a continuous medium. They do not understand that different substances have different properties **because** they are made of different atoms. Students describe solids, liquids, and gases as made up of smaller pieces of that same substance, which come together to form a whole. | Water is made up of particles of $H_20$. The particles of $H_20$ exist within the liquid water. Thus, in between the particles is liquid water. | To move to the next level student needs to develop an understanding that a substance is made up of particles. Moreover, they need to understand that there are empty spaces between the particles. |
| Descriptive Model | Students at this level see objects as being a continuous medium. When asked to describe what makes up a common substance, they are described exactly as they appear. Thus, substances always have the same properties because the student has no concept that the substance may have a structure made up of smaller pieces. | Water is a clear, colorless liquid. Ice is a "clear" solid. They have different structures and are described differently. Therefore, they are not the same substances. | The ideas that objects are made up of parts could be a useful piece of knowledge to help students realize that a pieces of a substance that looks continuous, can be broken down into smaller pieces. Student needs to realize that a substance is changeable (it can change phases), or in other cases, may be broken into smaller pieces. |

The sixth grade chemistry unit has three embedded assessments, which have students create models to explain how a smell can travel from its source to their nose. These models also show how a single student's view of matter changes over time (see Figure 3.1). For example, this student started with a "descriptive model." At this level of understanding, a phenomenon is depicted exactly as it appears (See Figure 3.1a). The student's model consists of a drawing of the odor with no particle ideas and writes, "The odor is coming out of the source." By lesson 5 (see Figure 3.1b), the student now represents the odor as ammonia molecules that are moving in all directions and describes what is happening as follows: " Molecules in the liquid come off the surface of the liquid and become a gas. They move around and change direction when they come in contact with another object." This is a "Basic Particle Model", because the student identifies the molecules as ammonia molecules and includes the random movement of the molecules both in the drawing and written portion of the model. By the end of the unit (see Figure 3.1c), the student has a "Complete Particle Model". The student represents molecules of air and ammonia and includes a more sophisticated representation of random movement through the use of arrows and writes:

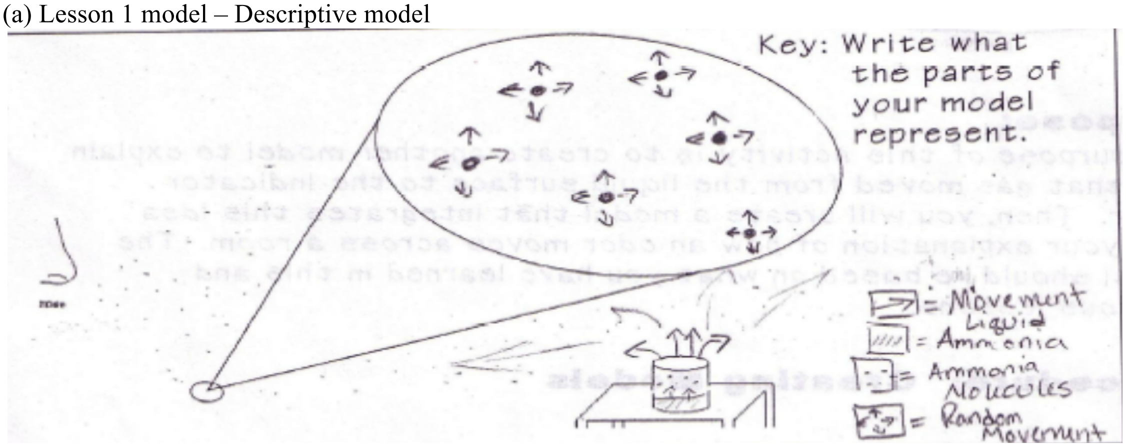
> My model shows that molecules go into the gas phase from the liquid and move outwards through the air in straight lines until they bump into something. When they bump something, they go in another direction in a straight line. The speed of the movement will change according to temperature and if there is an air current.

This example shows one student's path to a particle view of matter. The student started with a "descriptive model" of matter to a "complete particle model" by the end of the unit. In this case, the student's particle view of matter was represented in both the drawing and written portions of his model.

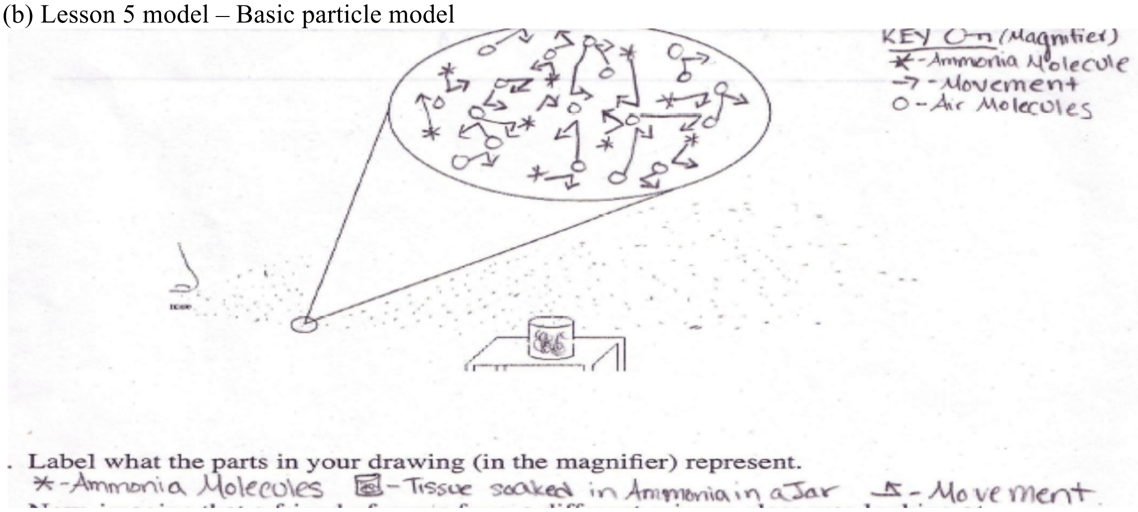
The development of the Particle Model of Matter construct map is the first block in the BAS. The ultimate goal of the BAS is to determine whether items are good measures of the construct, for this study the particle nature of matter. Ideally, items would be developed based on the levels of the construct map. In this case, the items were developed prior to the construct map before linking them to the construct map.




(a) Lesson 1 model – Descriptive model




(b) Lesson 5 model – Basic particle model




(c) Lesson 15 model: Complete Particle Model

**Figure 3.3. Changes in one student's model (drawing portion) over time.**

## Item Design

The second building block of the BAS is items design. The *Investigating and Questioning our World through Science and Technology (IQWST)* project (Krajcik, Reiser, Fortus and Sutherland, 2009) takes the approach of building student's ideas over time. In the sixth grade chemistry unit, students develop and use the particle model to explain phenomena, such as states of matter, phase changes, and properties (Merritt, Krajcik, & Shwartz, 2008). Thus, students experience various phenomena throughout this eight-week unit to help them to gain knowledge and understanding of the different aspects of the particle nature of matter. What follows is a discussion of the item design process, which included identifying learning performances and item development.

### Identifying and unpacking standards

For the development of this unit, we identified three standards (see Table 3.2) from the Benchmarks for Scientific Literacy (AAAS, 1993) and National Science Education Standards (NRC, 1996). The identification of a small number of standards sets the IQWST curricula apart because of our focus on depth instead of breadth.

Once the standards were identified, we underwent a process of unpacking what it means to teach them. Unpacking, in this instance, means we carefully read through the standard, identifying concepts within them which are important, what knowledge students may bring to these ideas, what misconceptions students have as well as to what depth these concepts should be covered for students, in this case, in sixth grade (Krajcik et al., 2008).

For example, the first standard (AAAS 4D/M1) begins with the idea: All matter is made up of atoms. We determined that this idea was composed of two concepts: 1) that matter is made up of particles (atoms) and 2) that these particles are atoms. Then, we determined that students need to understand what matter is – anything that has mass and takes up space. From research, we were able to identify that students would have difficulty in differentiating weight and mass as well as difficulty in identifying air and other gases as matter (Driver et al, 1985, Driver et al, 1994). Additionally, we looked at what prior knowledge students should have of matter based on the preceding national standards. In some instances, as we unpacked the standards, we also identified what concepts students would not be expected to learn at this time. For example, students are not expected to understand that a single atom has the *chemical* properties of that element, but it takes several atoms to give the element its *physical* properties.

**Table 3.2. National Standards (AAAS, 1993; NRC, 1996).**

| |
|---|
| **AAAS 4D/M1:** All matter is made up of atoms, which are far too small to see directly through a microscope. The atoms of any element are alike but are different from atoms of other elements. Atoms may stick together in well-defined molecules or may be packed together in large arrays. Different arrangements of atoms into groups compose all substances. |
| **AAAS 4D/M3:** Atoms and molecules are perpetually in motion. In solids, the atoms are closely locked in position and can only vibrate. In liquids, the atoms or molecules have higher energy, are more loosely connected, and can slide past one another; some molecules may get enough energy to escape into a gas. In gases, the atoms or molecules have still more energy and are free of one another except during occasional collisions. Increased temperature means greater average energy of motion, so most substances expand when heated. |
| **NRC B5-8: 1A: A substance has characteristic properties, such as density, a boiling point, and solubility, all of which are independent of the amount of the substance** |

This process of clarifying and elaborating the standards helped to ascertain what it means to teach sixth grade students the particle nature of matter and how the particle model is used to describe states of matter, as well as explain phase changes and properties. Unpacking process also helped to identify what ideas needed further support for students (Krajcik, McNeill and Reiser, 2008). For instance, helping students understand that matter is anything that has mass and volume is a fundamental concept for helping students to understand both states of matter as well as developing a particle view of matter. Students often conflate the terms mass and volume. Therefore, the decision was made to include activities for students to measure mass and volume as well as to include discussions of matter and volume on both macroscopic and microscopic levels when discussing states of matter. From this work, we were able to develop a unit that contains three learning sets and corresponding assessment items.

**Development of Items**

Assessment items for the unit were developed at the same time as the unit was being developed. One source of item development was the unit's learning performances. The standards that serve as the unit learning goals (Table 3.3) were used to construct learning performances. A learning performance results from combining the content standard with an inquiry standard. These learning performances clearly specify what students are expected to be able to do with the knowledge described in the benchmark. Moreover, they "serve as the learning goals that guide development of learning activities and assessments" (Krajcik et al., 2008, p.7).

**Table 3.3. Example of a Learning Performance.**

| Content Standard | Inquiry Standard | Learning Performance |
|---|---|---|
| **AAAS 4D/M3:** Atoms and molecules are perpetually in motion. In solids, the atoms are closely locked in position and can only vibrate. In liquids, the atoms or molecules have higher energy, are more loosely connected, and can slide past one another; some molecules may get enough energy to escape into a gas. | Develop…models using evidence. (NRC, 1996, A: 1/4, 5-8)<br><br>Models are often used to think about processes that happen…too quickly, or on too small a scale to observe directly… (AAAS, 1993, 11B: 1, 6-8) | Using the particle model, students will explain phase change from a solid to a liquid. |

Thus, in this unit students use the particle model of matter to explain phenomena related to states of matter, phase changes, and properties. For example, the particle model can be used to explain a property like boiling point. The boiling point of a substance occurs at a fixed temperature and involves the rapid evaporation of anywhere in a bulk liquid. During heating, particles gain energy and move faster. The energy of these molecules is enough to overcome the attractive forces of the other liquid molecules so that it goes from the liquid to the gas phase.

In some instances, the cognitive tasks that students might perform are characterized through the revised Bloom's taxonomy (Anderson & Drathwohl, 2001). Bloom's taxonomy is a classification of educational objectives and consists of six categories. These revised categories (Remember, Understand, Apply, Analyze, Evaluate, and Create) range from simple to complex and are a framework for developing items focused on what students should be able to do with their knowledge of the domain. In the example in Table 3.3, students are expected to be able to explain a phase changes from solid to liquid on a molecular level. In this example, an item could be developed in which students explain phase changes through the construction of their own models or model(s) provided to them. When we developed an item to assess this understanding, we decided to create an item where models representing a phase change from solid to liquid were presented to students (see Figure 3.4). In this item, students are expected to be able to distinguish the different phases of matter based on the spatial relationship between the different particles. In the solid models, the particles are close together. In the liquid models, the particles are close together, but loosely connected. In the gas models, the particles are far apart. Therefore, students should be identify that the correct answer is B, where the particles are close together in the solid state and together, but loosely connected in the liquid state.

**4. In the following models, each circle represents a wax molecule. Which model best represents what happens when a solid wax melts into liquid wax?**

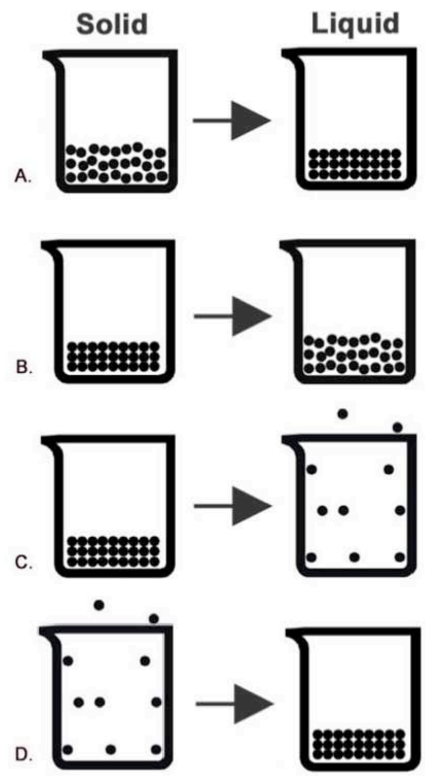



**Figure 3.4. Assessment Item developed from learning performance.**

The second source of item development was through the application of Bloom's Taxonomy (Anderson et al., 2001). We utilized this aspect of the taxonomy in order to develop a broad range of questions that could be termed high, medium and low level difficulty. In other words, we wanted to develop items that spanned from requiring students to recall knowledge that they had learned about the particle nature of matter to having them apply their knowledge. For example, we wanted to assess whether students understand what is happening on a molecular level as a substance changes phases. This question was developed for the learning performance: Students will explain that the particles are the same, but behave differently in the three phases. Figure 3.5 is the item that resulted. In this item, students need to recall their knowledge of two of the phases, liquids and solids, on the molecular level and compare the two. This item implies that the molecules are the same, but requires students to recall the difference in movement between the two phases.

**Figure 3.5. Assessment Item developed from learning goal.**

**When a substance changes from a liquid to a solid, which of the following is true?**

A. The molecules get colder.

B. The molecules of the solid move faster.

C. The molecules of the substance change from soft to hard.

D. The molecules move more slowly.

We then used the Project 2061 Item Analysis Procedure (DeBoer, 2007) to analyze whether the items aligned with the learning goals. This included determining whether the learning goal was: 1) needed to make a satisfactory response? (necessity) and 2) enough by itself, or do students need additional knowledge to solve the item? (sufficiency). In the second example item (Figure 3.5), the learning goal is both necessary and sufficient because students need to have a molecular understanding of a liquid to solid phase change. The result of the item development process led to the creation of identical pre- and posttests composed of multiple-choice and written response items. The existing pre/posttest consists of 15 multiple-choice items and three written response items (see Appendix A).

## Outcome Space

The third building block of the BAS is the outcome space. The outcome space is where the measurer decides how to make inferences and how aspects of the responses are to be categorized and scored. The progress variable identifies the qualitatively different levels of knowledge of a domain, including what students know and can do with said knowledge. The outcome space assists in identifying student responses that correspond to a specific level of the construct map, emphasizing the content of the responses that reveal a particular level.

For each item, I had to determine what content knowledge needed to respond to the item. Moreover, I read each item and its responses and compared the content knowledge required to answer that question with the content knowledge expressed in each level of the construct map to determine whether the item focused on one or more levels of the construct map. Thus, in this process I identified which items were dichotomous and which were polytomous. Dichotomous items only focus on one level of the construct map. For example, the item in Figures 3.4 was identified as a dichotomous item that aligned with the "Incomplete" level of the construct map. At the "Incomplete" level (see Table 3.1), students are able to recognize that solids, liquids and gases have different spacing between them. The answer choices for the item in Figure 3.4 are models of solids and liquids with incorrect answer choices A, C and D including incorrect representations for solids and/or liquids. As stated earlier, students should be able to recognize the correct answer as B based on the spacing of the molecules in the answer choices.

Polytomous items were identified as such because their item responses cover more than one level of the construct map (Wilson, 2005). Item 1 (see Figure 3.6) is a polytomous item. The responses for this item align with three levels of the construct map: "Descriptive", "Mixed" and "Incomplete". Answer choices B and C were aligned with the “Descriptive” level of the construct map because these responses correspond with students visualizing a gas as being a continuous medium. Answer choices B and C were designed to mirror the drawing portions of models students have created in previous studies, which were previously coded to represent a continuous view of matter. At the “Mixed” level, students have both a continuous and particle view of matter. Answer choice D represents a “Mixed” level response in that it contains both waves and particles. At the “Incomplete” level, students have a particle view of matter, with empty space between the particles. This view of matter corresponds with answer choice A. For each item, the answer choices were aligned with the levels of the construct map to determine whether an item was dichotomous or polytomous. All written response items were determined to be polytomous items because students provide responses at all levels of the construct map.

1. **Below are four possible models of a gas.** Which model would a scientist use to show how water vapor condenses to a liquid?

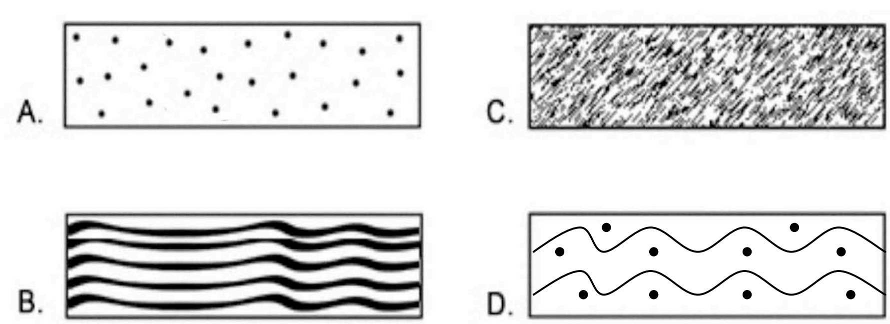



**Figure 3.6. Polytomous multiple-choice item.**

Scoring guides were developed for written response items. These guides take into account actual student responses to the item as a means for mapping them to the different levels of the construct map. When measuring a small and constrained construct, such as this one, the scoring guide will resemble the construct map. Table 3.4 is the scoring guide for the written response portion of an item in which students are creating models to explain to a friend how water vapor, water and ice are all the same substance. For example, the "Basic Particle Model" level of the construct map details students having difficulty in describing the motion of particles on a molecular level. This is reflected in the scoring guide by including how students are unable to accurately describe movement of the particles in all three states.

**Table 3.4. Scoring guide for Written Portion of Model.**

| Code | Part B |
|---|---|
| **0** | No response<br>Descriptive – describes water in each state exactly as it appears, defines what a phase change is. |
| **1** | Mixed Model<br>Although the student may mention atoms or molecules, student describes how a phase change occurs on a macro level. |
| **2** | Incomplete Particle Model<br>Although student may identify particles as molecules, they do not fully understand what an atom or molecule is. Student is able to distinguish spacing between molecules in each state OR difference in movement in each state. |
| **3** | Basic Particle Model<br>Student identifies particles as water molecules. Student is able to describe spacing between molecules in each state. *Student is unable to distinguish movement during the different phases. For example, can describe movement of a liquid and a gas, but a solid does not move.* |
| **4** | Complete Particle Model<br>Student identifies particles as water molecules. Student is able to describe spacing between molecules in each state. Student is able to describe movement during the different phases. |

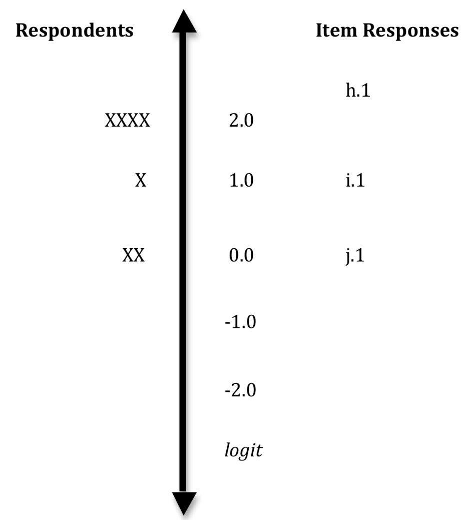
## Measurement Model

The final building block of the BAS system is the measurement model. The model is used to relate the scored data back to the construct map. Thus, it can be seen as a technical version of the construct map. This technical version of the construct map is developed through Rasch modeling, which is centered around analysis at the item level (Wilson, 2005; Embretson & Reise, 2000; Hambleton & Jones, 1993). Rasch modeling relates students' abilities to item difficulties, by placing items and persons on the same scale. The model is visualized through the use of Wright maps, which are an aggregation of all students' proficiency levels in relation to all the item difficulties. These abilities are measured on a logit scale. Relative locations are important on Wright maps because the probability of success is with respect to the items estimates. Differences between items or persons have the same meaning on a logit scale (Embretson & Reise, 2000). Table 3.5 displays how logit differences relate to probabilities.

**Table 3.5. Relationship between logit differences and probabilities.**

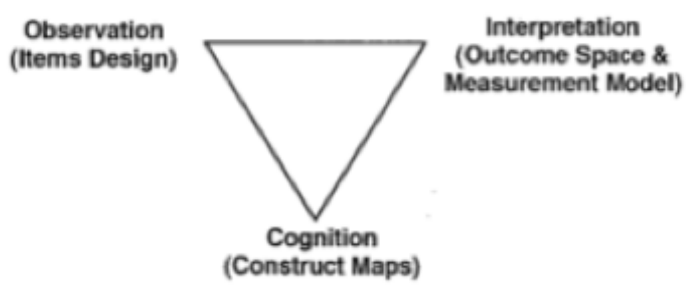
| Logit difference | Probability |
|---|---|
| -3.0 | 0.05 |
| -2.0 | 0.12 |
| -1.0 | 0.27 |
| 0.0 | 0.50 |
| 1.0 | 0.73 |
| 2.0 | 0.88 |
| 3.0 | 0.95 |

Figure 3.7 is a generic Wright map that can help to understand logit differences. The letter X represents respondents and each X is one respondent. When a respondent's location and the item difficulty are at the same location, there is a 50-50 chance of them getting the item correct. For example, the two respondents located at 0.0 logits have a 50% chance of getting item j.1 correct. When the respondent's location is above an item, they have a greater chance of answering correctly. Thus, the respondents located at 2.0 logits have an 88% chance of getting item j.1 correct and a 73% chance of getting item i.1 correct (l.0 logit difference). Respondents have lower chance of getting the item correct when the item is above the respondent's location. Respondents located at 0.0 have a 27% change of getting item i.1 correct (-1.0 logit difference).




(Wilson, 2005, p. 96)
**Figure 3.7. Generic Wright map.**

As stated earlier, the NRC Assessment Triangle is a model for developing good assessments of learning. But how does the BAS relate to the assessment triangle? Figure 3.7 illustrates how the BAS relates to the NRC assessment triangle. The "cognition" corner is represented by the construct map, which is built upon the theories of learning within a particular domain. The "observation" corner is realized through the previously discussed items design. The "interpretation" corner corresponds to both the development of the outcome space and the measurement model.




(Kennedy, Brown, Draney, & Wilson, 2006, p. 33)
**Figure 3.7. Relationship of BAS to Assessment Triangle.**

## Chapter Summary

This chapter described the Bear Assessment System (BAS). The BAS is a cyclical approach for the development of assessments. Moreover, its cyclical nature offers the opportunity to continuously revise the construct map, the items used to measure student understanding, and the outcome space. The purpose of this study is to track students' understanding of the particle nature of matter during instruction. The overall question informing this study is: How does middle school students' understanding of the particle nature of matter change during enactment of a model-based unit? To answer this question, two studies were conducted.

The first study is a calibration study aimed at answering the question: Are the assessments valid measures of students' knowledge of the particle model of matter? To determine whether the construct map truly represents student understanding, the items were aligned with the levels of the construct map. The Wright map can then being the process of assessing the validity of the items and construct map. The results of this study were then used to answer the overall question in the tracking study, as well as two sub-questions:

- What is students' prior knowledge of the particle nature of matter and how does their prior knowledge relate to the progress students make during instruction?
- What is the relationship between students' drawings and explanations of phenomena?

The next two chapters present the results from both studies.

# CHAPTER 4

## Calibration Study

## Introduction

The previous chapter describes the assessment system that I used to analyze student data. The purpose of the calibration study is to examine the question: **Do the assessment items of the particle model of matter progress variable validly measure students' knowledge of the particle model of matter?** In particular, this study allows me to ascertain how well the items relate to the Particle Model of Matter (PMM) progress variable. A map of the PMM progress variables, including the performance levels, was discussed in Chapter 3 (see Table 3.1).

## Calibration Study

### Study Setting & Participants

The calibration study involved 89 7th graders from Detroit, Michigan. These students were completing the 7th grade chemistry unit that builds off of the ideas that students learned in the sixth grade unit. I chose this sample of students because 18 students had not experienced the sixth grade unit and thus, hoped their lack of experience with the sixth grade unit would result in a wider range of responses.

Sixth grade students were not used for the calibration study because 1) they were just starting the chemistry unit for which this instrument is being evaluated and 2) would not be able to provide a wide range of responses to the items. This sixth/seventh grade teacher has taught both sixth and 7th grade units for more than three years and has more than five years teaching experience.

## Data Sources

There are two data sources for the calibration study: the assessment and exit interviews. Table 4.1 lists the data sources used in this study as well as their purposes.

**Table 4.1. Calibration Study Data Sources.**

| Data Source | Description | Purpose |
|---|---|---|
| Assessment | Multiple-choice and two open-ended assessment items | To calibrate items and evaluate the particle model of matter progress variable. |
| Exit Interviews | Interviews about the items on Form A & B | To assess item validity, including student difficulties with items |

### Assessment Items

To determine whether the items measured students' understanding of the particle nature of matter, students responded to a 25-item test. This test was composed of 13 items from the existing pre/posttest (see Appendix A) as well as twelve new items. Three items (two multiple-choice items and one written response item) from the pre/posttest were not used in this part of the study to insure the items were completed within the one-day time frame. The new items were created to insure that there were enough items to both evaluate the progress variable as well as to insure that the item responses covered the entire construct map.

The development of these new items (Items 6-11, 15-21 and 23-24) was identical to that of the original items (see Chapter 3), but with the additional step of writing the items to insure that they also aligned with a particular level of the construct map. There are 20 learning performances for the entire unit, but that does not mean there is a single item on the pre/posttest for each of these learning performances. In this case, I identified learning performances that aligned with the construct map, but did not have related assessment items developed for the pre/posttest. For example, there are four existing multiple-choice items that explicitly involve phase changes, a major part of the progress variable. I created one new open-ended item that explicitly asked students to create a model to explain a phase change (see item 24, Appendix B).

Alignment of the items with the different levels of the construct map represents hypotheses about how students are expected to respond to items (see Chapter 3). Where the responses to these items align with the Wright map as well as further analysis, help to identify whether the alignments align in the same pattern as my hypotheses.

Although two classes completed the assessment within the original one-day time frame, one class completed the test over a two-day window based on other school functions occurring due to the holiday season. Although not ideal, a sufficient subset of students (n = 4) was given exit interviews after administration of the exam.

**Exit Interviews**

A commonly used method for gathering evidence of validity is the exit interview (see section on Validity). The exit interview is completed after students complete an item or after they complete the assessment (Wilson, 2005). For this study, students were interviewed after they completed the assessment. The original interview was designed to last over a longer period of time. However, programs related to the holiday season resulted in my having to modify the interview on the spot to not interrupt other teacher schedules. During the interview, students talked about their thinking as they completed the item.

Four students (2 male, 2 female), chosen by their teacher, participated in the interviews. I interviewed students with a range of abilities, as determined by the teacher, and different genders so that I could get a range of responses in relation to the construct map. One male student had no prior experience with the sixth grade unit. During the interview, students talked about their thinking as they completed particular items as well as how they felt about responding to the items in general. In particular, I asked them "Were there any questions that were difficult or hard to understand? Why was the item difficult?". Their responses were audiotaped.

## Data Analysis

### The Outcome Space and Scoring Guides

As was discussed in Chapter 3, the outcome space provides evidence of students' knowledge related to a particular level of the construct map based on the content of their item responses. Thus, scoring guides were developed for written response items to take into account both student responses while mapping them to the different levels of the construct map. For each of the written response items for both forms, this involved the development of scoring guides for the drawing and explanation portion of the items (see Appendix B, see items 13 and 14 for both forms).

The scoring guides contain more details and one additional level, in comparison to the progress variable. This is because the scoring guides must be able to handle all possible student responses, including incomplete, incorrect and unusual responses. The PMM scoring guides have three columns. The first column is the code to designate the particular student response. The second column describes what knowledge the students at that level include in their responses. The third column represents examples of student responses at that level.

For the PMM scoring guides, there is a blank category for uncodable or blank responses. In addition, some levels incorporate common incorrect or incomplete responses. For example, in the scoring guide for Form A, item 13 (see Table 4.2), the descriptive level includes students providing a completely incorrect answer. In addition, common misconceptions for a particular level are also accounted for. For example, the code 3, “Beginning Particle Model,” accounts for students including an incorrect mechanism in their response. This is done to account for the actual student responses. For example, many students at this level will respond that in warm temperatures, particles move faster, but in cold temperatures particles will freeze. Although the student seems to understand what is happening on a molecular level when temperatures are warm, they are attributing macrolevel attributes (freezing) to their microlevel explanation (particles freezing in cold temperatures).

By defining the outcome space, the potential item responses are qualitatively linked to a performance level that corresponds to a particular level on the construct map. The analysis of the data will provide the specific cut scores that correspond to the particular levels of the progress variable.

**Table 4.2. Scoring guide for Item 22.**

| Code | Part C | |
|---|---|---|
| **0** | No response<br>Descriptive – describes model OR gives completely incorrect explanation (i.e. an external source creates movement) OR uses prior experience to explain what is happening. | I chose my answer in A because the temperature of a room doesn't affect how fast or slow an odor moves through a room. |
| **1** | Mixed Model<br>Student tries to explain what is happening, but uses the incorrect mechanism OR simply repeats correct choice "Warmer room air moves faster" | Molecules move faster at colder temperatures. The heat slowes the molecules down. |
| **2** | Beginning Particle Model<br>Student may identify particles as molecules, but focuses partially on a macro level explanation (odor/air). Student explanation focuses only on one gas (odor/air/gas/atoms/molecules) moving faster. Although mostly correct, student answer may include incorrect mechanism. | The warmer the temperature the faster atoms move. The colder the temperature the slower atoms move. |
| **3** | Basic Particle Model<br>Student is able to identify that (air and/or odor) molecules travel faster in a warm room (and/or slower in a cold room) in correct relation to temperature/energy | The smell reaches the door faster at 85$^{o}$F. The molecules are warmer and move faster in room A. In room B, the molecules are colder and moves slower. Therefore, the smell reaches the door faster at 85oF, because the warmer the room the faster the molecules and atoms move. |
| **4** | Complete Particle Model<br>Student is able to correctly explain on a microscopic level that in a warmer room, air moves faster because of higher energy, resulting in odor spreading/traveling thru room faster (and/or vice versa for cold room) | I chose the smell will reach the faster in 85$^{o}$C because there will be more heat energy in the room than in 50$^{o}$F. The heat energy will cause the molecules to speed up and reach the other side faster. |

**Measurement Model**

As mentioned earlier, BAS utilizes Rasch-based modeling because of its ability to use the same scale to generate estimates of person abilities and item difficulties. For this study, I used a one-dimensional Rasch-based model. Using a one-dimensional model means that I am not looking at a students' model of matter separate from their content knowledge, but that the particle model of matter is a single dimension of their overall understanding of the particle nature of matter. The Construct Map (Kennedy, Wilson, & Draney, 2009) software was used for calibrating items using a partial credit model

(Masters, 1982), which is a polytomous extension of the Rasch Model (Rasch, 1960). The partial credit model assumes that there are ordered steps for both person proficiency and item difficulty. For items that cover more than one level of the construct map, each level of response is a step. With a partial credit model, students with higher proficiencies should align with more difficult steps. I utilized this model because both test forms include dichotomous and polytomous items. In other words, the tests include items that measure only one level of the progress variable (dichotomous) and items that measure multiple levels of the construct map (polytomous). Furthermore, each step has its own associated difficulty.

There are three parameter estimation methods available using the Construct map software, of which two could be used for modeling the data: maximum likelihood estimation (MLE) and expected a posteriori (EAP). MLE is best for data that has many respondents and few items, while EAP is best for use with many items (>50) (Embretson & Reise, 2000). For both studies, MLE is utilized because of the number of items (32) in the study.

## Findings

Several analyses were conducted to determine whether the items validly measure student understanding of the particle nature of matter. The sections that follow discuss the evaluation of the model, including reliability and validity. To begin, I discuss the evaluation of the model.

### Evaluation of the model

The model is evaluated through examining the Wright map, the item fit and respondent fit. The Wright map is utilized to look at two different aspects of the data –

the Thurstonian thresholds and the step difficulties. The Thurstonian thresholds represent the point on the variable at which the probability of being observed in that level or above equals that of being observed in the levels below (Wilson, 2004). Thus, a respondent at zero logits has a 50% probability of getting an item at that level correct. For example, the respondents at zero logits in Figure 4.1 have a 50% chance of responding to h.1 and i.1.

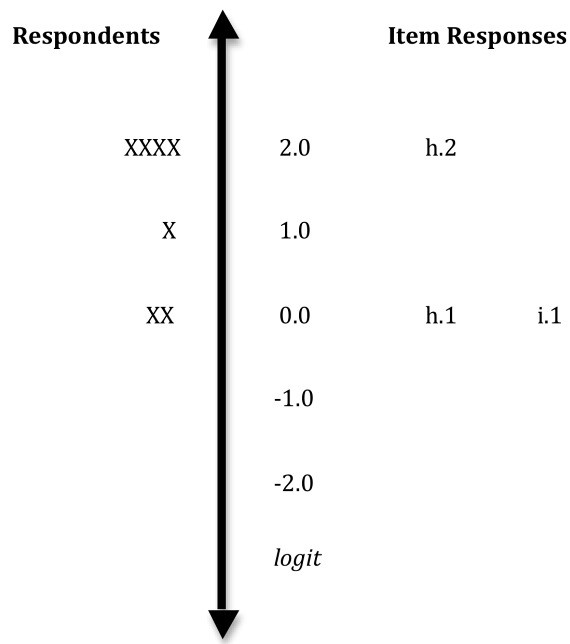



**Figure 4.1. Generic Wright map of polytomous items, where X = 1 respondent.**

Because a partial credit model was used, the expectation is that the order of the thresholds related to the item is also ordered. For example, if an item is designed to garner responses from students related to three levels of the construct map, these responses would be scored to represent three levels (0, 1 and 2). Thus, the item has two steps, or the transition from one level to the next. If an item has two steps h.1 and h.2, then h.1 should appear lower on the Wright map than h.2 (see Figure 4.1). This also means that step h.1 has a lower difficulty than step h.2. If the order of the thresholds were not correct, this would indicate that the entire construct map would need to be reconsidered.

Step difficulties represent the difficulty of achieving a score at that score level rather than at the preceding level (Wilson, 2004). Unlike Thurstonian thresholds, it is not always expected that the more difficult step appears higher on the construct map because they estimate the difficulty in transitioning form one level to the next. The step difficulties indicate the difficulty to transition from one level to the next. This means that as students progress from a step 0 to 1, this step could be easier than the step from 1 to 2 or vice versa. However, step difficulties are meaningless for dichotomous items as 0 and 1 represent a student getting the item correct or incorrect. For polytomous items, the step difficulties represent a degree of correctness and "some steps may be relatively easier or more difficult than others" (Embretson & Reise, 2000, p. 106). Therefore, only polytomous items were examined for step difficulties. Items that have disordered steps would need to be further investigated to determine if there is a pattern that emerges for why their steps are disordered. In addition, this could indicate that the levels of the construct map are not distinct and would need to be revised.

The investigation of fit is "the gathering of evidence that the mathematical models being used are appropriate" (Wilson, 2004, p. 127). Fit is investigated in terms of both items and respondents (discussed later). In addition, I determined how the levels of the progress variable correspond with the levels of the Wright map.

Determining the performance levels was accomplished using the Thurstonian thresholds of the item response categories. Thurstonian thresholds indicate the ability required to achieve a response at that level or above 50% of the time. The determination of these levels is discussed further later.

**Wright Map**

The Wright map of item thresholds for the particle model construct (see Figure 4.2) resulted in an almost normal person distribution around 1.00 logit. On the left side of the map are students represented by the letter X. Each X represents one student.

```
-----------------------------------------------------------------------------
                 Wright Map (MLE)     Variable: Particle Model of Matter
                               IRT Categories
       Map of person estimates and response model parameter estimates
=============================================================================

            students              Thurstonian Thresholds (Recoded)
-----------------------------------------------------------------------------
6  |                        |
   |                        |
   |                        |
   |                        |17.2
5  |                        |
   |                        |
   |                        |
   |                        |
4  |                        |                  22c.4
   |                        |                              24b.4       25a.4
   |                       X|                        23b.3 24a.4             25b.4
   |                        |            22b.3
3  |                        |
   |                        |                                          25a.3 25b.3
   |                       X|
   |                        |                  22c.3 23a.3
2  |                 XXXXXXX|                              24a.3 24b.3
   |                      XX|
   |               XXXXXXXXX|                                                25b.2
   |              XXXXXXXXXX|1.2   2.1               23a.2
1  |    ----XXXXXXXXXXXXXXXX|7.1   11.1  22b.2 22c.2 23b.2
   |          XXXXXXXXXXXXXX|19.1
   |          XXXXXXXXXXXXXX|22c.1
0  |                   XXXXX|6.1   17.1  22b.1             24a.2 24b.2
   |                    XXXX|16.1                    23b.1
   |                     XXX|10.1  12.1  20.1                    24b.1
   |                     XXX|3.1                     23a.1
-1 |                        |13.1  18.1                                25a.2
   |                        |1.1   15.1                                      25b.1
   |                        |5.1                             24a.1
   |                        |8.1
-2 |                        |9.1   21.1
   |                        |14.1                                      25a.1
   |                        |4.1
   |                        |
-3 |                        |
   ==========================================================================
          Each X represents 1 student, each row is 0.255 logits
```

**Figure 4.2. Wright map for the Particle Model of Matter construct.**

On the right side of the map are the items. Zero logits represents average ability and difficulty; thus, students located at zero logits have a 50% chance of getting items at that level correct and greater than 50% for items that are below them. Above zero logits, are the more difficult items and students with higher ability. Below zero logits are items that are less difficult and students with lower ability. In this case, most students will have a greater than 50% chance of getting most of the items correct because most students have proficiencies above zero logits. This was expected because most of these students had already experienced the unit, and were completing the 7$^{th}$ grade unit. It should be noted that students' having prior experience with unit could be an issue in terms of having a wide enough range of student abilities.

The Wright map is a visual means of interpreting the construct map (Wilson, 2005). As mentioned before, I am using a partial credit model, which means that on the Wright map each threshold of an item should be ordered. The results show that the steps are order as expected. For example, item 1 has two steps: 1.1 and 1.2 (see Appendix B It covers three levels of the construct map: 1) a descriptive response is the zero step, 2) mixed response is step 1 and 3) incomplete response is step 2. The importance of distance for a Wright map is demonstrated because the second step (1.2) is more difficult than the first step (1.1) because it is located at a higher logit (see Figure 4.1). The same desired pattern is reflected in the written response items (Items 22-25). This is also reflected in that the lowest level items are actually scoring as such (see Evidence Based on Internal Structure section). Moreover, the most difficult items for students were the open-ended items (22-25) on both forms.

As aforementioned, I examined the step difficulties. There are ten polytomous items, whose step difficulties are displayed in Figure 4.3. Figure 4.3 displays the steps of each item (22c.1, 22c.2, 22C.3, 22C.4), as well as the average for all steps for a particular item (22C). The four items that have disordered steps are indicated in bold on Figure 4.3.

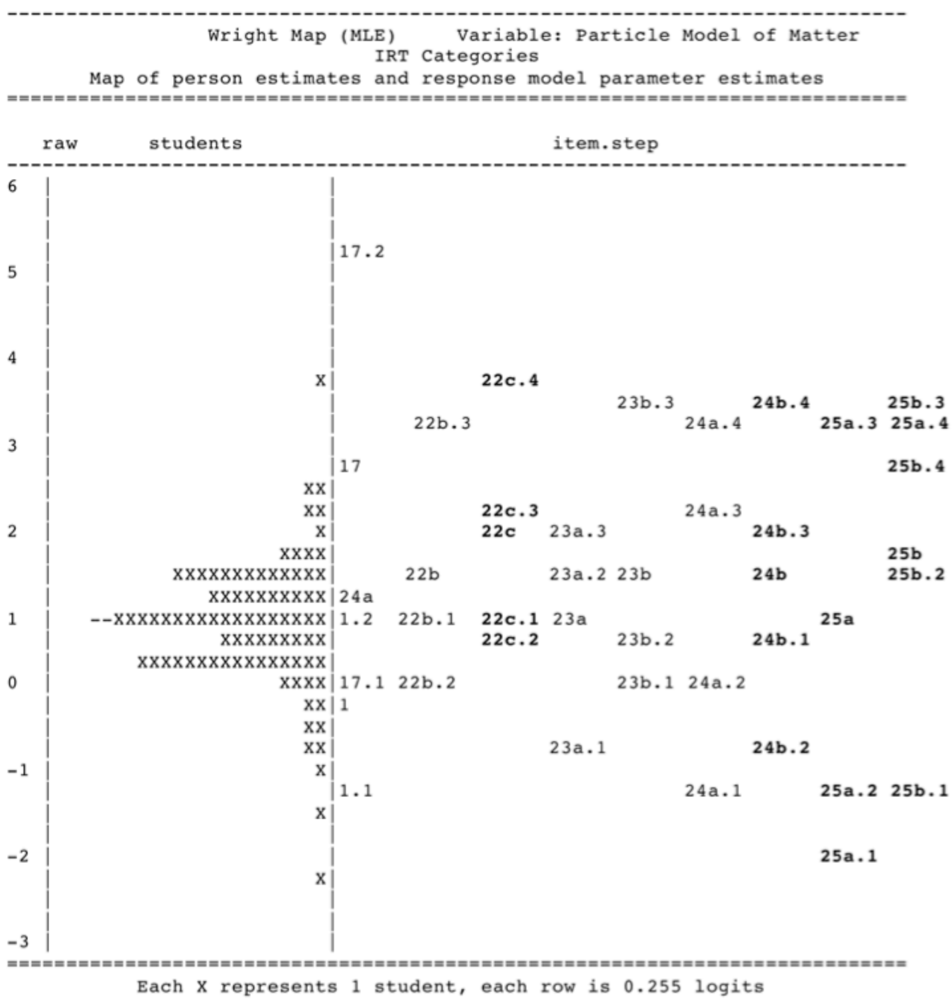



**Figure 4.3. Wright map of steps for polytomous items.**

Table 4.3 displays the actual step difficulties, where the higher the number, the more difficult the step. Thus, if the step difficulties were ordered, then they would be increasingly more difficult (higher in number). Item 22.C and item 24B have disordered steps in which, it is easier to go from step 1 to step 2 than to go from step 0 to step 1 (see

Table 4.3). Items 25.A and 25.B have their last two steps disordered. For example, for item 25.A it is easier to go from a step 3 to a step 4 (3.31) than to go from step 2 to step 3 (3.19).

**Table 4.3. Step Difficulties for 4 items with disordered steps.**

| | Step difficulties | | | |
|---|---|---|---|---|
| **Item** | **0 to 1** | **1 to 2** | **2 to 3** | **3 to 4** |
| 22, part C | 0.98 | 0.41 | 2.63 | 3.32 |
| 24, part B | 0.50 | -0.51 | 2.00 | 2.37 |
| 25, part A | -1.89 | -1.30 | 3.31 | 3.19 |
| 25, part B | -1.31 | 1.24 | 3.48 | 2.48 |

Items 22C and 24B have the same disordered steps. In both cases, it is easier to go from a "Mixed" level explanation of a phenomenon to a "Incomplete" level than from a "Descriptive" level explanation to an "Incomplete" explanation. This points to a lack of distinction between the "Mixed" and "Incomplete" levels. This is also reflected in the scoring guides for these items (see Appendix B) in that both levels are incorporating some of the same concepts.

Items 25A and 25B both have disordered steps where it is easier to go from a "Basic" level response to a "Complete" level response than to go from an "Incomplete" level response to a "Basic" level response. In each instance, this is primarily due to the sample of students involved in the study. For each item, there was only one student who reached a "Complete" level response. Thus, the lack of students responding at the highest level is reflected in the disordered step difficulties.

**Wright map discussion**

In sum, I have examined the Wright map using both Thurstonian thresholds and step difficulties. As expected, the Thurstonian thresholds were ordered. This insures that interpretations related to item difficulties and person abilities can be made. For example, easier items are at the lower end of the Wright map while more difficult items are at the higher end. This also means that persons and items can be interpreted together such that respondents that are located at 0.0 logits have a 50% chance of answering items at that location correctly.

Analysis of the step difficulties found that slightly less than half of the polytomous items have disordered step difficulties. Disordered steps for two of the items indicated that there are not enough students responding at the highest levels of the construct for those items. In addition, two items showed that there is a lack of distinction between "Mixed" and "Intermediate" level explanation responses. This is an indication that the levels of the construct map need to be modified.

Overall, the Wright map indicates that the model is meeting the technical requirements of the Wright map. This is indicated in that the locations of both items and respondents can be used to interpret performance. However, analysis of item steps indicates there is a lack of distinction between two levels and that the population sample does not include enough student observations at the highest level of the construct. What still needs to be determined is the relationship between the Wright map and the construct map.

**Determining Performance Levels**

The Thurstonian thresholds were used to determine the performance levels that correspond with the levels of the construct map. Cut-points mark the boundaries of the levels and they are the midpoint between the means of the Thurstonian thresholds of two consecutive levels. Thurstonian thresholds do not exist below the first step of an item, so there is no mean threshold value for the lowest category (Descriptive level); thus, there is no midpoint between the lowest two categories – the "Descriptive" and "Mixed" levels.

For this study, I used the 67% confidence interval around the mean for the "Mixed" level to define the cut-point between the first and second categories. Dichotomous items only have one step, so their step difficulties are their Thurstonian threshold (see Appendix D).

One way to look at Thurstonian thresholds is produced by the ConstructMap program (Kennedy, Brown, Draney & Wilson, 2005). The thresholds along with student proficiencies for all items are displayed in Figure 4.4. The histograms on the left side of the map are the student proficiencies and on the right side, the thresholds for each item are shown in columns. The name of each item is displayed on the x-axis. The thresholds for the open-ended items (22-25) are generally higher than those for the multiple-choice items (1-21).

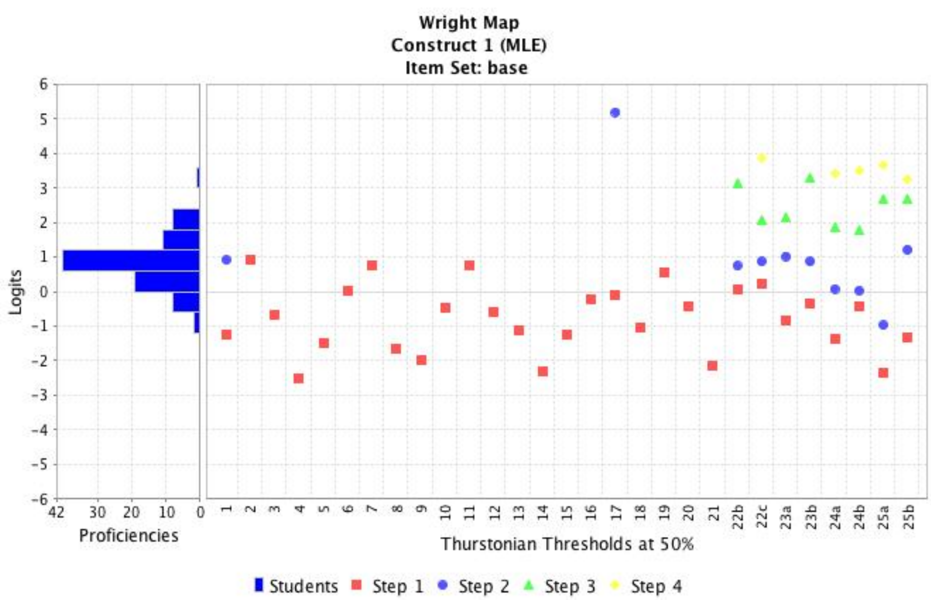



**Figure 4.4. Graphical Wright Map of Thurstonian thresholds for the items.**

For each level of the construct map, the corresponding item responses represent a range of difficulty level, as some items are easier than others. This is reflected in the range of thresholds for each level. Table 4.4 displays the threshold ranges and means for each level of the construct map. For example, the six items that measure responses at the "Complete" level have thresholds ranging from 3.36 to 5.29 logits and the average threshold value for these items is 2.92 logits. The twelve items with responses corresponding to the "Mixed" Level ranged from -2.23 to 0.64 logits, with a mean of -0.61 logits. Using the 67% confidence level of -0.33, the mean for the descriptive level is calculated to be -0.96 logits.

**Table 4.4. Threshold range and means for the different levels of the construct map**

| Level (# of items) | Range (in logits) | Mean (in logits) |
|---|---|---|
| Complete (6) | 3.36 to 5.29 | 3.92 |
| Basic (10) | 0.85 to 3.24 | 2.24 |
| Incomplete (24) | -2.43 to 1.31 | -0.36 |
| Mixed (12) | -2.23 to 0.64 | -0.61 |
| Descriptive | N/A | -0.96 |

This examination of the range and means of the items finds that the responses to the 24 items aligned with the incomplete level ranged from the descriptive to the basic level of the construct. The criteria zones, cut-points and means for the different levels are represented in Figure 4.5.

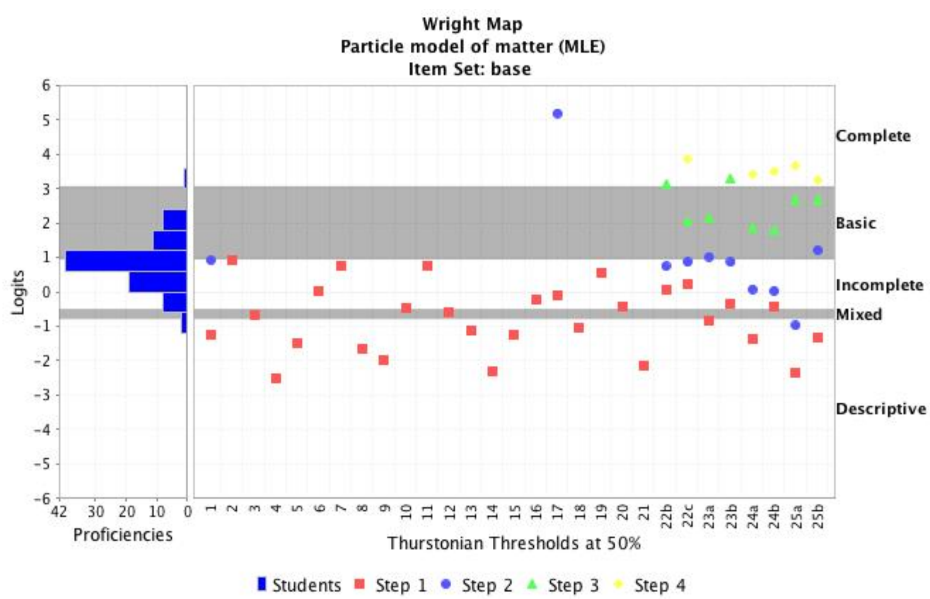



**Figure 4.5. Mapping performance levels to the Wright map.**

Figure 4.5 shows the relationship between the item steps and the construct map using a graphical Wright map. The map shows the range of the different levels, called criterion zones, to the corresponding student proficiencies that likely yield a specific level of response on most items. It shows that the average for the "Incomplete" level is very close to the "Mixed" level and that the "Mixed" level is very narrow in range. This is another indication that these levels are not distinct. In addition, Figure 4.5 further illustrates the range of thresholds for the "Incomplete" level encompasses both the "Basic" and "Mixed" levels of the construct map. There are 27 items that have responses related to the "Incomplete" level. Figure 4.5 shows that most items do not fall within the range of the "Incomplete" level; instead, most responses are found at the mixed or descriptive level. This is another indication that these levels are not distinct.

**Performance Level Discussion**

The performance levels of the items were used to link the Wright map to the progress variable. The design of the items (see Chapter 3) included the development of items of a range of difficulty and thus, it was expected that there would be a range of difficulties for each level, with some items being more difficult than others, perhaps even producing a few outliers. None of the items aligned with the "Complete" level of the construct map fall outside of the complete level, which is not unexpected because five of the six items that align with the "Complete" level are associated with written responses items (Items 22B to 25B). The single multiple-choice item aligned with the "Complete" level was item 17 (see Figure 4.5).

Of the ten items aligned with the “Basic” level, only three of these items fall outside of the “Basic” criterion zone. Item 11 (see Figure 4.5) falls slightly outside of the zone into the “Intermediate” zone. The other two items, 22B and 23B are explanation portions of written response items. These items appear in the “Complete” zone (see Figure 4.5). What is different about these two items is that there were no students who responded at a “Complete” level to these items, which resulted in these items ranging slightly higher.

There are 12 items that were aligned with the “Mixed” level of the construct map. Eight of these items are related to the written response items, which are more difficult by design. Therefore, it was expected that many of these responses could range into other levels (see red squares for items 22B to 25B in Figure 4.5). Although some of these items fall into the “Descriptive” zone, this could indicate that the “Mixed” level should be wider and that the 67% confidence interval is not wide enough for distinguishing between “Mixed” and “Descriptive” level responses.

Analysis also revealed that most of items aligned with the “Incomplete” level spanned from the “Descriptive” to the “Basic” level in terms of range of difficulty. Moreover, a majority of these items (15 items) were found to be outside the range for the “Incomplete” level. Five of these items are written response items, and most were more difficult, as expected. The other 10 items, most multiple-choice items, fell into the “Mixed” and “Descriptive” level zones. Thus, both the disordered step difficulties as well as linking of the construct map to the Wright map indicate that there are too many levels for the progress variable.

**Item Fit**

The item fit is investigated as evidence that the model being used is appropriate. The mean square fit statistic, or infit, is one means for measuring item fit. Residual are the difference between the observed score and expected score for a person responding to a particular item (Wilson, 2005). The infit is a ratio of the expected squared residual and the observed squared residuals. Thus, when the observed residuals vary as much as expected the mean square is 1. The infit was provided as a report in the ConstructMap program. A range of mean square value with a lower bound of 0.75 and upper bound of 1.33 is deemed acceptable (Wilson, 2005). Thus, items that fall outside of this range have a poor fit. Poor fit indicates that the items are not performing predictably; therefore, these items do not fit the model.

**Item Fit Discussion**

Overall, the items showed a good fit with average infit mean squares of 1.00. This indicates that overall, the items are behaving as expected. Figure 4.6 graphically shows that the infit mean squares fall in between the acceptable range for all items. The infits for all the items range from 0.85 to 1.18. Thus, no items ranged outside of the acceptable range. This indicates that the items are performing as expected and that the data fits the model. In sum, examination of item fit shows that the items are measuring the construct and provides evidence that the model being used is appropriate.

```
                               Item Estimates
                               Item Set: base
                           Variable: Construct 1
                             Infit Mean Squares
============================================================================
           0.58  0.67  0.75  0.83  0.92  1.00  1.08  1.17  1.25  1.33  1.42
-------------+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+--
Item 1                   .       *         |                       .
Item 2                   .            *    |                       .
Item 3                   .             *   |                       .
Item 4                   .              *  |                       .
Item 5                   .                 |*                      .
Item 6                   .                 |  *                    .
Item 7                   .                 |         *             .
Item 8                   .                 *                       .
Item 9                   .                 |     *                 .
Item 10                  .                 |  *                    .
Item 11                  .                 *                       .
Item 12                  .         *       |                       .
Item 13                  .                 |   *                   .
Item 14                  .            *    |                       .
Item 15                  .              *  |                       .
Item 16                  .          *      |                       .
Item 17                  .                 |    *                  .
Item 18                  .                *|                       .
Item 19                  .                 |    *                  .
Item 20                  .        *        |                       .
Item 21                  .                *|                       .
Item 22B                 .                 |   *                   .
Item 22C                 .            *    |                       .
Item 23A                 .             *   |                       .
Item 23B                 .     *           |                       .
Item 24A                 .                *|                       .
Item 24B                 .          *      |                       .
Item 25A                 .               * |                       .
Item 25B                 .               * |                       .
============================================================================
```

**Figure 4.6. Item Infit Mean Squares for the Particle Model of Matter.**

**Respondent fit**

A student sometimes will provide incorrect answers to items that the model predicts they should get correct and vice versa. Respondent fit is unique to Rasch modeling in that it allows for the determination of these unusual response patterns compared to expectations based on both item and respondent threshold locations (Wilson, 2005). Examination of respondent fit is not intended to determine causal relationships. For example, results could be due to student guessing or cheating. However, results can be used to identify patterns of responses, especially for those students for whom the model is not a good fit.

Respondent fit is similar in scale to item fit. Students who have infits below 0.75 are providing consistent responses to items. This means that the student is responding consistently with what the model predicts – getting items correct the model predicts they should and vice versa. Students who have infits between 0.75 and 1.33 have a “good” fit. This means that the students are responding to the items consistently, but have items that they are answering items correctly that the model predicts they should not and vice versa. Students who have infits above 1.33 have a “random” fit. These are the students for whom the model is not a good fit as they are getting items correct the items the model predicts they should get incorrect and vice versa.

Respondent fit was investigated through kidmaps, which are modified Wright maps, as well as examining infits related to students’ ability estimates. Kidmaps (see Figure 4.7) display the student’s location, the items they answered correctly and incorrectly as expected, as well as items they were expected to get wrong, but answered correctly. As examples of the respondent fit, two cases were chosen (see Table 4.4). These two cases represent the same ability level along the Wright map, but they have very different response patterns.

**Table 4.5. Two cases of responses for Particle Model.**

| Student ID | Score | Ability | Infit |
|---|---|---|---|
| 1064 | 30 | 0.95 | 0.97 |
| 1039 | 30 | 0.95 | 1.67 |

As shown in Figure 4.7, kidmaps display the item students have provided responses at levels that the model predicts they should reach appear in the left-hand column labeled “Level Responded” and those that they have not reached appear in the right-hand column labeled “Next Level” (Wilson, 2005). Figure 4.7 represents a student (1064) with a “good” fit. As discussed earlier, this means the student is responding consistently to items as the model predicts, but there are items the student is getting correct the model predicts they should not and/or vice versa.

```
----------------------------------------------------------------------------
Diagnostic Map (MLE)
Student    : 1064
Variable   : PMM
Ability    : 0.946
Infit ms   : 0.965  t stat : -0.009
-----------------Level Responded-----------------Next Level-----------------
                                   |   |
                                   |   |
                                   |   |
                                   |   |17.2
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |22c.4
                                   |   |25a.4
                                   |   |23b.3 24a.4 24b.4
                                   |   |22b.3 25b.4
                                   |   |
                                   |   |25a.3 25b.3
                                   |   |
                                   |   |22c.3 23a.3
                                   |   |24a.3 24b.3
                                   |   |
----------------------------------------------------------------------------
                                   |   |25b.2
                                2.1|   |23a.2
                    1.2 22c.2 23b.2|XXX|7.1 11.1 22b.2
                               19.1|   |
----------------------------------------------------------------------------
                                   |   |
                     6.1 17.1 22b.1|   |24a.2 24b.2
                               16.1|   |
                     10.1 12.1 20.1|   |24b.1
                          3.1 23a.1|   |
                    13.1 18.1 25a.2|   |
                         15.1 25b.1|   |
                      5.1 8.1 24a.1|   |
                                   |   |
                                9.1|   |
                          14.1 21.1|   |
                                4.1|   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
============================================================================
                          Each row is 0.255 logits
```

**Figure 4.7. Kidmap:  Good Fit to Particle Model construct.**

The “XXX” is the respondent’s location and the dotted lines are the “surprise lines” – the threshold for which a respondent is expected to answer correctly. The respondent has a 50% chance of answering items correctly when they are located within the surprise lines. In Figure 4.7, the items that appear in the section of the map where the “XXX” represents the items that student 1064 had a 50% probability of answering correctly, responding to five items correctly and incorrectly to five items. As expected for a student with “good” fit, most item responses for this student appear in the lower left-quadrant and the upper-right quadrant with a few items outside these areas. Student 1064 has not responded at the level the model predicted for three items: 24a.2, 24b.1, and 24b.2. Accordingly, these items appear in the lower right quadrant of Figure 4.7.

Student 1039 (see Figure 4.8) represents a “random fit,” which means a lot of unexpected outcomes in his or her responses. For the most part, he or she has not responded correctly to more difficult items as expected, but has random responses to both lower level items and some higher-level items. For example, the student scored higher than expected on items 22b.3, 25a.3 and 24a.3 (appear in upper left section of Figure 4.8). However, this student did not respond correctly to lower difficulty items, as the model predicted. This is reflected in that lower level item responses (i.e. 21.1) appear in the lower right quadrant of the kid map.

```
----------------------------------------------------------------------------
Diagnostic Map (MLE)
Student    : 1039
Variable   : PMM
Ability    : 0.946
Infit ms   : 1.668  t stat : 1.899
-----------------Level Responded-----------------Next Level-----------------
                                   |   |
                                   |   |
                                   |   |
                                   |   |17.2
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |22c.4
                                   |   |25a.4
                                   |   |23b.3 24a.4 24b.4
                              22b.3|   |25b.4
                                   |   |
                              25a.3|   |25b.3
                                   |   |
                                   |   |22c.3 23a.3
                              24a.3|   |24b.3
                                   |   |
----------------------------------------------------------------------------
                              25b.2|   |
                                   |   |2.1 23a.2
                     7.1 11.1 22c.2|XXX|1.2 23b.2
                               19.1|   |
----------------------------------------------------------------------------
                                   |   |
                               17.1|   |6.1 24b.2
                               16.1|   |23b.1
                         20.1 24b.1|   |10.1 12.1
                           3.1 23a.1|   |
                               13.1|   |18.1
                           1.1 15.1|   |
                            5.1 8.1|   |
                                   |   |
                                9.1|   |
                               14.1|   |21.1
                                4.1|   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
                                   |   |
============================================================================
                          Each row is 0.255 logits
```

**Figure 4.8. Kidmap: Random fit to Particle Model construct.**

## Respondent fit discussion

Respondent fit is examined to determine whether the model fits students' responses. Overall, there is a good respondent fit to the items, with an average infit of 1.01 (see Appendix B) and the model fits 85.3% of students. Although it did not happen with this study, student interviews would have needed to be conducted to find out what he or she was thinking to understand the randomness of the response patterns of the 13 students for whom the model does not fit.

## Summary

The technical side of the model, the Wright map has been analyzed to determine whether the model is appropriate for measuring the particle model of matter (PMM) progress variable. Analysis shows that the model is a good fit for the data. As expected, the item thresholds are ordered. In addition, the item fit and respondent are appropriate for measuring the (PMM) progress variable. However, analysis of the step difficulties indicated that there are too many levels for the PMM construct map. Linking of the PMM construct map to the Wright map also provided further evidence that there are too many levels. This represents the first step in the evaluation of the assessment items. The consistency of the items needs to be investigated by analyzing reliability.

## Reliability

Reliability seeks to answer the question: If provided the same assessment, would the scores agree? Thus, determining reliability is a means of investigating the consistency of the items for assessing (Cronbach, 1990; Wilson, 2004), in this case, students understanding of the particle nature of matter. The evidence for reliability was investigated through the standard error of measurement, the separation reliability, alternate forms reliability, and interrater reliability.

### Standard Error of Measurement

There is a degree of uncertainty associated with each estimate. The standard error of measurement (SEM) “tells the measurer how accurate each estimate is” (Wilson, 2004, p. 126). SEM is the error associated with the location of the respondents. For a single respondent, there is standard error associated centered at his or her location on the Wright map, with an approximately normal distribution (Wilson, 2004; Embretson & Reise, 2000). The student's estimate and standard error can then be used to determine their 67% and 95% confidence interval. Examination of these intervals identifies whether they are reasonable. For instance, you do not want the intervals spanning the entire range of the map. The mean estimate for all respondents is 0.89 logits, with an average standard error of 0.33 (see Appendix C). This means that a student with an ability estimate of 0.82 logits has uncertainty as to their exact location, but their location is centered at 0.82 logit, with an error of ±0.35. A student at this ability level would have a 67% confidence interval of 0.82 ±0.35 (0.47, 1.17) or alternatively, a 95% confidence interval of 0.82 ±1.96*0.35 (0.17, 1.51). This means, you could expect the student's ability estimate to be between 0.47 and 1.17 logits, 67% of the time and between 0.17 and 1.51 logits 95% of the time.

A second method for examining the SEM is to graph the standard error of respondent locations versus the estimates of students' estimates. This relationship is typically a "U" shape where the minimum is near the mean of the item thresholds and increasing values towards the extremes (Wilson, 2005). The SEM is smaller in the middle, such that students in the middle always have more items near them than the extremes. Thus, it is desired that the item thresholds be distributed in "a uniform way over a construct" (Wilson, 2005, p.142). With this analysis, one is looking for the range of logits that provide the most information as well as whether this range matches what the construct is measuring. For this study, most items are found between the mixed and basic levels; therefore, it is desired that the range of logits that provide the most information match these levels.

The SEM of the items did not produce a perfect "U"-shape (see Figure 4.9). However, there is a partial U-curve centered near 0.5 logits. This "low" point is where the items provide more information, and the peak is where the items are providing less information. This indicates that I do not have enough items that test students with really low ability (less than -0.75 logits). It also indicates that there are not enough items for assessing students with very high ability (greater than 1.25 logits). Overall, the items measure the construct, providing more information about students with abilities between -0.75 logits and 1.25 logits.


**Figure 4.9. The SEM for the particle model of matter construct.**

Analysis of the standard of measurement indicates the items are reliable measures of students with ability estimates between -0.75 and 1.25 logits, which is a large proportion of the sample. The SEM curve also indicates that the items are consistent measures of student abilities as there is a narrow range of error between 0.33 and 0.49 logits.

**Separation reliability**

The separation reliability is an internal consistency coefficient that measures the difference in the observed total variance of the estimated locations and the variance accounted for by the errors (Wilson, 2004; Cronbach, 1990). Separation reliability is generated as a part of the "Ability Estimates with Fit Statistics" report of ConstructMap. This measure of reliability provides how much variance is accounted for by the estimate of a respondent's location. There are no absolute standards for determining what an acceptable level of reliability (Wilson, 2004; Cronbach, 1990).

However, the separation reliability can provide insight into different aspects of the sample and the assessment. Higher reliability coefficients are associated with samples that have a wider range of abilities of students in the sample. Higher reliability coefficients are also associated with longer tests (more items), as well as the number of levels per item (Cronbach, 1990).

The separation reliability of was calculated to be 0.75. Thus, the estimator of a respondent's location accounts for 0.75 of the variance of the distribution of ability estimates. First, this indicates that the students in this sample do not have a wide range of different abilities. Second, there may not be enough items of varying difficulties. This is in agreement with the item distribution discussed in determining the relationship between the construct map and the Wright map, which shows that 24 of the 27 total items have item responses related to the "Incomplete" level of the construct map. This is twice as many items as those associated with the "Mixed" level (12 items). Finally, the reliability indicates that there are too many levels for the progress variable; therefore, the levels of the construct map need to be modified. This was already indicated during analysis of the model where it was determined that the item levels were not distinct.

**Interrater reliability**

Interrater reliability was determined for the scoring of the written response items. Both forms were scored using the scoring guides detailed in Appendix B. The open-ended items were scored with one other rater who was familiar with the assessment and my research. Two scoring rounds of ten tests each were conducted. We obtained inter-rater reliability of 90% for both runs, which was determined by coding ten tests and dividing the number of items coded identically by the total number of items coded. This indicates that there was a high level of agreement between the raters and that the levels were modified for the scoring guides appropriately.

## Summary

Each analysis provides different information about the reliability of the items. The SEM indicates that the items are better at measuring students with abilities ranging from --0.75 to 1.25 logits. Items with difficulties in this range provide the most information about student performance. The separation reliability determined that the sample does not include a wide range of students and needed to include more items, as most items are associated with the "incomplete" level of the construct map. In addition, it provides further evidence that there are too many levels of the construct map. In this next section, the validity of the items is discussed.

## Validity

Test validity refers to the extent to which the items actually measure the particle model construct. It also determines to what degree inferences based on test scores are appropriate and meaningful (Cronbach, 1990; Kane, 2001). The validation process is important for collecting evidence to support the claims interpreted from student scores. The validity of the items is evidenced through several measures. The *Standards for Educational and Psychological Testing* (1999) identifies five categories for gathering evidence of validity:

1. Evidence based on instrument content
2. Evidence based on response processes
3. Evidence supporting the internal structure
4. Evidence based on relations to other variables
5. Evidence based on consequences of testing

In this study, the first four categories are utilized to analyze validity. The fifth category is not investigated because this form of validity is still debated; including what evidence is used to argue the consequences of testing (Kane, 2001). The focus of this study is to examine whether the items actually measure the particle nature of matter construct and while it is important to always be aware of consequences of any instrument, this would require an additional study.

## Evidence Based on Instrument Content

Evidence based on instrument content is accomplished through "an analysis of the relationship between a test's content and the construct it is intended to measure" (AERA et al., 1999) as well as analysis concerning validity of the test (Wilson, 2005). Content validity has been evidenced through detailing the process of developing the construct map, design and development of items, determining how items will be scored and will also be evidenced through calibration of the Wright map, which represents a technical version of the construct map.

## Evidence Based on Response Processes

Evidence based on response processes refers to analysis of individual responses. The evidence gathered through the exit interviews occurs when students explain how they arrived a particular answer or why they found a particular item(s) difficult to respond to. The exit interviews also serve to set up "many of the expectations that are to be compared to outcomes during the investigations into other aspects of validity" (Wilson, 2005). Thus, the exit interviews helped to identify items students found difficult (and why) as well as their thought process for answering particular items.

Conducting exit interviews of four students around each instrument provided evidence based on response processes. These interviews were designed to prompt student responses to the items, especially the open-ended items. Students were not expected to respond to every item because of sensitivity to the fact that they are middle school students. If more time had permitted, the researcher would have interviewed more students to get more information on any issues pertaining to items.

Results from the interview indicated that two of the students had issues with the thermometer items (items 6 and 17). Item 6 responses focus on the explaining how thermometer works on a macrolevel (describing what happens) while item 17 focuses on explaining how a thermometer works on a molecular level. Both students expressed concern with question 6, with one boy student stating: "maybe in the answers it should say like what the substance is then maybe we would have a little more understanding of how it is going to be affected". The second student stated: "I don't really remember going over like how the red stuff moves up and everything." He also expressed difficulty with the second temperature item as well. In sum, both students expressed discomfort with answering the question because of lack of familiarity with how thermometers work. This discomfort with the thermometer related items is reflected in item 17 being the most difficult item and item 6 as the 14th most difficult item (see Evidence Supporting the Construct Map)

These interviews also indicated that three of the four students found the items overall to be "easy." One student, who had experienced the unit the previous year stated: "Last year it was more difficult." This is reflected in students' performances on the assessment centered above zero logits, meaning most students had a greater than 50% probability of responding correctly to most items. The fourth student did not express his thoughts on the overall test, just on items he found difficult. For example, the first item (see Figure 4.10) is about what model best represents a gas. The student found this item difficult because: "it did not have any actual keys or anything to uh show like uh how to find the right answer." This item was found to be of moderate difficulty for most students (see Evidence Supporting the Construct Map).

1. **Below are four possible models of a gas.** Which model would a scientist use to show how water vapor condenses to a liquid?

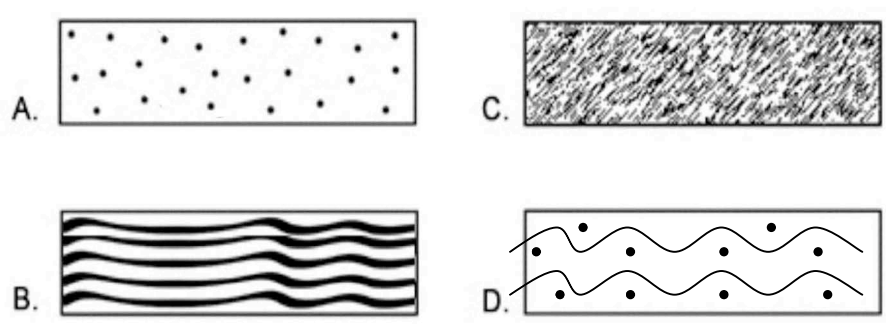



**Figure 4.10. Item 1 of the assessment.**

All four students expressed a difficulty in remembering how to depict a phase change, especially in responding the written response items. When probed about these items, they expressed difficulty in remembering how to answer these items. A common response that all students had were expressed by a student who stated about why these items were more difficult, “I guess cause you had to think more than the other ones, you know. It wasn’t just like, oh your teacher told you this”. This is also reflected in that some of the most difficult items are related to the open-ended items (22B/C, 23A/B, 24A/B and 25A/B), which appear in the upper levels of the Wright map.

## Evidence Based on Internal Structure

Evidence based on internal structure has been gathered in relation to two different aspects of the instrument – the instrument level and the item level. At the instrument level, there will be a correlation of the expected order of item difficulty with Wright map estimated locations to determine the Spearman rank-order correlation. At the item level, the mean locations will be examined to insure that students with higher abilities score higher on each item.

### Evidence Supporting the Construct Map

The general rule for a good instrument is that the items span the full range of respondent locations (Wilson, 2005). Thus, it is important to examine whether items spanned the construct as expected. In other words, does the order of the items, from low to high difficulty, match the construct? Table 4.5 lists the expected ranking for item difficulties, with 29 being the easiest and 1 being the most difficult. These rankings were based on both prior research results and hypothesizing results for new items based on similarity to content and structure of current items. The last column is the ranking of the estimated locations according to the Rasch model.

The relationship between the expected order and the estimates can be quantified using the Spearman rank-order correlation. The correlation for this data was found to be 0.92. In other words, the two rankings are highly correlated.

**Table 4.6. Theoretical Expectation versus Actual Item Outcomes.**

| Item# | Topic(s) | Expected | Estimated |
|---|---|---|---|
| 17 | Particle model, liquids, how thermometer works | 5 | 1 |
| 22C | Particle model explanation, movement of gases in relation to temperature | 1 | 2 |
| 25B | Particle model explanation, mixing of gases | 2 | 3 |
| 22B | Particle model drawing, movement of gases in relation to temperature | 6 | 4 |
| 23B | Particle model explanation, phase change from liquid to gas | 3 | 5 |
| 24B | Particle model explanation, states of matter – solid, liquid and gas | 4 | 6 |
| 24A | Particle model drawing, states of matter – solid, liquid and gas | 8 | 7 |
| 2 | Explanation of condensation on a molecular level | 7 | 8 |
| 23A | Particle model drawing, phase change from liquid to gas | 11 | 9 |
| 25A | Particle model drawing, mixing of gases | 14 | 10 |
| 11 | Identify element from compounds, given molecular formula | 10 | 11 |
| 7 | Biology, gases | 9 | 12 |
| 19 | Evaporation | 13 | 13 |
| 6 | Explanation of how a thermometer works | 12 | 14 |
| 1 | Identify particle model of gas | 18 | 15 |
| 16 | Particle model of a gas after removal of gas | 15 | 16 |
| 20 | Phase change | 20 | 18 |
| 10 | Explanation of condensation | 21 | 17 |
| 12 | Particle model, empty space | 17 | 19 |
| 3 | Identification of elements, given names of molecules and compounds | 19 | 20 |
| 18 | Liquids, evaporation | 20 | 21 |
| 13 | Particle model, phase change | 27 | 22 |
| 15 | Particle model, explain liquid movement | 25 | 23 |
| 5 | Particle model, properties | 26 | 24 |
| 8 | Liquids, evaporation | 23 | 25 |
| 9 | Particle model, evaporation | 28 | 26 |
| 21 | Identify compound, given molecular formulas | 16 | 27 |
| 14 | Particle model, phase change (solid to liquid) | 29 | 28 |
| 4 | Particle model, phase change | 22 | 29 |

The Wright map (see Figure 4.2) provides evidence that the rankings were highly correlated, as the item steps indicate a positive relationship. The lower level responses (step 1) for items are listed in the lower portion of the map. For example, item 1 has two steps. Its lowest step, 1.1, is found just below -1.0 logit. Items with higher difficult, like the second step for item 1 (1.2) are found in the just above 1.0 logits on the map. This range also includes more difficult items that have one step, such as item 2.1, which as expected, is meant to measure a more difficult level of the construct map. As expected, the top half of the map features the highest steps for items. This indicates that the levels of the construct are being measured from low to high as expected. These results also reflect the overlapping of the "Incomplete" level responses to both higher and lower levels of the construct map in that there are large differences between the expected and estimated rankings. For example, I expected items 4 and 21 to be more difficult than they were estimated to be (see Table 4.6).

## Evidence Supporting the Item Design

### Item Analysis

Item analysis provides a means for assessing the consistency of the items. The items were designed, such that students with higher ability are more likely to score higher on more difficult items. Thus, as scores increase the average location of each group should increase (Wilson, 2004). For the most part, this was true. However, there was one item (see Table 4.7) that did not fit this pattern. This item was further examined to determine whether these results could be explained.

Item 22.c is the explanation portion of question 22, part C. A scoring guide was used to determine the code for each response level. This item shows an interesting response pattern between response categories 1 and 2 because the mean ability for category 2 is lower than that for category 1. This is another indication that the levels of the construct map are not discriminating between a level 1 and level 2 response.

**Table 4.7. Item Statistics for Select Particle Model Construct Items.**

| | Response Categories | | | | |
|---|---|---|---|---|---|
| **Statistics** | **0** | **1** | **2** | **3** | **4** |
| Item 22.c | | | | | |
| Count | 29 | 20 | 28 | 10 | 1 |
| Percent (%) | 32.95 | 22.73 | 31.82 | 11.36 | 1.14 |
| Pt-Biserial | -0.37 | 0.07 | 0.02 | 0.31 | 0.33 |
| Mean Ability | 0.46 | 0.91 | 0.83 | 1.43 | 3.53 |
| SD Abilities | 0.35 | 0.35 | 0.35 | 0.37 | 0.49 |

**Item Analysis Discussion**

Items are designed such that students of higher ability are more likely to respond correctly to more difficult items. Most of the items that comprise the assessment were found to have met this expectation. Only one item did not meet this expected pattern.

The explanation portion of written response item 22, again pointed out issues with the construct map. The explanation item showed the average ability of students responding at level 1 being higher ability than the ability of the group of students responding at level 2. Examining the scoring guide for this item (see Appendix B) also demonstrates that "Incomplete" level responses may not be distinct from the other levels, especially "Mixed" level responses.

In sum, item analysis demonstrates that the items are consistent in that for most items students higher on the construct are scoring higher on each item. In addition, written response items showed the greatest difference between mean ability for each response category. This suggests that the open-ended items may provide a better measure of the construct.

## Evidence Based on Relations to Other Variables

Theory indicates there should be a strong relationship between the instrument and external variables (Wilson, 2005). In this case, students' grades were collected and analyzed in relation to students' performance on the instrument. In addition, a Pearson correlation was completed to determine if student performance was related to whether students had prior experience with the sixth grade unit.

The evidence for a relationship to other variables can be determined using the Pearson correlation coefficient, which in this case is a correlation between the students estimated locations and their grades. The expectation was to see convergent evidence, or in this case a positive relationship between the two variables (grade increases as students' estimated location increased). The correlation was found to be 0.05, which is positive, but there is basically no correlation. This relationship between the particle model and grade is also shown in Figure 4.11.

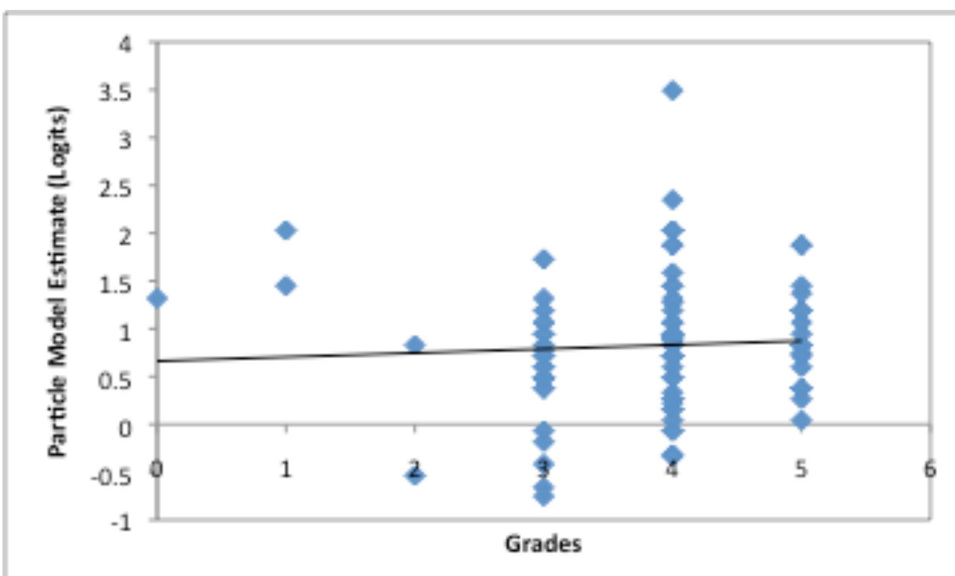



**Figure 4.11. Relationship between the Particle model construct and grades (A=5, B=4, etc).**

In addition to the relationship between students' grades and their proficiency estimates on the particle model construct, a group correlation focused whether students had experienced the sixth grade unit was completed. This was done to determine whether there was a curriculum participation effect.

Two groups of students were involved in this study, those who had experienced the sixth grade unit (n = 71) and those who had not (n = 18). Students who experienced the sixth grade unit had a higher average raw score and ability than those who had not (see Table 4.8). The correlation of raw scores to students experiencing the sixth grade unit is 0.40. In addition, the PMM construct was weakly correlated (0.38) to students having experienced the unit. The higher separation reliability for students with no experience is also further evidence that the items are better at measuring students with abilities within the range of -0.75 and 1.25 logits. There is a moderate correlation between proficiencies and experiencing the sixth grade unit.

**Table 4.8. Averages for students who did and did not experience the sixth grade unit.**

| | Total # students | Average Raw Score | Average Ability | Separation reliability |
|---|---|---|---|---|
| Experienced unit | 71 | 30 | 0.95 | 0.69 |
| No unit | 18 | 24 | 0.29 | 0.76 |

**Discussion**

Results show that the total raw scores are only moderately correlated with students having experienced the sixth grade chemistry unit. This means that those who experienced the unit performed better on the items. In addition, there is a low correlation between the particle model of matter construct and students' experience with the sixth grade unit. This suggests that the particle model of matter construct is only slightly related to students experiencing the unit.

Teacher grades were also examined as an external variable that could be related to the particle model of matter construct. Analysis found no relationship between the two variables. This is not an unexpected result as student grades take into account more than just test grades, such as class work, homework and quizzes.

## Summary

Overall, the many forms of evidence gathered indicate that the instrument is valid and measures the PMM construct. Content validity has been evidenced through describing the construct, the item development process, the outcome space and evaluation of the Wright map. Evaluation of the Wright map indicated that the model is appropriate; this was reflected in that the expected difficulty of the instrument is highly correlated to the difficulty estimates generated by the model. This relationship was also evident in that the items students expressed as being the most difficult were found to be some of the most difficult items based on model estimates.

Item analysis indicated that most items were consistent in measuring the construct. Only one of the 29 items was determined to not be consistent. This item provided further evidence that there are too many levels that comprise the PMM construct map.

It was also found that higher scores are moderately correlated to students having experienced the sixth grade unit. However, it was found that the instrument was better at measuring the performance of students who had not experienced the unit. No relationship was found between students' grades and their proficiency estimates.

## Study Limitations

There were limitations to this study. First, I was limited to conducting interviews during class time on one of the two days due to school events and an upcoming holiday break.  I had originally planned to conduct lengthier interviews both days about more items, but this had to be quickly revised once I arrived the first day of data collection. This limited both the number and types of questions I had originally planned to ask, as well as, the number of students I was going to interview.

Second, I wanted to determine whether the items grouped together using an exploratory factor analysis. However, there were not enough students in the sample to conduct the test. If there were more students in the study, I would be able to determine whether a unidimensional or multidimensional model is a best fit for the data.

Finally, analysis has revealed that the items are better able to test students who have abilities centered around zero logits. This was evident in both the development of the SEM curve, which indicated that more information could be gathered about students with abilities between -0.50 and 1.0 logits. If I had the opportunity to conduct this study again, I would hope to include more students who had no experience with unit. I would still those who had experienced the unit to insure I have a wide range of students to measure. A wider range of students would also help to better determine the range of abilities that the items are best at measuring.

## Conclusion

This chapter presented the results of the calibration study and discussed how this analysis has shown that the items do measure the PMM construct. Moreover, the model does not violate the assumptions of item response modeling. Each subscale is unidimensional and higher overall ability estimates are associated with higher item scores. The Wright map shows the items span the entire construct and that the polytomous items are ordered. In other words, items that span more than one level of the construct elicited responses that increased in difficulty as expected. Analysis of the item step difficulties was the first indication that the "Incomplete" level was not distinct. Further evidence of this was found when relating the construct map to the Wright map. This process determined that items related to the "Incomplete" level of the map span in difficult from the "Descriptive" to the "Basic" level.

Results indicated that there was good item fit, as no item spanned outside of the acceptable range of infit mean squares. In addition, examination of the respondent fit also showed good fit, as the model fits 85.3% of respondents.

The standard error of the mean indicates the items are good measures of students with ability estimates between -0.75 and 1.25 logits. However, there were other issues related to reliability, as the separation reliability was lower than expected because the sample of students does not include a wide range of students of differing ability. The separation reliability was also another indication that the "Incomplete" level of the construct map is not distinct.

Analysis also indicates that items are valid. Only one item was found to have disordered in the average ability of students to respond to the item at the different levels. In addition, the items do measure the construct as students' scores on the items are highly correlated to their ability estimates on the particle model of matter construct. These results also suggest that the written response items may be better measures of the construct.

As aforementioned, the performance levels of the items determined by the model were linked to the progress variable. Analysis indicates that the levels of the construct map are not distinct. Therefore, I revised the construct map such that particle aspects of the “Incomplete” level were collapsed into the “Basic” level and any macrolevel ideas were collapsed into the “Mixed” level. Table 4.9 is the modified construct map containing four levels. The bolded text highlights the main difference(s) between the levels, starting from the “Descriptive” level. For example, at the descriptive level, students have a continuous view of matter. At the “Mixed” level, the first highlighted portion shows that at this level students have a particle and continuous view of matter.

This study details the process of calibrating the Particle Model of Matter progress variable. It shows that the Rasch model is a good fit for both the items and students. This study also helped to identify that the proposed progress variable was unidimensional, but had too many levels for distinguishing students’ understanding. Moreover, it shows that the calibrated Particle Model of Matter (PMM) progress variable can be used to assess students’ knowledge of the particle model of matter. Though this is not a learning progression, results from this empirically validated study can be used to refine any hypothetical learning progressions of students understanding of matter.

In the next chapter, the modified PMM progress variable is calibrated. This variable serves as the basis for tracking student understanding during instruction as well as for answering the overarching question guiding this study: *How does student understanding of the particle nature of matter change during enactment of a model-based unit?*

**Table 4.9. Revised Particle Model of Matter Construct Map.**

| Category | Description | Example | Progressing to Next Step |
|---|---|---|---|
| Complete Particle Model | Students use particles (molecules) to explain phenomena. There is empty space between the particles. The students are able to **distinguish spacing AND motion relevant to the particular state** they are in. Different substances have different properties because they are made of different atoms OR have different arrangements of same atoms. | Water vapor, liquid water, and ice are all made up of water molecules. The molecules in water vapor are far apart and move around freely. In a liquid, they are closer together, but move around each other. In a solid, they are close together and vibrate.<br><br>Sugar and water are not the same because they are made up of different molecules. | |
| Basic Particle Model | Students use **particles** (may use atoms and/or molecules) to explain phenomena. There is empty space between the particles. Students have difficulty in explaining the difference in spacing in different states and/or are unable to distinguish the difference in movement for all states. **Different substances have different properties because they are made of different atoms or have different arrangements of same atoms.** | Water vapor is made up of water molecules that are spaced far apart and move freely everywhere. Liquid water is made up of water molecules that are moving, but are closer together than in water vapor. In ice, the molecules are even closer together.<br><br>Sugar and water are not the same because they are made up of different molecules. | Students need to understand the difference in movement of a substance in different phases. For example, a simulation of the same substance as a solid, liquid and a gas should include the same representation for water molecules, but with different spacing and movement, including how movement changes as temperature changes. |
| Mixed Model | Students use both **particle and descriptive views** when explaining everyday phenomena. When asked to describe what makes up a substance, students at this level often describe particles within a continuous medium. They do not understand that different substances have different properties **because** they are made of different atoms. Students describe solids, liquids, and gases as **made up of smaller pieces of that same substance, which come together to form a whole**. | Water is made up of water particles. The water particles exist within the liquid water. Thus, in between the particles is liquid water. | The idea that water is made up of the same atoms/molecules no matter the state should help students to realize that a substance's atoms/molecules do not change. For example, a model of ice, water and water vapor should include the same representation for water molecules, but with different spacing and movement. |
| Descriptive Model | Students at this level see objects as being a **continuous medium**. When asked to describe what makes up a common substance, they are **described exactly as they appear**. Thus, **substances always have the same properties** because the student has no concept that the substance may have a structure made up of smaller pieces. | Water is a clear, colorless liquid. Ice is a "clear" solid. They have different structures and are described differently. Therefore, they are not the same substances. | The ideas that objects are made up of parts could be a useful piece of knowledge to help students realize that a pieces of a substance that looks continuous, can be broken down into smaller pieces. Student needs to realize that a substance is changeable (it can change phases), or in other cases, may be broken into smaller pieces. |

# CHAPTER 5

## Tracking Study

### Introduction

The purpose of this study is to investigate how students' understanding changes as they engage in a contextualized model-based chemistry unit aimed to help them to develop a particle understanding of matter. The overarching question guiding this study is:

**How does middle school students' understanding of the particle nature of matter change during enactment of a model-based unit?**

This chapter reports on the analysis and findings of the Tracking study. The Calibration study determined that the items were good measures of the construct. However, analysis revealed that the progress variable needed to be modified because the levels were not distinct (see Chapter 4). Therefore, the pre/posttest items (see Table 4.7, Chapter 4) were calibrated and aligned with the construct map. Using this calibration, students learning gains from pretest to posttest are analyzed. This calibration of the PMM progress variable also serves as the basis for tracking a subset of these students as they experience the unit.

In addition to tracking students' performance using the PMM progress variable, there are two sub-questions that are answered in relation to students' performance:

a. **What knowledge of the particle nature of matter do students bring and how does this relate to the progress students make during instruction?**

b. **What is the relationship between students' drawings and explanations of phenomena?**

The first sub-question seeks to determine whether students initial understanding of matter influences their learning during instruction. During the unit, students construct models of phenomena. These models have two parts – a drawing and an explanation. The second sub-question seeks to determine the relationship between students' drawings and their explanations as they experience the unit.

## Tracking Study

## Study Setting & Participants

### Research Participants

This study collected pre/posttests from 602 sixth grade students taught by seven teachers from five schools in the Midwest and Southwest United States. School 1 is located in a rural town of varying SES in the Midwest. Two teachers were teaching the unit for the second time, while it was the first year of teaching the unit for the third teacher. School 2 is located in a suburb of a large Midwest city and the teacher has taught the unit for three years. School 3 is located in a mid-size urban city in the Southwest, whose student population is 62% Hispanic. The teacher in this school was teaching the unit for the second time. Schools 4 and 5 are located in the same large urban Midwest school district, with very different student populations. The teacher in school 4 is teaching the unit for the second time in a school with a >90% Hispanic population. The teacher in school 5 has taught the unit for four years in a school with a >90% African-

American student population. These sites were based on teacher's voluntary participation. Of these seven teachers, three were able to provide the embedded assessments for 122 students. Table 5.1 summarizes the student and teacher participation by school.

**Table 5.1. Participation of students and teachers by school.**

| *School* | *Number of Teachers* | *Student pre/posttest* | *Embedded Assessments* |
|---|---|---|---|
| 1 | 1 | 78 | 43 |
| 2 | 3 | 310 | 58 (one teacher) |
| 3 | 1 | 95 | 21 |
| 4 | 1 | 55 | |
| 5 | 1 | 64 | |
| | | N=602 | N=122 |

## Setting

### Curriculum

The study took place during the 2008-2009 school year as students experienced the sixth grade chemistry unit over an eight-week period. For a full description of the curriculum unit, see Chapter 2.

## Data Sources

Two different assessment types serve as data for this study. These assessments include the pre/posttests and the embedded assessments from lessons 1, 5, and 15 (see Appendix E). Table 5.2 includes a description of the different data sources as well as its purpose.

**Table 5.2. Tracking Study Data Sources.**

| **Data Source** | **Description** | **Purpose** |
|---|---|---|
| Pre/Posttest | Multiple-choice and open-ended assessment items given at the beginning and end of the unit | To describe students prior knowledge of the particle nature of matter. Written response items used to evaluate students explanations of their models. |
| Embedded Assessments | Lesson 1, 5, and 15 modeling activity sheets | To describe the pathway students take to a particle model. To evaluate students explanations of phenomena as they progress through the unit. |

**Student Artifacts**

Data collection included students' identical pre- and posttests. These assessment items focus on both the content and processes (modeling) students learn during the course of the unit. The results of the pretest served to provide insight into students' prior knowledge in answering the two sub-questions. For the second sub-question, for which students' models are the focus, I utilized the first open-ended test item (Items 16 and 17) as well as students' embedded assessments. This particular open-ended test item was chosen because of its similarity to the modeling tasks of the embedded assessments and serves as a means to infer students' initial models of matter. Teachers mailed back the pre- and posttests as well as student activity sheets (embedded assessments).

**Items**

The pre/posttest was used in this study (see Appendix A). These tests include identical assessment items. It is composed of 11 multiple-choice items, 9 of which were used in the calibration study and 3 open-ended items, of which two were used in the calibration study. The open-ended items are renumbered for the study because the drawing and explanation sections of the modeling items are treated as separate items. For example, item 16 is renumbered as items 16 and 17, where item 16 is the drawing and Item 17 is the explanation (see Appendix A). Because there was only a two-week turn around between data collection and when the pretests were being sent out, none of the new open-ended items from the Calibration study appear on the pretest. None of the modified multiple-choice item appear on the pre/posttest.

The sixth grade chemistry unit has three embedded assessments that have students create models to explain how odors can travel from its source to a detector, a nose. These embedded assessments occur during Lesson 1, Lesson 5 and Lesson 15 (see Appendix E). Although the questions on activity sheet 5.2 provide more scaffolding in terms of questions related to the practice of modeling, the activity remains the same in all three lessons because students have to include a key, their drawing and explanation in all three activities (1.1, 5.2 and 15.1). In addition, these activities are a way of examining students' explanations because they are the same activity that can be coded with the same rubric. The advantage of using the same rubric is that one can score the models in a manner that identifies changes in students' particle ideas.

**The Outcome Space and Scoring Guides**

Because the progress variable has been modified for the tracking study, the scoring guides were also modified for both the multiple-choice and written response items to take into account the aforementioned changes. In addition, scoring guides for the embedded assessment were developed based on the new construct map (see Appendix E).

The scoring guides, just as with the original scoring guides, contain more details and one additional level, in comparison to the progress variable. Again, this is so the scoring guide can handle all possible student responses, including when students are unable to or do not respond. The PMM scoring guides have three columns. The first column is the code to designate the particular student response. The second column describes what knowledge the students at that level include in their responses. The third column represents examples of actual student responses at that level. For both the PMM and embedded assessment scoring guides, there is a blank category for uncodable or blank responses. By defining the outcome space, the potential item responses are qualitatively linked to a performance level that corresponds to a particular level on the construct map. The analysis of the data will provide the specific cut scores that correspond to the particular levels of the progress variable.

## Scoring

The pre/posttest were scored using the scoring guides detailed in Appendix A. The open-ended items were scored with one other rater familiar with the assessment and my research. We obtained inter-rater reliability of 94.4%, which was determined by coding nine pre/posttests, then dividing the number of items coded identically by the number of items coded. The embedded assessments were scored using the scoring guide detailed in Appendix E. We obtained inter-rater reliability of 89%, which was determined by coding nine embedded assessments, then dividing the number of items coded identically by the number of items coded.

## Measurement Model

I utilized Rasch-based modeling to use the same scale to generate estimates of person abilities and item difficulties. For this study, I used a one-dimensional partial credit Rasch-based model with MLE estimation (see Chapter 4). The partial credit model is used because the test includes items that cover a single level of the construct map as well as items that cover multiple levels of the construct map. The Rasch model did not violate the assumptions of item response modeling that each subscale is unidimensional and higher scores are associated with higher abilities and the data fit the model sufficiently (see Chapter 4).

### *Item Calibration*

The Construct Map (Kennedy, Wilson, & Draney, 2009) software was used to calibrate items. The item calibration was conducted using both pretest and posttest forms.

In addition, I calibrated the embedded assessments with those of the pre/posttest items with Activity sheet 1.1 data attached to the pretest and Activity sheet 15.1 data attached to the posttest. Thus, I calibrated all the items together. Therefore each student observation was treated as two different students. For example, a student's pretest is one student and their posttest is another "student." I was then able to anchor the difficulties generated for the entire set and look at each of the items separately.

Calibration of the embedded assessments was completed with the anchored items using only the 122 students involved in this part of the study. The drawing sections of the embedded assessments were identical, so they were calibrated as multiple repetitions of a single item. The explanation questions of the embedded assessments were also identical and were also calibrated as multiple repetitions of a single item. In order to track students understanding during the unit, I first needed to determine how the calibration of the items mapped onto the construct map.

**Determining Performance Levels**

To determine cut-points between the performance levels, I used the Thurstonian thresholds of the item response categories. As a reminder, the Thurstonian thresholds indicate the ability required to achieve a response at that level or above 50% of the time. Thurstonian thresholds are computed for each step of an item. The cut-point is the midpoint between the means of the Thurstonian thresholds of two levels. Since, Thurstonian thresholds do not exist below the first step of an item, there is no mean threshold value for the lowest category or a midpoint between the lowest two categories. For this study, I used the 67% confidence interval around the second category's mean to define the cut-point between the first and second categories. Dichotomous items only have one step, so their step difficulty (item difficulty) is the Thurstonian threshold.

Once these levels were established, I was able to utilize the ConstructMap software to determine the different starting points for students. In addition, this allowed me to track students' understanding on the same scale from pretest to posttest.

## Findings

### Calibration

The items were calibrated using the ConstructMap software. As mentioned earlier, the pretest and posttest were calibrated together using a unidimensional partial credit Rasch model. This was done in order to capture a full range of student responses, as well as to determine that the Rasch model is still a good fit after changing the PMM progress variable, which also meant changes to scoring of the items.

The thresholds for all item responses to the pre/posttest are displayed in Figure 5.1. The histograms on the left side of the map are the student proficiencies and on the right side, the thresholds for each item are shown in columns. The number of each item is displayed on the x-axis. In general, there are more thresholds for the open-ended items (12-18) that are higher than those for the multiple-choice items (1-11), which are mostly dichotomous.

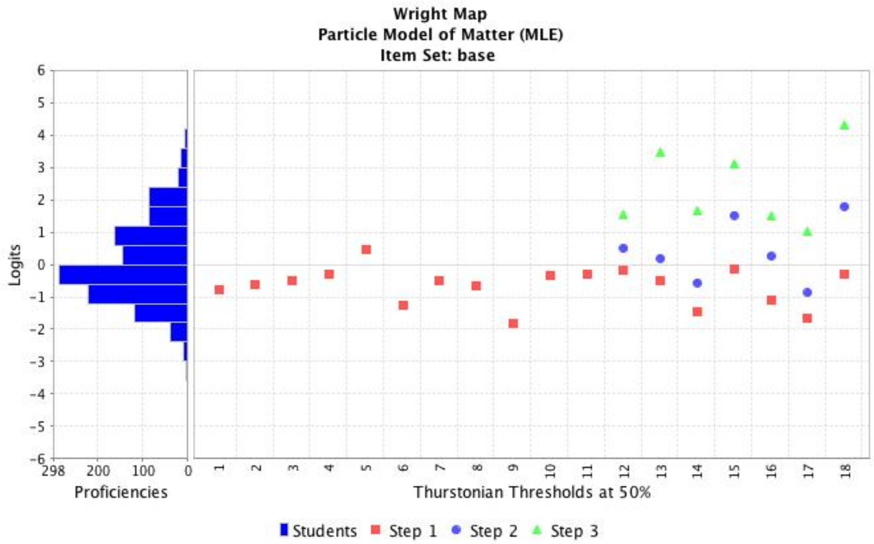



**Figure 5.1. Graphical Wright Map of Thurstonian thresholds for pre/posttest**

First, I investigated the technical aspects of the Wright map. The thresholds are ordered as expected. Overall, the items showed a good fit with average infit mean squares of 1.13. This indicates that overall, the items are behaving as expected. The infits for all the items range from 0.87 to 1.33. Thus, no items ranged outside of the acceptable range. Overall, there is a good respondent fit to the items, with an average infit of 0.93, which is slightly less than the ideal 1.00. Results also show that the model fits 87.7% of students.

I also examined the reliability of the items. The average standard error of measurement (SEM) was 0.49. The SEM of the items produced a “U”-shaped (see Figure 5.2) centered near zero logits. The SEM curve indicates that the items best measure students with abilities between -1.50 and +1.50 logits. In other words, the items provide more information about students with abilities between -1.50 and +1.50 logits.


**Figure 5.2. The SEM for the particle model of matter construct.**

The separation reliability of the items was calculated to be 0.85. Thus, the estimator of a respondent's location accounts for 0.85 of the variance. Thus, there are a wide range of students of different abilities responding to the items and the levels are much more distinct than those in the Calibration study.

In sum, the Rasch model is a good fit for the data. The item fit and respondent are appropriate for measuring the (PMM) progress variable. Moreover, the items are measuring student performance reliably.

To track students using the progress variable, I needed to determine the criterion zones associated with the corresponding levels. The Thurstonian thresholds computed for each step of an item were used to determine the criterion zones. To accomplish this, cut-points, which mark the boundaries of the levels, must be determined. The cut-point is the midpoint between the means of the Thurstonian thresholds of two levels. However, Thurstonian thresholds do not exist below the step one of an item, so there is no mean threshold value for the lowest category or a midpoint for the lowest two categories. For this study, I used the 67% confidence interval around the second category's mean to define the cut-point between the first and second categories. Dichotomous items only have one step, so their step difficulties are their Thurstonian threshold.

This process resulted in the determination of the criteria zones. The establishment of these zones provides the context for mapping student progress. Figure 5.2 shows the criteria zones mapped onto the graphical Wright Map of Thurstonian thresholds for the pre/posttest items. Most items measure students' knowledge at the basic level, and as expected, most item steps fall within the "Basic" criteria zone. Comparing the hypothesized levels for items (see Appendix A) with the criterion zone the associated item step is located indicates that most of the item steps associated with a particular level fall within the hypothesized level. The exception is the mixed level items, which range from the descriptive to basic level. This is partially due to the fact that most of the higher and lower ranging "Mixed" level responses (red square for items 12-18) are associated with the more cognitively demanding written response tasks. Moreover, most of the higher-level responses tend to be explanation items, and the lower level responses tend to be the drawing portion of items.

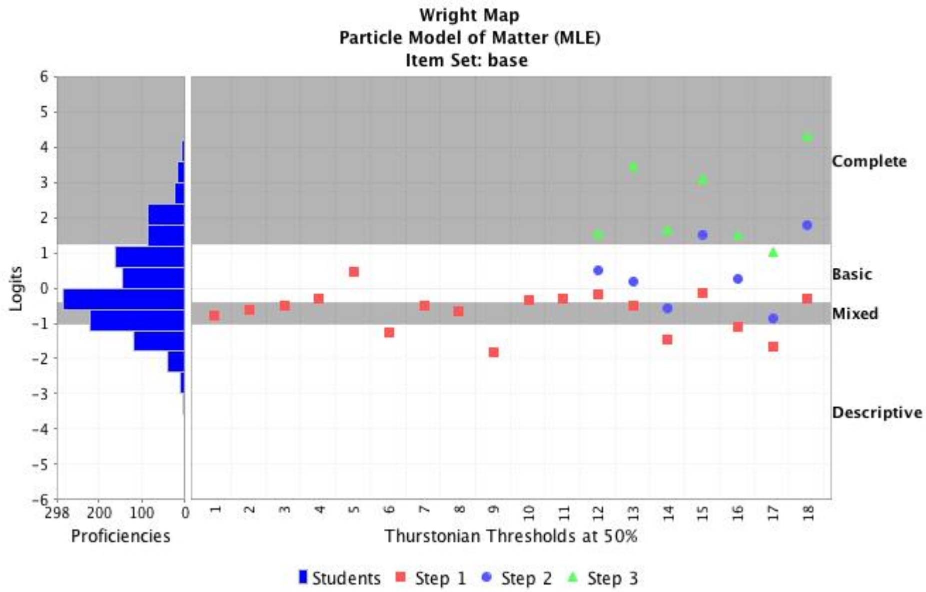



**Figure 5.3. Graphical Wright Map with criterion zones.**

The items were originally designed to have varying degrees of difficulty (see Chapter 3). Therefore, items that fall outside of their hypothesized level are demonstrating that each level has items of varying difficulties associated with them. As stated earlier, most items are designed to test knowledge at the "Basic" level. These results point to a need for more items at the other levels of the construct map, including the descriptive level. Moreover, additional items related to other levels of the construct map could potentially result in more accurate determination of the criteria zones.

The correlation between pretest and posttest scores was a moderate 0.51. This indicates that students' performance on the pretest, and their performance 8 weeks later after a unit of instruction, are unlikely to be strongly dependent, and thus treating them as not dependent for the purposes of estimation is probably reasonable. Further examination of this relationship using a subset of students, as well as its relationship to student performance during instruction, is examined in the section titled "Effect of students' initial understanding on performance during the unit."

**Student Pre- to Posttest Gains**

Estimates of the student proficiencies were calculated using the ConstructMap software. Calibration of the pretest, posttest, multiple choice, written response and models items were rerun, anchoring the item difficulties to those determined with the calibration. Therefore, the item difficulties remained the same whether I was analyzing pretest or posttest multiple-choice items.

Table 5.3 reports gains from pretest to posttest along with results of paired-sample t-tests. Students participating in the sixth grade chemistry curriculum experienced large and significant proficiency gains. On average, students' overall abilities increased a significant ($p < 0.001$) 1.73 points from pre- to posttest. Therefore, if a person with an average proficiency (0 logit) responded to an average difficulty item (0 logit) on the pretest, such that they had a 50% probability of getting the item right, then if that same person had a 1.73 logit change in proficiency at the posttest, this means they would now have a greater than 76% better chance of getting the item right. This gain also corresponds to an effect size of 2.28, meaning a change in proficiency equivalent to two and one-quarter standard deviations. Students' proficiencies also increased for the multiple-choice, written response and modeling items. Moreover, the modeling items had an effect size of 2.06 ($p< 0.001$) – a two standard deviation change in proficiency.

**Table 5.3. Sixth Grade Student Gains, (N=602)**

| **Items** (# items) | **Pretest Mean** (SD) | **Posttest Mean** (SD) | **Gain (SD)** | **Effect Size**[a] |
|---|---|---|---|---|
| **ALL (20)** | -0.83 (0.76) | 0.95 (0.99) | 1.36 (0.85) | 2.28*** |
| Multiple Choice (15) | -0.54 (0.76) | 0.98 (1.22) | 1.52 (1.12) | 1.88*** |
| Written Response (5) | -0.79 (0.98) | 1.01 (1.18) | 1.80 (1.18) | 1.87*** |
| Models (4) | -0.92 (1.00) | 1.13 (1.25) | 2.05 (1.31) | 2.08*** |

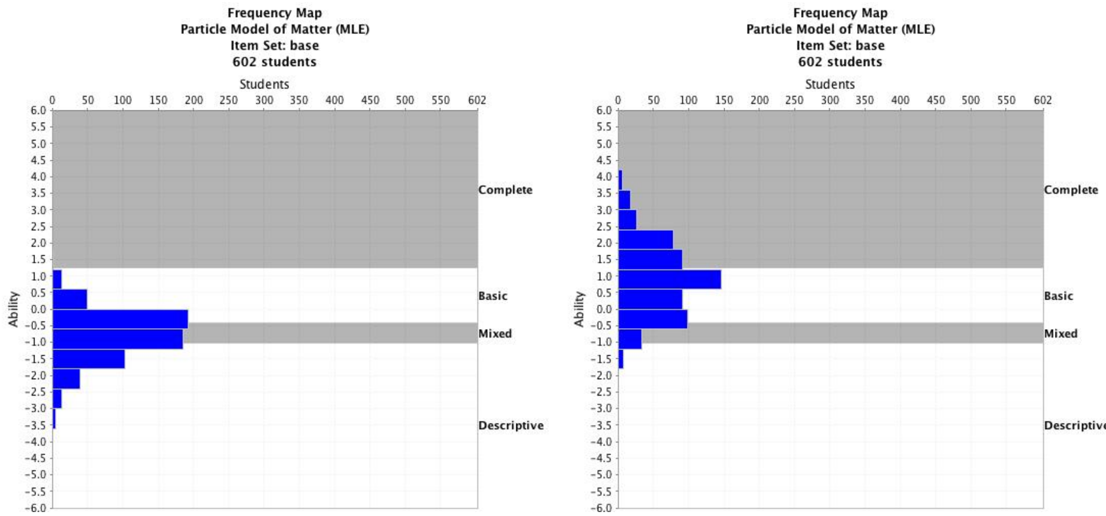
[a]Effect size calculated by dividing gain score by standard deviation (SD) of the pretest.
***p< 0.001

Table 5.4 displays the distribution of students at each level of the progress variable at the beginning and end of the unit. At the pretest, 67.8% of students are at the "Mixed" or "Descriptive" level. By the end of the unit, 90.4% of students have a particle view of matter, with most of these students (53.5%) at the "Basic" level. The table also shows that the average ability for each level has increased from pretest to posttest.

**Table 5.4. Distribution of student abilities from prettest to posttest.**

| | **Pretest** | | | **Posttest** | | |
|---|---|---|---|---|---|---|
| *Level* | *Number of Students* | *% of Students* | *Average Ability* | *Number of Students* | *% of Students* | *Average Ability* |
| Complete | 2 | 0.3 | 1.42 | 222 | 36.9 | 2.01 |
| Basic | 192 | 31.9 | -0.03 | 322 | 53.5 | 0.45 |
| Mixed | 187 | 31.1 | -0.75 | 42 | 7.0 | -0.73 |
| Descriptive | 221 | 36.7 | -1.69 | 16 | 2.7 | -1.40 |

Frequency maps are the left side of a Wright Map. Figure 5.4(a), the map on the left, shows the distribution of student abilities on the pretest, while Figure 5.4(b) shows the distribution of student abilities on the posttest (Figure 5.4). These figures are consistent with the findings in Table 5.3 that the mean improves from pretest to posttest. Moreover, there is a shift in distribution from the lower levels of the progress variable to higher levels.



(a) pretest (b) posttest

**Figure 5.4. Distribution of student abilities on pretest and posttest.**

## Summary

The development of criterion zones has allowed me to analyze students' progress at two times: before starting the unit and eight weeks later, after the end of the unit. I have found that most students from disparate school districts, cities, backgrounds, and geographical locations across the United States move from less sophisticated to more sophisticated views of matter. Thus, students progressed as anticipated from the beginning to the end of the curriculum unit. This shows that coherent curriculum and assessment can help students achieve a particle view of matter. Furthermore, students' progress can be followed through the use of the PMM progress variable. Student work from a smaller subset of these students was analyzed to further investigate how students' views of matter change as they experience the curriculum.

## Tracking Student Understanding During Instruction

The tracking study involves three teachers from 3 different schools in three different cities and 122 students (see Table 5.1). These students are not a representative sample of students in the study, but represent teachers who sent student work. Of those students, only students with work completed on all three embedded assessments (AS1.1, AS5.2, and AS15.2) were included in the study. Students' understanding of the particle model of matter was tracked using pretests, embedded assessments and posttests. Table 5.5 details the mean and sample variances of the student ability estimates for the particle model of matter (PMM) variable. This group of students starts out with a higher average ability estimate on the pretest than those in the overall study of student performance. The wide variance in student results for the embedded assessments indicate the many different models that students created for each assessments.

**Table 5.5. Means, and variances of person ability estimates for the PMM progress variable (N= 122).**

| Assessment | Mean (in logits) | Sample variance |
|---|---|---|
| Pretest | -0.54 | 0.80 |
| AS1.1 | 0.82 | 3.39 |
| AS5.2 | 1.03 | 2.75 |
| AS15.1 | 1.27 | 3.79 |
| Posttest | 1.08 | 0.62 |

Students performed consistently better from the pretest to posttest. Gains in students' proficiency estimates are reported in Table 5.6, as well as results of paired-sample t-tests. There were significant gains from the pretest to AS1.1. There were gains from AS1.1 assessment to the AS15.1 assessment, but they were not significant. There is a slight, but insignificant, drop in performance from AS15.1 to the posttest.

**Table 5.6. Student proficiency estimate gains from each successive assessment.**

| | Gain | p |
|---|---|---|
| Pre-AS1.1 | 1.36 | <0.001 |
| AS1.1-AS5.2 | 0.21 | 0.31 |
| AS5.2-AS15.1 | 0.24 | 0.20 |
| AS15.1-Post | -0.18 | 0.26 |

Another way of looking at students' progress from pretest to posttest is the Performance Map (Figure 5.5). The Performance Map shows students' ability estimates over time. A Performance Map can be generated for a single student, for an entire class, or entire groups of students. Figure 5.5 shows the average progress for all students from pretest to posttest. Overall, this indicates that student conception of matter improved during instruction. It also shows that, on average, students progress to a "Complete" model of matter during instruction.

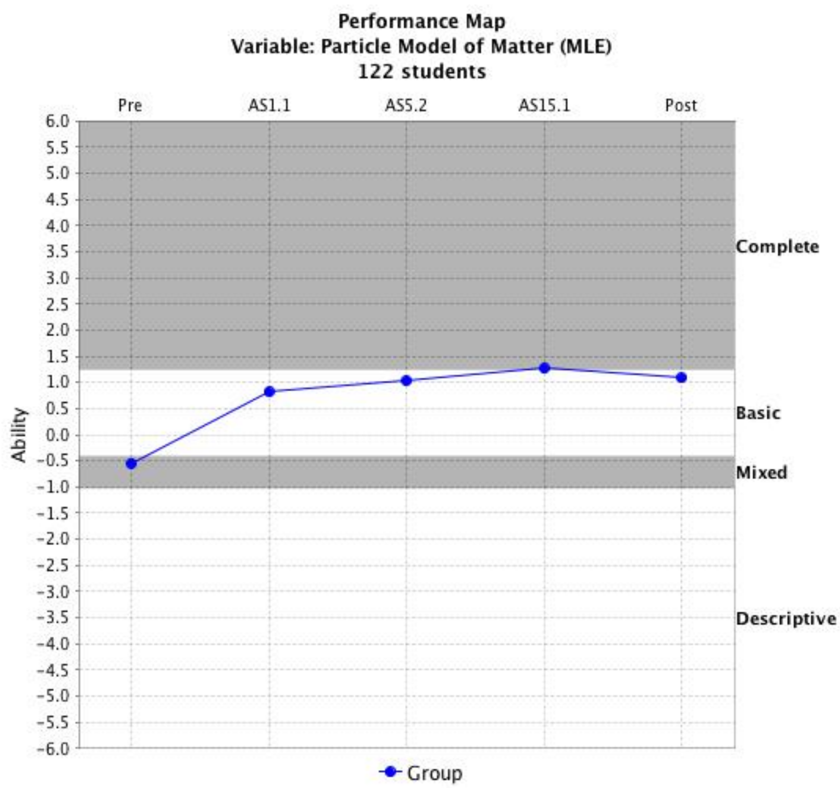
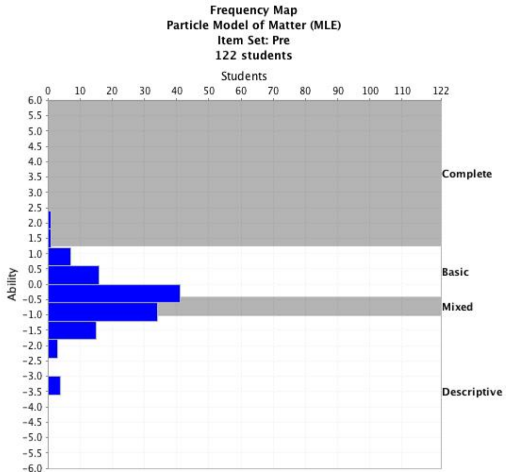
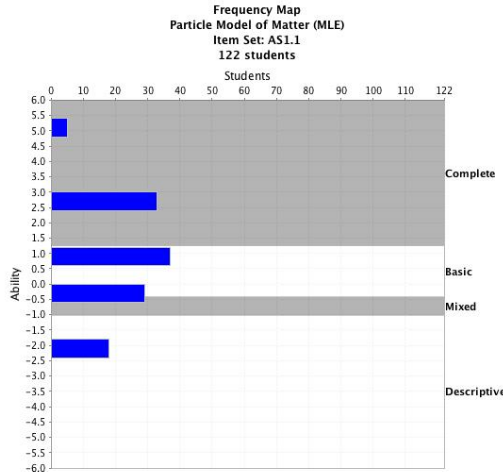



**Figure 5.5. Map of average student progress on the PMM progress variable.**

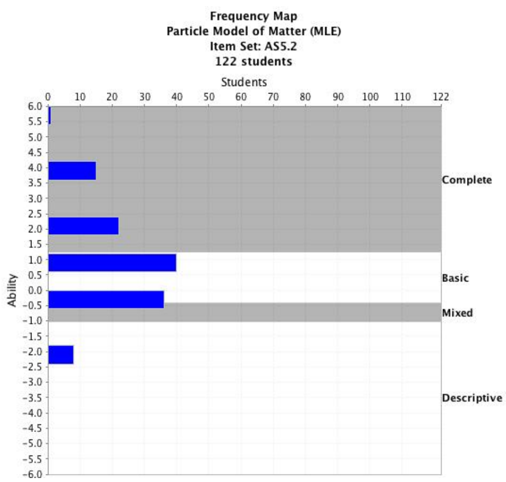
Yet another way to examine how students' views of matter change during instruction is through looking at the frequency maps of student abilities at each assessment point. Figure 5.6 displays these frequency maps from pretest to posttest. This is also consistent with the mean improvements as well as the Performance Map, showing that the distribution shifts towards higher levels of the progress variable. These displays also indicate that most students are beginning at either “Descriptive” or “Basic” levels at the pretest and move to “Basic” and “Complete” levels during instruction. They also show the varied responses that students provided to the embedded assessments, as was indicated by the variances found for these assessments shown in Table 5.6. At the posttest, students display a wide array of knowledge from the “Mixed” to “Complete” levels of the progress variable.



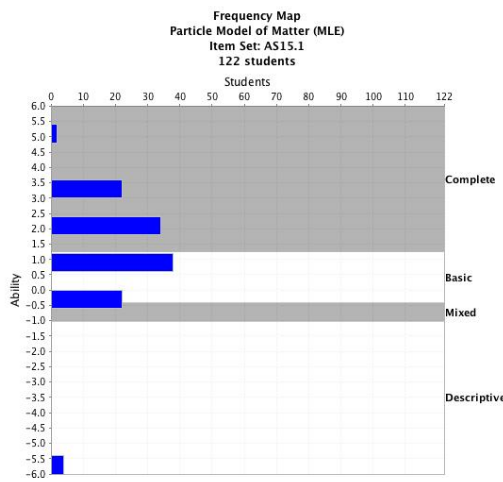

pretest



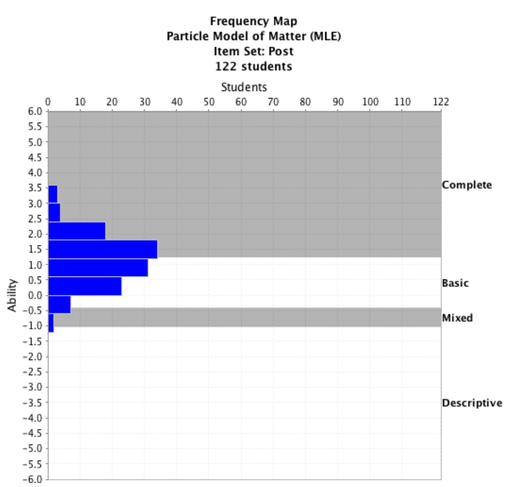

AS1.1




AS5.2




AS15.1




posttest

**Figure 5.6. Map of student (N=122) ability estimates from Pretest to Posttest.**

**Discussion**

Tracking students' understanding of the particle model of matter during instruction provides greater insight to students' development of the particle model. The embedded assessments (AS1.1, AS5.2, AS15.1) occur at different points during instruction. The first assessment, AS1.1 occurs at the beginning of the unit and results indicate that most students have a basic particle model of matter. Results indicate that students have a significant growth in performance from pretest to AS1.1. This could be due to the discussions students have prior to creating their models. During this discussion, students have the opportunity to talk about how they think odors are able to travel. This may also be due to the ability to talk with peers as they create their models.

In the three lessons that follow, students investigate matter, its different states and study gases by investigating the behaviors of gases. These investigations are then used as evidence of a particle model of matter. In lesson five, students are introduced to the idea that everything is made up of particles. They investigate the ability of an acid and a base to change the color of indicator paper without being dipped in the liquid. Through creation of models and discussions around the phenomenon and the models, students develop an understanding of evaporation and that the particles of the liquid are the same as those of the gas. Then, students are supposed to be introduced to a different model, a computer simulation to explain how smells travel across a room before constructing their own models of smell. Results seem to indicate that these preceding instructional strategies have helped students to further develop a particle view of matter.

Prior to assessing student views of matter in Lesson 15, students experience phenomena to help them understand properties and phase changes on a molecular level.

Although there are several instructional strategies that are supposed to occur between lesson 5 and lesson 15, it is difficult to pinpoint which of these have contributed most to student learning gains from lesson 5 until lesson 15. Since the unit was written with a particular sequence of learning performances and their associated learning activities, it can be postulated that this learning sequence helped students to develop a particle view of matter.

Student performance, on average, dropped between AS15.1 and the posttest. AS15.1 occurs before a class review of all the big concepts students have learned. During this review students create models of phenomena before coming to a class consensus model that can explain all the phenomena that they have reviewed. The lack of scaffolding prior to the modeling activities of the posttest may in some part explain the drop in student performance, although there could be other reasons for drop in student performance.

In sum, well-aligned curriculum and assessment can provide insight into instructional strategies and instructional sequencing that help students in developing a particle model of matter. As students experience the unit, results indicate that assessments can be used to track students' understanding during instruction. In addition, results show that students can develop a "complete" particle view of matter during instruction. Although tracking student performance has found that student performance improves during instruction, it does not indicate whether student performance on each of the assessments is related. The first sub-question of this study seeks to examine whether students initial understanding of the particle nature of matter influences their performance on the assessments that follow.

### Effect of students' initial understanding on performance during the unit

The PMM progress variable has allowed students' progress to be tracked from pretest to posttest. However, it does not provide insight into whether students' initial understanding of the particle nature of matter relates to their subsequent performance. This part of the study seeks to answer the sub-question: *What knowledge of the particle nature of matter do students bring to the unit and how does this relate to students progress students towards a particle model?*

Table 5.7 shows the number of students at each level of the progress variable at the pretest, using students' proficiency estimates in relation to the established criterion zones. It was found that a majority of the students being tracked (51.6%) are at a basic model level. This is also reflected in the pretest frequency map in Figure 5.6, which shows the distribution of students is mostly at the lower portion of the Basic model level. These results indicate that most students are at the beginning of developing a basic particle model. In addition, both the pretest frequency map in Figure 5.6 and the student pretest performance found in Table 5.7 indicate that students are beginning at all levels of the progress variable.

**Table 5.7. Student pretest performance (N =122).**

| Level | Number of Students | % of Students |
|---|---|---|
| Complete Model | 2 | 1.6 |
| Basic Model | 63 | 51.6 |
| Mixed Model | 25 | 20.5 |
| Descriptive Model | 32 | 26.2 |

To determine how students' beginning knowledge of matter influences their development of a particle model of matter, I first needed to see if there was a correlation between students pretest proficiency estimate and their proficiency estimates at each of the embedded assessments and posttest. Table 5.8 displays the correlations and significance of those correlations of each assessment. Results indicate that there is a weak, but significant, correlation between student ability at the pretest and each of the successive embedded assessments. Students' proficiency estimates on the pretest are moderately correlated ($p<0.001$) with students' proficiency estimates on the posttest. Although this is less correlated than the overall pretest to posttest performance for all students (N=602), it still indicates that student posttest performance is moderately dependent on their pretest performance after 8-12 weeks of instruction for this group of students. Students' proficiency estimates between embedded assessments AS15.1 and the posttest ($p<0.001$) are also moderately correlated.

**Table 5.8. Correlation of students' pretest ability with later assessments (n =122)**

| | Pre | AS1.1 | AS5.2 | AS15.1 | Post |
|---|---|---|---|---|---|
| Pre | - | 0.333** | 0.222** | 0.221** | 0.437*** |
| AS1.1 | 0.333*** | - | 0.122 | 0.176 | 0.083 |
| AS5.2 | 0.222** | 0.122 | - | 0.361*** | 0.281*** |
| AS15.1 | 0.221** | 0.176 | 0.361*** | - | 0.399*** |
| Post | 0.437*** | 0.083 | 0.281*** | 0.399*** | - |

***p< 0.001, **p< 0.05

Student performance on AS1.1 is moderately and significantly correlated to their performance on the pretest. A weak and insignificant correlation exists between student performance on AS1.1 and the assessments that follow. This further indicates that the instruction that occurs in relation to this modeling activity has an influence on student performance.

The embedded assessment for lesson 5.2 has a moderate and significant correlation to student performance on AS15.1 assessments and weak, but significant, correlation between AS5.2 and the posttest. As mentioned earlier, AS5.2 occurs after students have learned the basic parts of the particle model: all matter is made up of particles, there is empty space between the particles, and the particles move. This indicates that the instruction that follows the learning of these particle ideas are related to student learning and performance on the final two assessments.

In sum, student pretest performance is moderately related to student posttest performance. Student performance on the posttest also related to student performance on embedded assessments AS5.2 and AS15.1. Results indicate that instructional strategies that occur between each of the assessments, especially between the pretest and AS1.1 as well as between AS5.2, AS15.1 and the posttest influence student performance. In sum, the design of the curriculum materials assists students in developing a particle view of matter.

To further examine these relationships, I performed a General Linear Model (GLM) repeated measures. Through this procedure, I was able to model students' assessment during the unit multiple times using analysis of variance. A test of sphericity was conducted to determine whether a univariate or multivariate model was most appropriate for modeling the data. Sphericity seeks to determine whether the data is correlated. In this case, the multivariate model was used because there was a violation of sphericity ($p<.000$).

Results indicate that there is an insignificant relationship between the pretest and successive assessments ($p=0.332$). There is a main effect of the successive assessments ($F= 66.052$, $p<0.001$), which indicates students are becoming more proficient during instruction. There is also a main effect of student ability on the pretest ($F=6.628$, $p< 0.001$). The Wilks Lambda multivariate test, which examines the overall differences among assessments, indicated was none of the main interaction effects of the assessment are significant ($p =0.52$). This further indicates that student pretest performance is moderately related to their performance on subsequent assessments and that instruction is having the greater impact on performance.

## Summary

The first sub-question of this study was: *What knowledge of the particle nature of matter do students bring to the unit and how does this relate to students progress students towards a particle model?* Results show that students began the unit at different levels of the progress variable, including all levels of the PMM progress variable. This indicates that students enter the unit with varying knowledge of the particle model of matter.

Students' pretest performance is weakly correlated to their performance on successive assessments. Moreover, students' performance on the pretest is moderately correlated with their performance on the posttest. This indicates that the knowledge students begin the unit has a weak influence on how they perform on the rest of the unit, and may be a moderate indicator of how students will perform on the posttest.

A moderate relationship also exists between student performance on the AS5.2 and AS15.1 assessment and performance posttest. This indicates that the design of the instructional sequence of the unit influences student performance on the posttest.

Further analysis determined that there is an insignificant relationship between the pretest and students' performance on the embedded assessments throughout the unit. This indicates that instruction and the assessment of student understanding during instruction has an important impact on student performance.

Thus far, I have been able to track students' understanding during instruction. I was also able to describe how students' concept of matter improved from pretest to posttest. As discussed during the description of the unit in Chapter 2, student models are composed of two parts – the drawing and the explanation. These two parts of the model have been treated as separate items in this study. Thus, I wanted to take further examine the models students construct to determine the relationship between the two components of student models: the drawing and the explanation during instruction.

**Assessing students' models of matter**

This part of the study answers the second sub-question: *What is the relationship between students' drawings and explanations of phenomena?* The scientific practice of modeling is an important feature of the sixth grade chemistry unit. As aforementioned, student models are comprised of two parts: the drawing and the explanation.

The pre/posttests and embedded assessments were used to track the relationship between the drawings and explanations. The embedded assessments utilized in this study involve students explaining how they are able to smell an odor from across the room (see Appendix E). There are two written response items on the pre/posttest that could be used to assess the relationship. I chose the first of the written response items (see Appendix A) because it is proximal to the embedded assessments. This item deals with students explaining in which room, a cold room or a warm room, they would smell an odor from an air freshener faster.

First, I investigated how each drawing and explanation was coded from pretest to posttest. Table 5.9 presents the results of this analysis. For each assessment, there is a drawing column represented by the letter "D" and an explanation column represented by the letter "E." For each assessment, the percentage of student responses at each level of the construct map for each part of the model is listed. For example, at the pretest, 21.3%

of students are at a “Mixed” level for their drawing and 24.6% of students have explanations coded at the “Descriptive” level.

**Table 5.9. Types of models students created.**

| | Pretest (%) | | AS1.1 (%) | | AS5.2 (%) | | AS15.1 (%) | | Posttest (%) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **D** | **E** | **D** | **E** | **D** | **E** | **D** | **E** | **D** | **E** |
| Complete | 0.8 | - | - | - | 37.7 | 0.8 | 59.8 | 2.5 | 49.2 | 11.5 |
| Basic | 6.6 | 12.3 | 48.4 | 4.9 | 54.1 | 18.9 | 33.6 | 23.8 | 19.7 | 63.9 |
| Mixed | 21.3 | 24.6 | 27.0 | 48.4 | 7.4 | 33.6 | 6.6 | 36.1 | 20.5 | 13.1 |
| Descriptive | 71.3 | 59.0 | 24.6 | 46.7 | 0.8 | 46.7 | - | 37.7 | 10.7 | 11.5 |

As previous results have shown, students produce more sophisticated models (drawing (D) + explanation (E)) as they progress through the unit. The data also shows that students are progressing faster with their models than with their explanations. Before starting the unit, 71.3% of students are at a “Descriptive” level for their drawing with 59% of students also explaining the phenomena on a descriptive level. By AS5.2, most students (54.1%) are at a “Basic” level for their drawing and 46.7% still have a “Descriptive” level of explanation. By the end of the unit, 49.2% of students are at a “Complete” level for their drawing and 63.9% are at a ”Basic” level for their explanation.

The relationship between student drawings and explanations was investigated by coding them into the three categories shown in Table 5.10: drawing more sophisticated, same level and explanation more sophisticated. These categories were designed to capture the different relationships that could exist between students’ drawings and explanations. For example, when students’ drawings are coded at the same level as their explanation, then the relationship is coded zero, indicating no difference between the two parts of the model.

**Table 5.10. Coding of difference in students models versus explanations.**

| Code | Category |
|---|---|
| 1 | Drawing is more sophisticated than explanation. |
| 0 | Drawing and explanation on the same level |
| -1 | Explanation is more sophisticated than drawing |

Figure 5.7 illustrates this relationship by displaying the distribution of these relationships during instruction. At the pretest, 52.5% of students' explanations were at the same level as their drawing. By Lesson 5 (AS5.2), almost all students (91.8%) have drawings that are at a higher level of the scoring guide than their explanation. By the end of the unit (posttest), this percentage has decreased as 51.6% of students' drawings are at a higher level of the scoring guide than their explanation. To further describe the relationship between student drawings and models, one student's pathway is discussed.

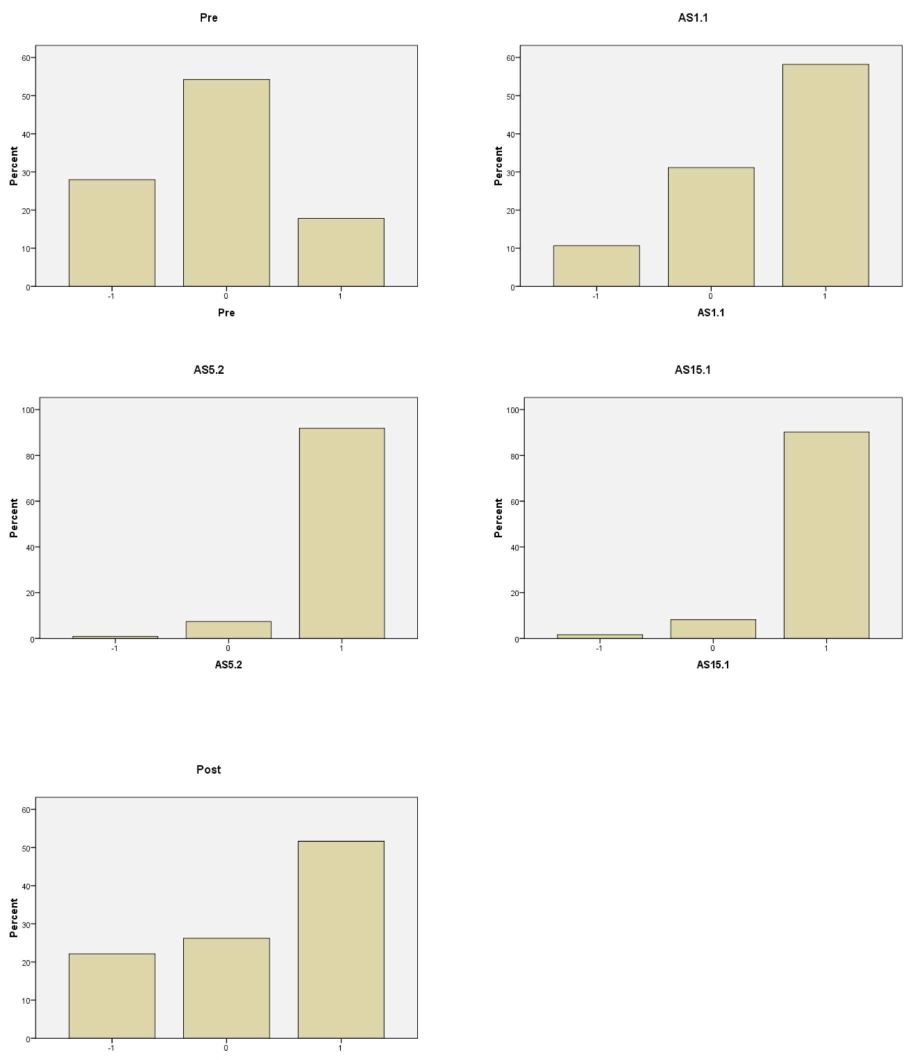
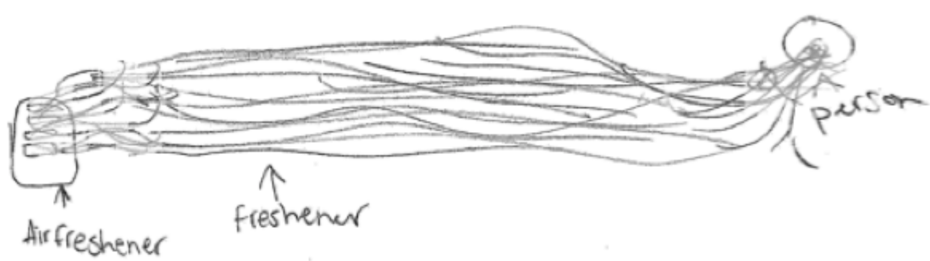



**Figure 5.7. Distribution of model versus explanation difference (n=122).**

**The case of Sarah**

Sarah is a student in one of the schools in the Midwest. Sarah was chosen because her drawing improved much faster than her explanation. To illustrate this change over time, three of the assessments were chosen (pretest item, AS5.2, and AS15.1).

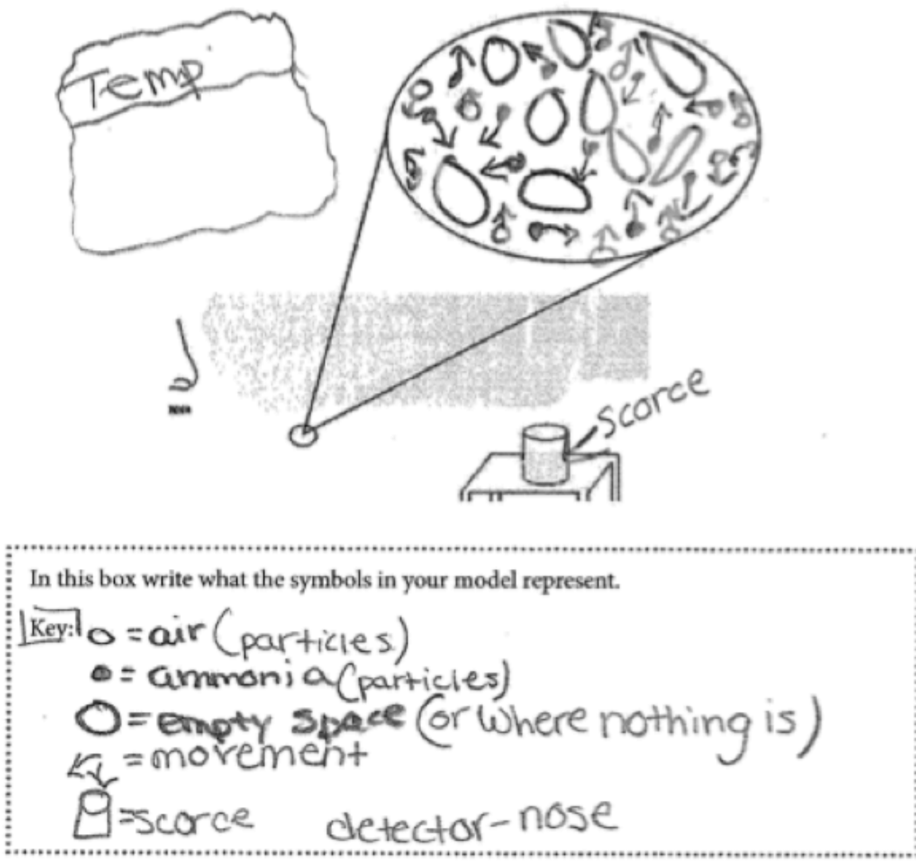
Figure 5.8 displays Sarah's pretest model. Sarah's drawing and explanation were both coded to be at the "Descriptive" level. This model is representative of the zero, where the drawing and explanation are coded at the same level.




Explanation: *They move faster because the room is lose and warm. Unlike when the room is cold it is stiff and you and the smell don't want to move.*

**Figure 5.8. Sarah's pretest model.**

Figure 5.9 displays Sarah's model for AS5.2. Sarah's model has increased in sophistication from the pretest and was coded at a "Complete" level. Her explanation has also improved, going from a "Descriptive" level at the pretest to "Mixed" level at this assessment. Thus, her drawing has improved much more than her explanation.



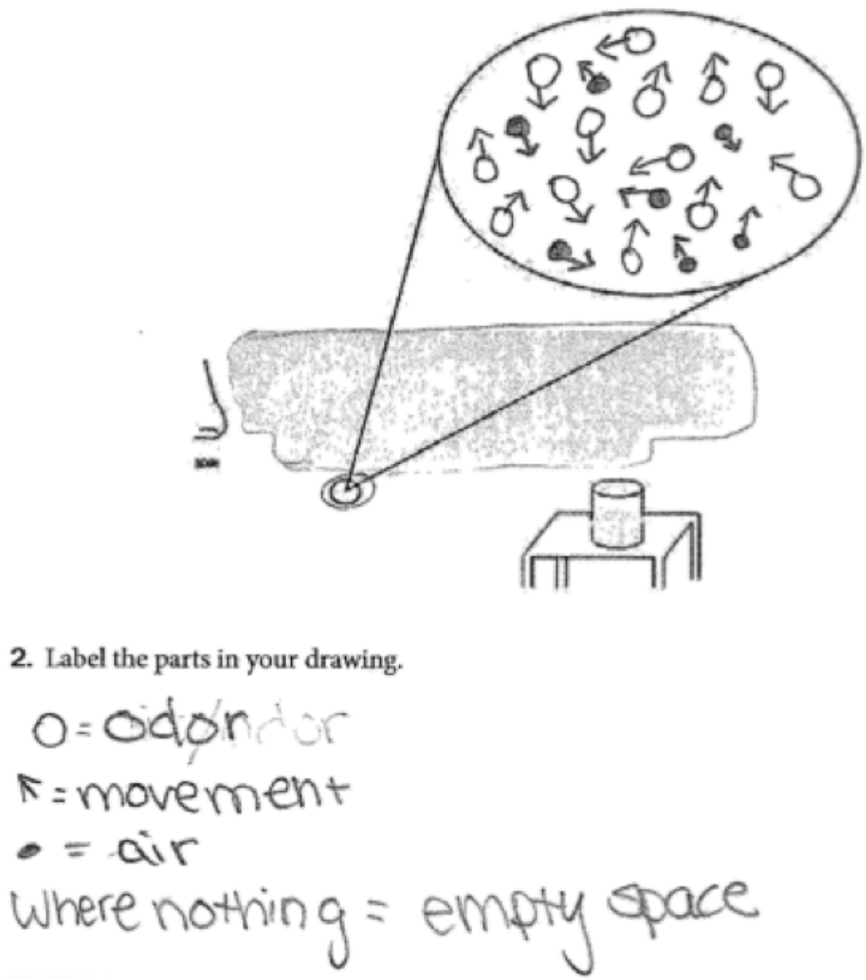

Explanation: *The odor and air mix and have empty space and move to our nose so we can smell them.*

**Figure 5.9. Sarah's AS5.2 model.**

As mentioned earlier, this embedded assessment occurs after students have watched a simulation of odor particles mixing with air particles in a room. The simulation also has the ability to show temperature affect on particle movement. However, this feature is not supposed to be used in this lesson. This might explain the inclusion of an objected labeled "Temp" in her drawing, though it is not referenced in the key for her drawing or in the explanation.

Finally, Figure 5.10 display's Sarah's model for AS15.1 which occurs near the end of the end of the unit. Sarah's drawing was again coded at the "Complete" level.

Although her explanation has now gone from the “Mixed” to “Basic” level, it still lags behind the improvement of her model.




Explanation: *The particle move around and collide.*

**Figure 5.10. Sarah’s model for AS15.1**

## Discussion

The second sub-question of this study was: *What is the relationship between students’ drawings and explanations of phenomena?* Students’ models are composed of their drawing and explanation. Results indicate that students’ explanations lag behind students’ drawings. As students progress through the unit, the drawing portion of the models progress faster than their explanations of the same phenomena. This is also reflected in that the drawing items (items 12, 14 and 17) tend to be less difficult than those of the explanation items (items 13, 15 and 18) as shown in Figure 5.2.

This is not totally unexpected as the instructional sequence focuses on students' development of a particle model of matter during the first learning set, especially their drawings, which culminates with embedded assessment AS5.2. Earlier results also show that instruction related to the drawing portions of the models help students to develop a particle view of matter. This does not mean that students' explanations are not a focus during instruction, but it does indicate that students need further scaffolding to help them to develop more sophisticated explanations. Thus, the unit could be refined to address this issue. What cannot be taken into account with this analysis is whether students are provided enough time to create their models during instruction, what types of discussions are had in relation to these modeling assessments, or other aspects of instruction.

## Study Limitations

There are several limitations to this study. It is expected that there could be teacher and school effects related to students' performance. I did not collect videos of instruction or interview the teachers about their experience with the unit. This information could provide more insight into differences in student performance as they experience the unit as well as their performance on the posttest. Although this information was not collected, it does not affect the ability to track students understanding as they experience the unit.

Second, the embedded assessments consist of only two questions that are identical. Although having only two items does not limit the ability to determine estimation parameters for these items, having more items would provide better estimates of the construct.

Thirdly, student posttest performance is moderately related to their pretest performance. Although this did not hinder my ability to estimate student proficiencies, it could have an impact on the accuracy of these estimates. If I had more time between the calibration study and the beginning of the unit, I would have done a separate calibration of the test items, including one of the embedded assessments.

Finally, all the teachers that participated in the study were supposed to return the embedded assessments as well as the tests. Despite constant communications and reminders to send these items, many teachers did not send complete data. As a result, only three teachers returned complete materials. Although this was less than ideal, it did not inhibit my ability to track students understanding in three different locations.

## Chapter Summary

In this chapter, I presented and discussed the findings of the tracking study. This study shows that an empirically validated progress variable for the particle model of matter can be used to track students' understanding during instruction for students from diverse backgrounds. Moreover, students' understanding of the particle model improved from pretest to posttest.

The alignment of items with particular performance levels needs to be improved. The raw data for interpreting the models shows more students at the lowest levels of the progress variables at the pretest and lesson 1.1 than those of the model estimates at the same points. This may be partially due to the fact that there are few items at the lowest two levels of the progress variable on both the pretest and posttest as well as low number of items on the embedded assessments. This will also necessitate that there be an improvement in the accuracy of the cut-points. However, results show that interpretations of items developed through scoring guides and design of outcome space can be applied to actual student data as they participate in the unit.

Students' posttest knowledge is moderately related to their performance on the pretest. However, results suggest that instruction is impacting students' developing knowledge and is also reflected in students' gains from pretest to posttest as well as their performance at each of the embedded assessment results.

Moreover, students' explanations of phenomena lag behind their drawings of matter during instruction. The emphasis of the sixth grade chemistry is for students to create models of phenomena, which include both their drawing and explanation. Results show that most students begin the unit with a match between their drawing and explanation. Although more students' explanation of a phenomenon matched their drawing by the end of the unit, a majority of students' have drawings that are more sophisticated than their explanations. Instruction around students' models would need to be further explored to fully understand why this is happening, such as whether students are given enough time to develop both the drawing and written portions of their models.

# CHAPTER 6

## Conclusion

### Introduction

The No Child Left Behind Act of 2001 (NCLB) has placed a greater emphasis on the use of assessments to track students' education progress as well as serving as an accountability measure of schools. These proximal assessments are based on out-dated learning theories and do not take into account modern learning theories or research on students' misconceptions (NRC, 2001). Often, these assessments do not provide feedback to teachers or students about where students are having difficulties. Simultaneously, urrent science education reform has focused on how to help all students develop scientific literacy. Both the *Nationals Science Education Standards* (NRC, 1996) and *Project 2061: Science for all Americans* (AAAS, 1989) have served as the basis for assessing students' understanding of the big ideas of science. However, they are broad statements about what students should know over large grade bands.

Learning progressions and progress variables have been proposed as a means to address the need for curriculum and assessments that can help teachers' improve their practice as well as to inform both students and teachers about students' performance. This study provides evidence that curriculum and assessment based on modern learning theories, can lead to the development of progress variables that are able to track middle school students' understanding of the particle nature of matter over time.

The notion of learning progressions is not new, as several studies have suggested progressions in students' understanding of the particle nature of matter (Renstrom, Andersson, & Marton, 1990; Johnson, 1998; Nakhleh, Samarapungavan & Saglam, 2005; Liu & Lesniak, 2006; Margel, Eylon & Scherz, 2007). The first learning progression hypothesized was by Smith et al. (2006) about understanding atomic molecular theory for K-8 students. Kennedy, Brown, Draney and Wilson (2006) were the first to empirically validate a progress variable tied to a curriculum. Since then, there have been a few studies to validate progressions, including within the context of curriculum and curriculum development (Alonzo & Steedle, 2008; Claesgens, Scalise, Wilson & Stacey, 2008; Songer, Kelcey, & Gotwals, 2009). The Berkeley Evaluation and Assessment Research (BEAR) Assessment System (BAS; Wilson, 2005; Wilson & Sloane, 2000) has been proposed as one method for linking assessments to learning progressions through the use of progress variables (Wilson, 2009).

The aim of this study was to answer the research question: **How does middle school students' understanding of the particle nature of matter change during enactment of a model-based unit?** This chapter summarizes the results of the two studies that were conducted and how these results add to the research literature. I then discuss implications for curriculum developers as well as directions for future research.

## Validation and Application of a Progress Variable

Both *Taking Science to School* (Duschl, Schweingruber & Shouse, 2007) and *The Learning Progressions in Science: An Evidence-based Approach to Reform* (Corcoran, Mosher & Rogat, 2009) expressed the need to build coherent curriculum and assessment systems. The major finding of this study is that the particle model of matter (PMM) progress variable, which is linked to coherent curriculum and assessments can be used to track student understanding.

The PPM progress variable was specifically developed to track student progress during instruction of the “How can I smell things from a distance?” IQWST curriculum unit. The development of the PMM progress variable was based on the research literature and empirical data from previous studies of students’ understanding during the unit (Merritt et al., 2007; Merritt et al., 2008). This also meant that the framework for learning and development that served as a basis for developing instructional materials also served as a basis for item development.

The first study was conducted to empirically validate the PMM progress variable. Results showed that the model was a good fit. The Wright map shows that the items spanned the entire construct and that more difficulty items are associated with higher proficiency estimates. Analysis of fit determined that the items were a good fit and that the Rasch model was consistent for measuring the progress of 86% of students.

The calibration study also identified that the items are reliable. However, there were issues related to reliability. The separation reliability was not as high as expected. This indicated that there was not a wide enough sample of students of varying ability and that the construct map was comprised of too many levels.

Determination of criterion zones mapped the Wright map to the construct map, which served to identify that the “Mixed” and “Incomplete” levels of the construct map were not distinct. Elimination of the “Incomplete” level resulted in more distinct levels. Thus, the calibration study was vital in establishing that the items were good measures of students understanding of the particle nature of matter. In addition, the evaluation of the reliability and validity of the items helped to determine that the PMM progress variable needed to be modified.

Findings from the tracking study showed that mapping the items onto the modified progress variable resulted in a highly reliable instrument. Moreover, the establishment of criterion zones showed that the levels are now more distinct. This study showed that students from varying backgrounds and regions of the United States experienced significant gains in proficiency estimates from pretest to posttest.

Empirically validating the PMM progress variable also allowed me to track students' understanding during instruction, especially their models of matter. Prior studies have indicated the need to explore students' hybrid models of matter (Renstrom et al., 1990; Johnson, 1998; Nakleh et al., 2005; Liu & Lesniak, 2006; Margel, Eylon, & Scherz, 2008). A smaller subset of students (n=122) was tracked from pretest to posttest through the use of embedded assessments. Results indicate students performed consistently better from pretest to the last lesson of the unit. Student performances on the embedded assessments indicate that students make significant gains in their proficiency as they experience the unit, achieving higher levels of proficiency. Although average student performance dips to the "Basic" level on the posttest, the results show that students are able to develop a particle view of matter by the end of the unit. Thus, the development of this progress variable provided the opportunity to identify the range of models students created prior to and during instruction. Results show that the instructional strategies of the sixth grade IQWST chemistry unit to help students develop a particle view of matter.

The first sub-question of the study sought answer the question: *What knowledge of the particle nature of matter do students bring and how does this relate to the progress students make during instruction?* The Rasch model indicated that a majority of the students involved in this part of the study began the unit with a “Mixed” model of matter. In addition, students’ pretest performance is weakly correlated to their performance on successive assessments. A moderate relationship exists between student performance on the AS5.2 and AS15.1 assessment and performance posttest. However, students’ performance on the pretest is moderately correlated with their performance on the posttest. This indicates that students’ prior knowledge has a weak influence on how they perform during instruction. Although student prior knowledge can be a moderate indicator of how students will perform on the posttest, the instructional sequence of the unit influences student performance on the posttest. Therefore, both instruction and assessment of student understanding during instruction have an important impact on student performance.

The second sub-question of this study sought to answer the question: *What is the relationship between students’ drawings and explanations of phenomena?* Further analysis of students’ models showed that students’ explanation of phenomena lagged behind their drawings of the phenomena during instruction. Gotwals (2006) found that students might develop content and reasoning skills differentially. These results suggest that student ability to reason about a phenomenon through their written explanations lags behind their ability to create drawings to explain that same phenomenon.

Studies have suggested that students understanding of the particle nature of matter would improve if a sequenced, developmental approach was taken to supporting students understanding of particle theory. This study demonstrates that a progress variable developed for a curriculum unit designed to be coherent can be used to track student progress towards a particle model of matter during instruction. This study focused on the validation of the PMM progress variable, which was developed in relationship to a particular curriculum intervention. This meant that the framework for the development of the instructional materials matched that of the development of the assessment items and that the PMM progress variable encapsulates the aims of instruction (Wilson, 2009). The validation of the PMM progress variable provided a common basis for tracking students' understanding during instruction. The validation process also substantiates that curriculum and assessment are aligned. I conducted a calibration study to evaluate the reliability and validity of the progress variable. This evaluation provided the information necessary to determine that the instrument is a valid measure of student understanding of aspects of the particle nature of matter, as outlined by the curriculum and detailed in the construct map.

A second study was conducted to utilize the modified progress variable for tracking student progress during instruction. The calibration of the modified progress variable demonstrated that the assessments are able to measure understanding of the particle model of matter when respondents are from diverse backgrounds and geographical locations. Moreover, the ability to track student progress during instruction demonstrates that the sequencing of the learning performances helps students in developing a particle view of matter. In the next section, I discuss the implications of these results.

## Implications

### Curriculum developers

The particle nature of matter is a big idea of science (Smith et al., 2006; Stevens, Delgado, & Krajcik, 2009). It is also the foundation for understanding a myriad of science concepts including properties, phase change, and chemical reactions. Previous interview studies (Novick & Nussbaum, 1978; Stavy, 1991; Nakhleh, Samarapungavan & Saglam, 2005) have outlined the difficulties students have with understanding particle theory and its related concepts. Current large-scale tests that assess students' knowledge of particle theory and its related concepts do not provide information to teachers or students that would help to improve teaching or learning. In addition, the large-scale tests tend to ask questions that require rote memorization of facts.

The development of progress variables is an opportunity to improve teaching and learning by providing feedback to students' and teachers (Corcoran, Mosher & Rogat, 2009). Scoring guides were developed for this study that aligned with the construct map. As the construct map also reflects the learning goals of the curriculum, it also points out the importance of embedded assessments tied to the learning goals of a curriculum (Kennedy et al., 2006). Therefore, the validated PMM progress variable could now be used to track students understanding during instruction. Teachers could then use the PMM progress variable to track students during instruction and provide feedback to students.

These results can also be used to identify improvements to IQWST, such as necessary changes to the curriculum to improve student performance. Analysis of student models indicated that students' development of explanation skills lagged behind their drawing skills. This points to a need for greater emphasis in both the curriculum unit and in teacher professional development on helping students to develop a more scientifically accurate explanation of phenomena.

Many studies proposed the development of curriculum materials that focus on students' development of a particle model of matter and the application of that model to explain macroscopic phenomena to help them understand the particle nature of matter (Ben-Zvi, Eylon & Silberstein, 1986; Kozma, Chin & Marx, 2000; Justi & Gilbert, 2002; Harrison & Treagust, 2002; Snir, Smith & Raz, 2003). The sixth grade chemistry IQWST unit, entitled "How can I smell things from a distance?" has demonstrated that this approach can help students to develop a "Basic" particle model of matter. Therefore, curriculum developers should create materials that focus on students developing and applying their models of matter to understand phenomena.

In sum, the validation of the PMM progress variable has many implications for curriculum development. First, the progress variable can now be used in conjunction with the $6^{th}$ grade IQWST unit to provide feedback to both teachers and students about student progress during instruction. Second, the unit needs improvements to help students develop more sophisticated explanations of phenomena. Finally, the unit demonstrates that students' development and application of their own models of matter can help them to develop more sophisticated views of matter. Therefore, curriculum developers should include opportunities for students to develop and apply their models of matter to explain phenomena. In the next section, I discuss future areas of research.

## Future Research

This study focuses on the sixth grade chemistry unit of the IQWST curriculum entitled "How can I smell things from a distance?" Teachers were provided professional development for the unit, although not all teachers attended. One limitation of this study is that it did not include observations of instruction. As *Knowing What Students Know* (NRC, 2001) points out,

> Ideally, an assessment should measure what students are actually being taught, and what is actually being taught should parallel the curriculum one wants students to learn (p.52).

The linking of instruction to assessment is vitally important to obtaining a complete picture of how closely teachers are following the curriculum, what modifications do they make to the curriculum and how are they utilizing embedded assessments to inform their practice. This includes whether they use scoring guides provided to them to inform instruction as well as for providing feedback to students about their progress. Therefore, classroom observation of teachers who are trained on using the scoring could provide more insight into the impact of assessment of student learning. Moreover, observations could help to identify the learning activities and teaching strategies that both help and hinder student development of particle notions.

A second limitation of this study is that only a small subset of students was tracked from pretest to posttest because of a lack of student work. A broader range of students could provide more information about student performance during instruction. Moreover, interviews of the students related to these assessments could provide information on the strengths and weaknesses of the assessments.

Finally, three identical embedded assessments were used to track students' understanding for this study. The curriculum includes several other embedded assessments in which students create models of particular phenomenon. The inclusion of additional embedded assessments could be used to identify particular learning activities that help students to progress. Using these other assessments, especially those related to states of matter and phase change could provide broader insight into students' knowledge of the particle model of matter. These modeling activities occur between AS5.2 and AS15.1 and could also help to identify learning activities that are helping student progress. They could also potentially identify where students need additional instruction.

Learning progressions are still in their early stages of development (Duncan & Hmelo-Silver, 2009). Their process of development includes many stages, from their development as a hypothetical learning progression based on what is known from research about student learning in a particular domain, or from cross-sectional studies of students' understanding across multiple grade levels (Schwarz et al., 2009, Songer et al., 2009; Claesgens et al., 2010; Alonzo and Steedle, 2008; Kennedy et al. 2006). The PMM progress variable could represent only one level of a larger progression. The findings from this research could be used as evidence for a larger progression (Shin, Stevens, Short, & Krajcik, 2009). The PMM progress variable focuses on student understanding of the particle model of matter and its application for understanding phase changes and properties. But how would a progress variable developed in relationship to the seventh grade IQWST unit, which focuses on the use of particle theory to explain chemical reactions and conservation of matter, be built upon the PMM progress variable?

Future research needs to be conducted on units developed with the same goals for instruction as the PMM progress variable. Moreover, this research needs to expand to the development of other curriculum units spanning kindergarten to twelfth grade to determine how students understanding progress during instruction. Depending on the grade level foci, these units could potentially range from four weeks to a year. This research could further examine the different types of hybrid models students create and whether these models mirror the range found during this study.

## Chapter Summary

If assessments are to be used to both track student progress and measure accountability, then they also need to provide feedback to teachers and students about that progress (NRC, 2001). The development of progress variables using the BAS is one approach that could meet both needs. The PMM progress variable was developed for the purpose of tracking student progress during instruction. The coherency of the curriculum and assessments provided the opportunity to align the PMM progress variable with the goals of instruction. This study shows that curriculum and assessment based on modern learning theories, can lead to the development and utilization of progress variables that are able to track students development of a particle model of matter over time.

# APPENDICES

# Investigating and Questioning our World through Science and Technology

## *How can I smell from a distance?*

**First and Last Name: ______________________________________**

**Date: ____________**

**Teacher Name: ______________________________**

**Class Hour: ______________________________**

## Part 1 -Multiple Choice

This test is an opportunity for you to show what you understand about chemistry concepts. Please try your best even if you are unsure of your answers.

**Please use a pencil to answer the questions.**

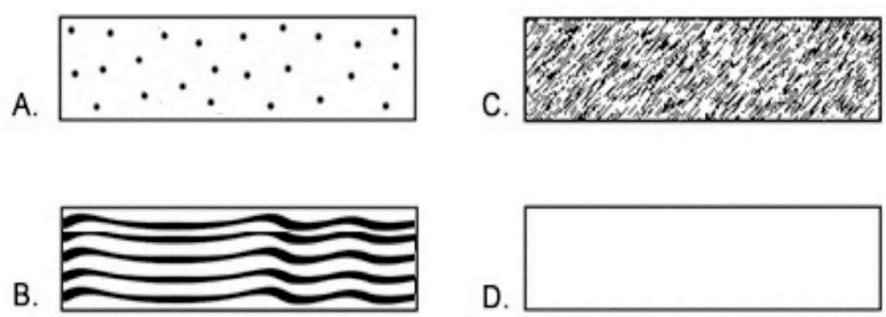
For the multiple-choice questions**,** record your answers on your ANSWER SHEET by filling in the circles. If you are not sure of the answer to a multiple-choice question, choose the BEST answer and go on to the next question. If you change your answer, be sure to erase your first answer completely. Choose only one answer for each question.

If you do not understand any of these instructions, please raise your hand.

**Remember; do NOT write anything in this test booklet!**

## Multiple-Choice Questions

1. **Below are four possible models of a gas.** Which model would a scientist use to show how water vapor condenses to a liquid? (Tests for basic level, correct answer is A)




**2. Both you and your friend can smell popcorn from different places in a room because the molecules that make up the odor:** (Checks basic model)

A. send a signal to your nose.
B. compress in the air.
C. move in all directions. (Correct choice)
D. expand in the air.

3. **If you could use a powerful microscope to see the particles in a gas, what would you see between the particles?** *(Tests for basic model)*

A. More particles
B. Air
C. Empty space (Correct choice)
D. Liquid

**4. Oxygen, hydrogen, and water are substances. Which of these substances are elements?** *(Checks basic – distinguishing elements from compounds)*

A. Oxygen, hydrogen, and water
B. Oxygen and hydrogen only (correct answer)
C. Oxygen only
D. Water only

**5. Which of the following is always true when a substance undergoes a phase change?** *(Tests for basic model)*

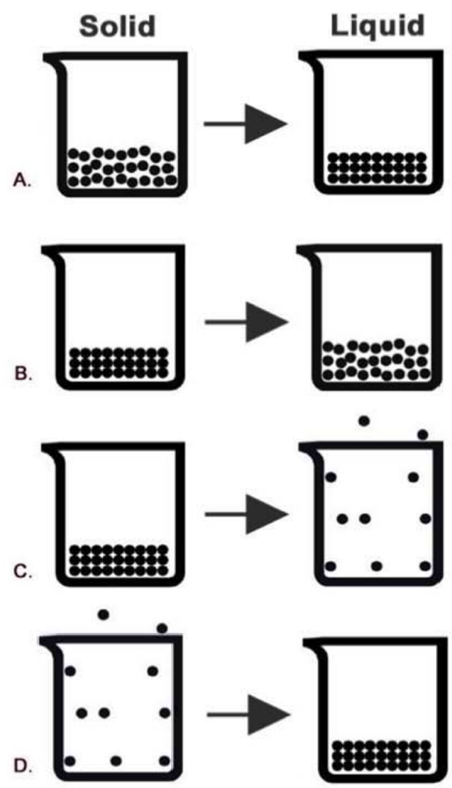
A. A new substance will form that has new properties.

B. The substance becomes liquid and heats up.

C. The substance's melting point becomes lower.

D. The substance still has the same type of molecules. (Correct answer)

**6. In the following models, each circle represents a wax molecule. Which model best represents what happens when a solid wax melts into liquid wax? (**Tests for Basic model)




**7. When a substance changes from a liquid to a solid, which of the following is true?** *(Tests for basic model)*

A. The molecules get colder.

B. The molecules of the solid move faster.

C. The molecules of the substance change from soft to hard.

D. The molecules move more slowly.

**8. Susan smells two bottles of perfume. They each smell different. Which of these answers does NOT explain why the odors are different?** *Checks at Basic Particle model level: different properties result from different arrangement of molecules in different substances.*

A. They have different properties.
B. They have different arrangements of atoms into molecules.
C. They are made up of different substances.
D. They have different mass. (Correct choice)

**9. Jason is trying to decide whether or not feathers are matter. Do you think feathers are matter?** *(Tests for mixed model)*

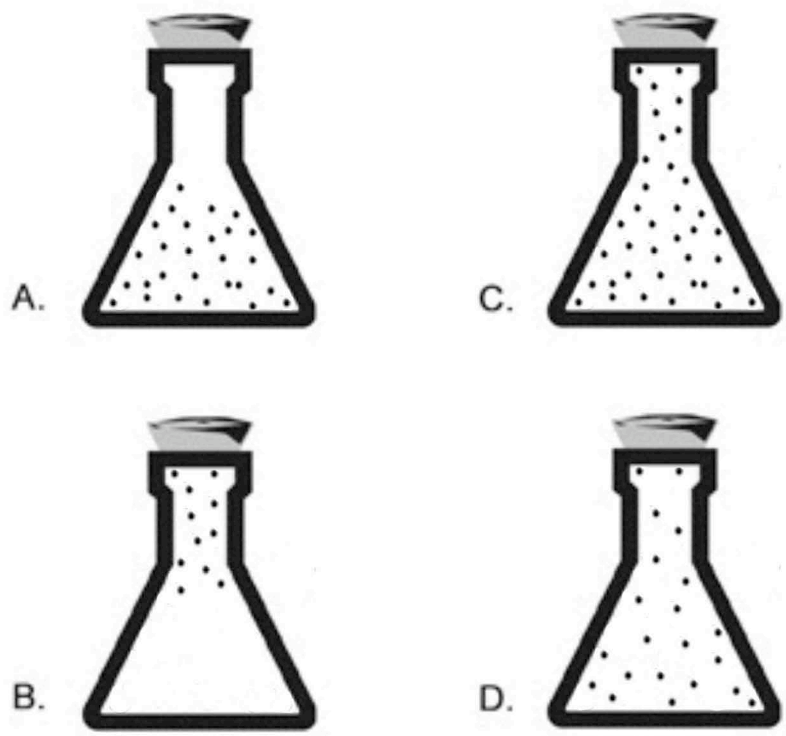
A. Yes, they are matter because you can see feathers.
B. No, they are not matter because they are too light to be matter.
C. Yes, they are matter because they have mass and occupy space. (correct answer)
D. No, they are not matter because they grow on birds.

**10. If a container of water is sealed and kept at the same temperature, then what can you say about the motion of the water molecules? (***Checks for basic model***)**

A. Keep moving at the same speed. (correct answer)
B. Slow down and eventually stop.
C. Slow down, but they won't stop.
D. Move faster over time.

***Use this model to answer question 11***
**Here is a model of a gas in a flask.**

**11. Imagine that some of the gas in the flask was removed. Which one of the following models best represents the gas that remains in the flask?** *(Tests for basic model- Answer D)*




**You have now completed Part 1.**

**Please go on to Part 2 and answer the Written Response questions**

## Written-response questions

## Please write your answer for question 1 on THIS SHEET.

**1. Bill and Shauna wondered if they could smell an air freshener faster in a cold room or a warm room. They decided to do an experiment: They made the room cold (50 F), plugged an air freshener in, and measured the time it takes for the smell to reach the door. The next day, they made the same room hot (85 F), plugged in a new air freshener, and again measured the time it takes for the smell to reach the door.**

A. What do you think would be the results of Bill and Shauna's experiment? Circle one of the following options:

1. The smell reaches the door at the same time in both temperatures
2. The smell reaches the door faster at $85^o$ F
3. The smell reaches the door faster at $50^o$ F

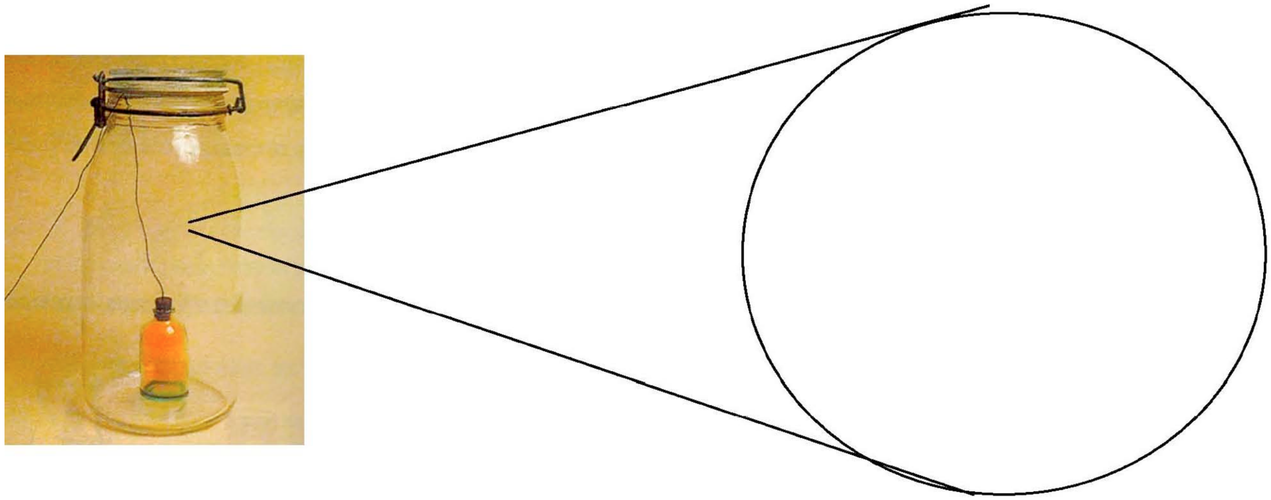
(**12**)B. Draw models that can help you explain your choice in part A. (Your models should show why the odors reach the door faster at one temperature than the other.)

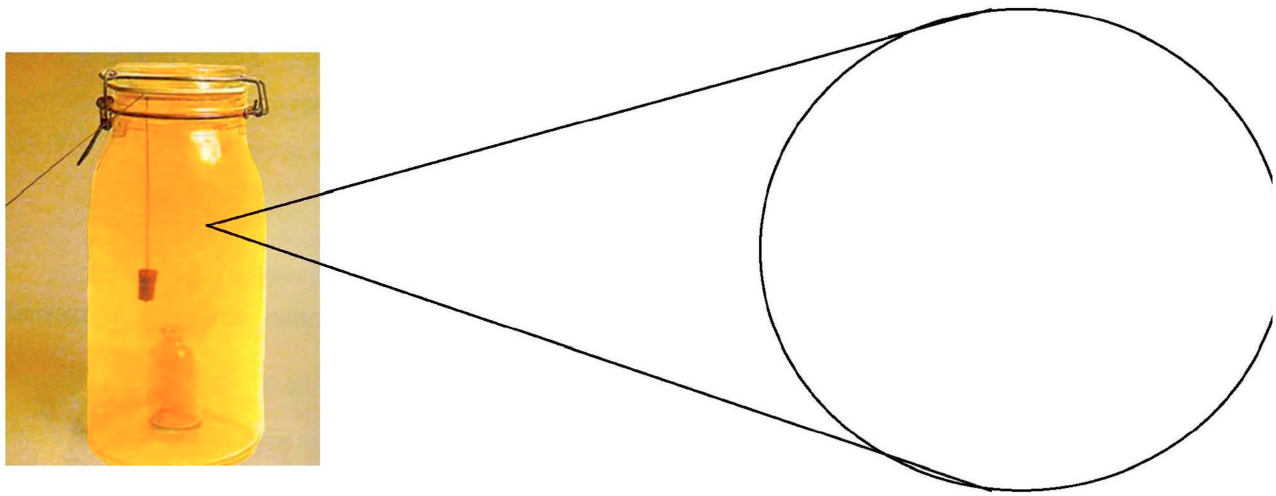
| Code | Model |
|---|---|
| | No drawing |
| **0** | Descriptive – describes in words or macro symbols of gas/air; smaller version of phenomena (a room with bill and shauna and air freshener |
| **1** | Mixed Model<br>- Particles and Descriptive (i.e. air particles, but waves of odor)<br>- Descriptive can also be including unnecessary macrolevel objects<br>- Movement may be included<br>- Relationship to temperature |
| **2** | Basic Particle Model<br>- air particles and/or odor particles<br>- Movement may be included<br>- Relationship to temperature |
| **3** | Complete Particle Model<br>- Air particles<br>- odor particles<br>- random motion (movement in all directions; collisions between particles) is required<br>- movement, correct relationship to temperature |

Make sure to label the different parts of your model. Key:

**(13)**C. Use your model to explain why you chose your answer in part A. Your statement should why the odors reach the door faster at one temperature than the other.)

| **Code** | **Part B** | |
|---|---|---|
| | No response | |
| **0** | Descriptive – describes model OR gives completely incorrect explanation OR uses prior experience to explain what is happening. | The smell would reach at the same time, because smells are not affected by the temperature in an area. OR The air is thicker when it is hot and thiner when tit is cold so therfor it is easier to travel when it is cold. |
| **1** | Mixed Model<br>Student tries to explain what is happening, but uses the incorrect mechanism or repeats correct choice " Warmer room air moves faster". Student may identify particles as molecules, but focuses on a macro level explanation (odor/air). | Because heat slows particles/molicules down and when colder particles/molicules speed up faster than hot air in the room. OR The oders reach the door faster in a 50o room because there is less molicules blocking its path to the door. |
| **2** | Basic Particle Model<br>Student is able to identify that (air and/or odor) molecules travel faster in a warm room (and/or slower in a cold room) in correct relation to temperature/energy | Hot air is also hot molecules and warm molecules are fast. So if the warmth heats up the molecules of the smell it will reach the door faster. OR The scent reaches the door faster at 85oF, because molicules speed up and spread apart. Moving faster, and making it to the door faster. |
| **3** | Complete Particle Model<br>Student is able to correctly explain on a microscopic level that in a warmer room, air moves faster because of higher energy and /or temperature, and collisions with odor molecules results in odor spreading/traveling thru room faster (and/or vice versa for cold room) | The odor would reach the door faster when it's hot because the molecules will gain energy and move all around and bounce off of things harder. When cold the molecules lose energy and don't bounce off of things as hard. OR I chose answer B because if the room is at 85oF then the odor makes it to the door faster. When the odor comes out of the jar and the warm air collides with the odor molecules that's what gives the odor energy to travel. The less hot air the odor has the less energy the odor has to travel. |





**(14)**

| Code | Model |
|---|---|
| | No drawing |
| **0** | Descriptive – describes in words or macro symbols of bromine/air; smaller version of phenomena (draws bottle with bromine in it); describes drawing |
| **1** | Mixed Model<br>- Particles and descriptive<br>- Descriptive can also be including unnecessary macrolevel objects<br>- Movement may be included |
| **2** | Basic Particle Model<br>- air particles (both models); bromine particles (2nd model)<br>- Movement may be included |
| **3** | Complete Particle Model<br>- Air particles (both models); bromine particles (2nd model)<br>- random motion (movement in all directions; collisions between particles) is required |

**(15)**B. Use your models to write a statement about what happened to the bromine, when the cork of the small bottle was opened in Figure 2.

| Code | Part B | |
|---|---|---|
| | No response | |
| **0** | Descriptive – describes model OR describes substances exactly as they appear and/or gives incorrect explanation | Figure 1 had gas with nothing messing it up. Figure 2 had less gas because of the Bromine gas. Or The bromine gas added more gas to the jar it was in. |
| **1** | Mixed Model<br>Student explains gases mixing on a macro level OR bromine gas entering the larger bottle and taking up empty space. Student may refer to atoms/molecules (i.e. there are atoms and molecules), but not as a means to explain what is happening. For example, a student could explain movement of particles out of the smaller bottle (leaving the smaller bottle, taking up space), but not in reference to mixing of bromine and air on a macro level). | The bromaine got mixed in with the air and it spread apart. OR<br>At first there were very few molicules, then when she pulled the cork, a bunch of molicules were in the jar. |
| **2** | Basic Particle Model<br>Student identifies particles as air and bromine molecules. Student is able to identify that air molecules and bromine molecules are mixing. | When the cork came off the bottle the odor molecules went in the air with the empty space and air molecules. |
| **3** | Complete Particle Model<br>Student identifies particles as air and bromine molecules. Student is able to correctly explain that bromine particles are mixing with air particles AND explains movement of particles (spreading/scattering/bouncing). | In figure two when Shayna opened the bottle, the bromine molecules escaped the jar. Then the air and bromine molecules collided into each other, making the molecules spread out faster throughout the jar. |

**Please write your answer for question 3 on THIS SHEET.**

**(16)3. Anna investigated two metal samples. Here is her data table she made:**

| | Sample 1 | Sample 2 |
|---|---|---|
| Mass | 10 grams | 15 grams |
| Color | Shiny gray | Shiny gray |
| Hardness | Scratched by an iron nail | Scratched by an iron nail |
| Melting point | 962°C | 661°C |
| Shape of object | Round circle (a ring shape) | Flat strip |

**Anna concluded that the two samples are two different metals. Which data in the data table help her tell that the two samples are different? Explain your answer.**

| Code | Category | Content | Example |
|---|---|---|---|
| | No response | Student leaves question blank | |
| **0** | Descriptive | Student identifies characteristics other than melting point as the data that identifies the difference. | "It has different detales" OR I help tell her that the two samples are different because the shape of object are round circle and flat strip. |
| **1** | Mixed | Student combines melting point with other characteristics to explain difference. Student fails to identify melting point as the only evidence of difference between the two samples. | The 10 grams and the 15 grams because they are different numbers and the melting different numbers and the shape describes them different. |
| **2** | Basic | Identify melting point as the property that can be used to conclude that the two samples are different, but gives no or incorrect explanation. | Melting point |
| **3** | Complete | Student identifies only melting point as evidence to conclude that the samples are different. Student clearly explains why only melting point can be used to conclude that the two samples are different (i.e. to be the same sample, they would have the same melting point OR they are made up of different atoms/molecules) . | The melting point tells Anna that the two metals are different because if the two metals were the same they would melt at the same temperature. It doesn't matter how big or small the metals are. |

**Appendix B: Calibration Study Test**

## Investigating and Questioning our World through Science and Technology

# *Sixth Grade Chemistry:*

# *How can I smell from a distance?*

**First and Last Name:** ________________________________________

**Date:** _____________

**Teacher Name:** _______________________________

**Class Hour:** _______________________________

**Gender:** (Place an X in the correct box) ☐ **Female** ☐ **Male**

# Part 1 -Multiple Choice

This test is an opportunity for you to show what you understand about chemistry concepts. Please try your best even if you are unsure of your answers.

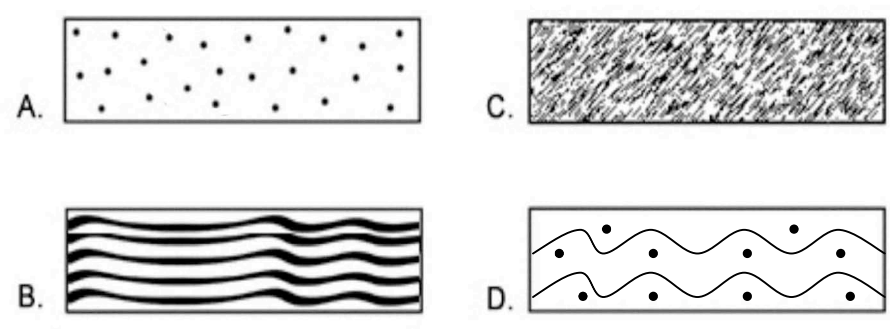
**Please use a pencil to answer the questions.**

For the multiple-choice questions**,** record your answers on your ANSWER SHEET by filling in the circles. If you are not sure of the answer to a multiple-choice question, choose the BEST answer and go on to the next question. If you change your answer, be sure to erase your first answer completely. Choose only one answer for each question.

If you do not understand any of these instructions, please raise your hand.

## Multiple-Choice Questions

1. **Below are four possible models of a gas.** Which model would a scientist use to show how water vapor condenses to a liquid?



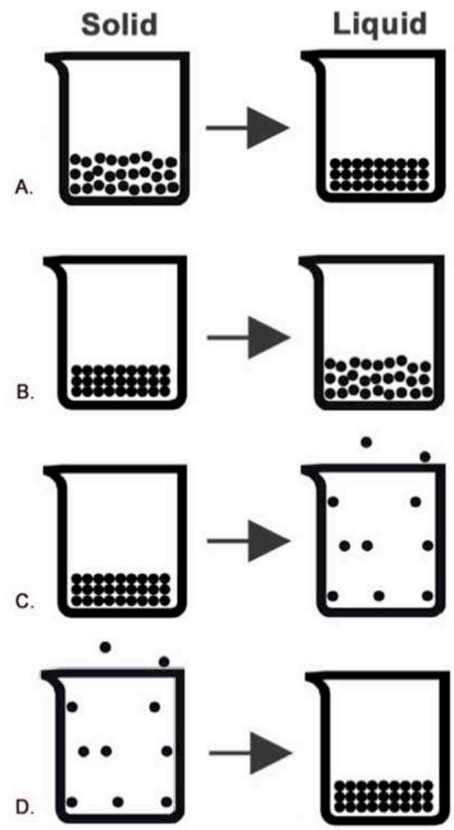

**2. When water condenses on a glass, the water molecules**

A. move faster. (Incomplete)
B. move slower. (basic)
C. do not move. (mixed)
D. get bigger. (mixed)

**3. Oxygen, hydrogen, and water are substances. Which of these substances are elements?** *(Checks Incomplete particle model – distinguishing elements from compounds)*

A. Oxygen, hydrogen, and water
B. Oxygen and hydrogen only (correct answer)
C. Oxygen only
D. Water only

**4. In the following models, each circle represents a wax molecule. Which model best represents what happens when a solid wax melts into liquid wax?**




basic particle

Complete particle

basic particle

basic particle

**5. Susan smells two bottles of perfume. They each smell different. Which of these answers does NOT explain why the odors are different?** *Checks at Incomplete Particle model level: different properties result from different arrangement of molecules in different substances.*

A. They have different properties.
B. They have different arrangements of atoms into molecules.
C. They are made up of different substances.
D. They have different mass. (Correct choice)

6. **Tom's younger brother is learning how to read a thermometer and asks, "Why does the red stuff in the thermometer go up when it gets hot outside?" What is a correct explanation that Tom can give to his brother?** *(Test for incomplete level)*
   - A. When the red stuff gets warmer, it increases in volume. Since it is confined in the tube, it must go up. (correct)
   - B. The red stuff in that little tube rises up because it is really sensitive to heat.
   - C. The red stuff goes up because the pressure of coldness is not there and the red stuff is free to move.
   - D. The heat hits the bottom of the thermometer and boosts up the temperature.

7. **If you breathe on a mirror, part of the mirror clouds up. What are you actually seeing when you see the mirror cloud up?** *(Tests for Incomplete model)*
   - A. Water droplets that formed from condensing water vapor from your breath
   - B. Carbon dioxide that you are breathing out of your lungs
   - C. Oxygen that you are breathing out from your lungs
   - D. Cooled nitrogen in the air around you

8. **Johnny puts water in a glass to drink. Before he drinks the water, he realizes he is late for school and leaves the glass on the counter. Johnny does not look at the glass until the next morning. The water in the glass:** *(Test for incomplete)*
   - A. The water evaporated into the air. (correct)
   - B. The water molecules shrank during the day.
   - C. The water molecules became larger.
   - D. The water sat out all night and no one touched it.

9. **Molly drops a small bottle filled with perfume in the corner of the room. She sweeps up the broken bottle and uses paper towel to clean up the remaining perfume. Is there any perfume left behind?** *(Test for incomplete)*
   - A. No. Molly cleaned up all the perfume.
   - B. No. Any left over perfume disappeared.
   - C. Yes. The perfume molecules made bigger perfume molecules.
   - D. Yes. There are a lot of perfume molecules in the air. (correct)

10. **Sam takes a cold bottle of water out of the refrigerator. He leaves it on the counter to run an errand for his mother. When he gets back, he notices drops of water on the outside of the bottle. Where did the water come from?** (Test for mixed)
    - A. Water came through the bottle
    - B. Air turned into a liquid
    - C. The coldness came through the bottle and made water
    - D. The water in the air condensed. (correct)

**11.** **Which of the following is an element?** *(Tests for incomplete model)*

A. $H_2O$
B. $CH_2OH$
C. $O_2$
D. $CO_2$

**12. If you could use a powerful microscope to see the particles in a gas, what would you see between the particles?** *(Tests for incomplete model)*

A. More particles (mixed)
B. Air (descriptive)
C. Empty space (Incomplete)
D. Liquid (descriptive)

**13. Which of the following is always true when a substance undergoes a phase change?** *(Tests for Incomplete model)*

A. A new substance will form that has new properties.
B. The substance becomes liquid and heats up.
C. The substance's melting point becomes lower.
D. The substance still has the same type of molecules.

**14. When a substance changes from a liquid to a solid, which of the following is true?** *(Tests for incomplete model)*

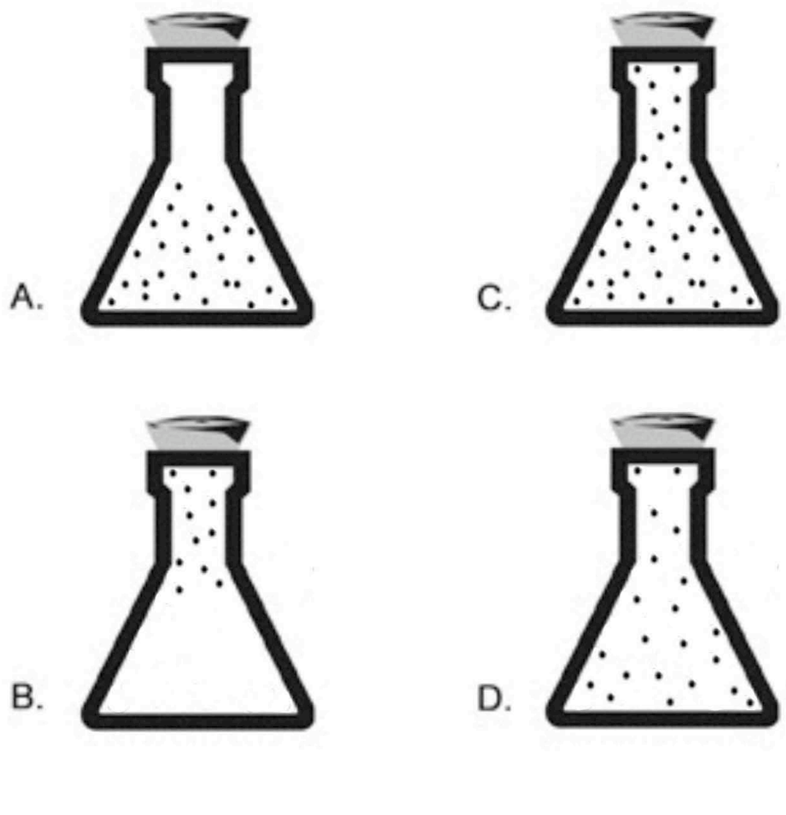
A. The molecules get colder.
B. The molecules of the solid move faster.
C. The molecules of the substance change from soft to hard.
D. The molecules move more slowly.

**15. If a container of water is sealed and kept at the same temperature, then what can you say about the motion of the water molecules?**

A. Keep moving at the same speed. (basic)
B. Slow down and eventually stop. (mixed)
C. Slow down, but they won't stop. (incomplete)
D. Move faster over time. (mixed)

*Use this model to answer question 16*
**Here is a model of a gas in a flask.**

**16. Imagine that some of the gas in the flask was removed. Which one of the following models best represents the gas that remains in the flask?** *(Tests for Incomplete model)*



**17. Tom's younger brother wants to know, "How does the red stuff in the thermometer know when it gets hot outside?" What is a correct explanation that Tom can give to his brother?**

a. The red stuff is a liquid that is made up of molecules. The molecules move faster as it gets warmer and the volume increases. (complete)
b. The red stuff is a liquid that is made up of smaller bits of the liquid. The smaller bits get together as it gets warmer making more liquid. (Incomplete)
c. The red stuff is a liquid. The coldness doesn't let it move. As it gets warmer, it can move more because the coldness goes away. (descriptive)
d. The heat is making the red stuff warmer. So the red stuff is showing the heat. Then the temperature goes up. (descriptive)

**18. Johnny puts water in a glass to drink. Before he drinks the water, he realizes he is late for school and leaves the glass on the counter. Johnny does not look at the glass until the next morning. The water in the glass is:** *(Tests for mixed model)*

a. Higher
b. Lower (correct)
c. The same

**19. Molly drops a small bottle filled with perfume in the corner of the room. She sweeps up the broken bottle and uses paper towel to clean up the remaining perfume. Jason walks in the room just as Molly finishes cleaning up. Can Jason smell the perfume?** *(Tests for mixed model)*

a. He can't smell the perfume
b. He smells the perfume a little (correct)
c. He can smell the perfume a lot

**20. Which of the following is NOT an example of a phase change?** *(Tests for Incomplete model)*

a. Water boiling
b. Wax melting
c. Wood burning
d. Gas condensing

**21. Which of the following is a compound?** *(Tests for Incomplete model)*

a. $O_2$
b. $N_2$
c. $H_2$
d. $NH_3$

## You have now completed Part 1.

## Please go on to Part 2 and answer the Written Response questions

## Written-response questions

## Please write your answer for question 1 on THIS SHEET.

**22. Bill and Shauna wondered if they could smell an air freshener faster in a cold room or a warm room. They decided to do an experiment: They made the room cold (50 F), plugged an air freshener in, and measured the time it takes for the smell to reach the door. The next day, they made the same room hot (85 F), plugged in a new air freshener, and again measured the time it takes for the smell to reach the door.**

A. What do you think would be the results of Bill and Shauna's experiment? Circle one of the following options:

1. The smell reaches the door at the same time in both temperatures
2. The smell reaches the door faster at 85$^{o}$ F
3. The smell reaches the door faster at 50$^{o}$ F

B. Draw models that can help you explain your choice in part A. (Your models should show why the odors reach the door faster at one temperature than the other.)

| Code | Part B |
|---|---|
| | No drawing |
| **0** | Descriptive – describes in words or macro symbols of gas/air; smaller version of phenomena (a room with bill and shauna and air freshener |
| **1** | Mixed Model<br>- Particles and Descriptive<br>- Descriptive can also be including unnecessary macrolevel objects<br>- Movement may be included<br>- Relationship to temperature |
| **2** | Incomplete Particle Model<br>- particles (odor or air)<br>- Movement not included<br>- Relationship to temperature |
| **3** | Basic Particle Model<br>- air particles<br>- odor particles<br>- Motion<br>- Relationship to temperature |
| **4** | Complete Particle Model<br>- Air particles<br>- odor particles<br>- random motion (movement in all directions; collisions between particles) is required<br>- movement, correct relationship to temperature |

Make sure to label the different parts of your model. Key:

C. Use your model to explain why you chose your answer in part A. Your statement should why the odors reach the door faster at one temperature than the other.)

| **Code** | **Part C** | |
|---|---|---|
| | No response | |
| **0** | Descriptive – describes model OR gives completely incorrect explanation (i.e. an external source creates movement) OR uses prior experience to explain what is happening. | I chose my answer in A because the temperature of a room doesn't affect how fast or slow an odor moves through a room. |
| **1** | Mixed Model<br>Student tries to explain what is happening on a molecular level, but uses the incorrect mechanism OR simply repeats correct choice "Warmer room air moves faster" | Molecules move faster at colder temperatures. The heat slowes the molecules down. |
| **2** | Incomplete Particle Model<br>Student may identify particles as molecules, but focuses partially on a macro level explanation (odor/air). Student explanation focuses only on one gas (odor/air/gas/atoms/molecules) moving faster. Although mostly correct, student answer may include incorrect mechanism. | The warmer the temperature the faster atoms move. The colder the temperature the slower atoms move. |
| **3** | Basic Particle Model<br>Student is able to identify that (air and/or odor) molecules travel faster in a warm room (and/or slower in a cold room) in correct relation to temperature/energy | The smell reaches the door faster at $85^{o}$F. The molecules are warmer and move faster in room A. In room B, the molecules are colder and moves slower. Therefore, the smell reaches the door faster at 85oF, because the warmer the room the faster the molecules and atoms move. |
| **4** | Complete Particle Model<br>Student is able to correctly explain on a microscopic level that in a warmer room, air moves faster because of higher energy, resulting in odor spreading/traveling thru room faster (and/or vice versa for cold room) | I chose the smell will reach the faster in $85^{o}$C because there will be more heat energy in the room than in $50^{o}$F. The heat energy will cause the molecules to speed up and reach the other side faster. |

## Please write your answer for question 2 on THIS SHEET.

**23. You are trying to explain to a friend how bromine can go from a liquid to a gas.**

a. Create a model that shows what happens when bromine goes from a liquid to a gas

| Code | Part A |
|---|---|
| | No drawing |
| **0** | Descriptive – describes bromine in words or macro symbols (drawing a test tube filled with bromine) |
| **1** | Mixed Model<br>- Contains both descriptive and particle ideas: Particles within liquid bromine; squiggly lines representing gas leaving bromine liquid surface<br>- Descriptive can also be including unnecessary macrolevel objects<br>- Movement may be included |
| **2** | Incomplete Particle Model<br>- particles represent bromine (implicit)<br>- Movement not included |
| **3** | Basic Particle Model<br>- bromine molecules<br>- Motion of molecules included (but is not correct for each phase) |
| **4** | Complete Particle Model<br>- bromine molecules<br>- Correct motion of molecules in liquid vs. gaseous state |

b. Using your model, explain to your friend how this happens.

| Code | Part B | |
|---|---|---|
| | No response | |
| **0** | Descriptive – describes their drawing, defines what a phase change is as when matter changes state and/or includes incorrect explanation | This happens by freezing the bromine. |
| **1** | Mixed Model<br>Student describes heat/warm needed for phase change from liquid to gas or evaporation as the cause for a phase change. | The liquid gets warm and it undergoes a phase change making a gas. |
| **2** | Incomplete Particle Model<br>Although student may identify particles as molecules, they do not identify them as bromine molecules. Student is able to describe spacing between molecules in gaseous versus liquid state OR student describes relationship of movement to different states (liquid vs. gas or change of state) | Bromine can go from a liquid to a gas bv adding heat energy to speed up the molecules. |
| **3** | Basic Particle Model<br>Student identifies particles as bromine molecules. Student is able to distinguish spacing between molecules in gaseous versus liquid state. Student incorrectly describes movement during the different phases. Student may identify temperature/energy affect. | The atoms and molecules are far apart in the bromine's liquid state. They move around a little and the atoms are farther apart from each other than the atoms in a solid. When the liquid is heated, a gas form of the liquid bromine forms. The atoms in a gas move more quickly and are farther apart than the atom in a liquid. |
| **4** | Complete Particle Model<br>Student identifies particles as bromine molecules. Student is able to distinguish spacing between molecules in gaseous versus liquid state. Student is able to correctly distinguish movement during the different phases and how temperature/energy affects movement. | |

**24. Your friend does not understand how water vapor, water and ice can all be the same thing.**

a. Create models that show the differences of water in these states (gas, liquid, solid).

| Code | Part A |
|---|---|
| | No drawing |
| **0** | Descriptive – describes in words or macro symbols of water (i.e. drawing a glass of water, ice cube, etc.) |
| **1** | Mixed Model<br>- Particle and descriptive (i.e. particles within water vapor, water, and ice)<br>- Descriptive can also be including unnecessary macrolevel objects<br>- Movement may be included |
| **2** | Incomplete Particle Model<br>- particles represent water in three phases<br>- Empty space between particles (may not be correct for all phases)<br>- Movement not included |
| **3** | Basic Particle Model<br>- Water molecules<br>- Empty space between particles (may not be correct for all phases)<br>- Motion included, but not correctly indicated for each of the phases |
| **4** | Complete Particle Model<br>- Water molecules<br>- Motion is correctly indicated for each of the phases |

b. Using your models, explain why water looks different in these different states.

| **Code** | **Part B** | |
|---|---|---|
| | No response | |
| **0** | Descriptive – describes water in each state exactly as it appears, defines what a phase change is; describes drawing | Water looks different in these states because all the atoms rearrange depending on the temperature. |
| **1** | Mixed Model<br>Although the student may mention atoms or molecules, student describes how a phase change occurs on a macro level. | Water looks different in those states because it was frozen and boiled. |
| **2** | Incomplete Particle Model<br>Although student may identify particles as molecules, they do not fully understand what an atom or molecule is. Student is able to distinguish spacing between molecules in each state OR difference in movement in each state. | Water looks different in these different states because water atoms are moving faster and slower as the phase changes so, the water will look different. |
| **3** | Basic Particle Model<br>Student identifies particles as water molecules. Student is able to describe spacing between molecules in each state. Student is unable to distinguish movement during the different phases. For example, can describe movement of a liquid and a gas, but a solid does not move. | Water looks different because the molecules are moving at different seeds and they are spaced out differently |
| **4** | Complete Particle Model<br>Student identifies particles as water molecules. Student is able to describe spacing between molecules in each state. Student is able to describe movement during the different phases. | Water in the phases of a solid liquid and gas are all different. This is because the atoms that make up the water molecules are all moving faster or slower with a different amount of spacing in each state. In a gas the molecule are moving very fast with a lot of space, bumping into each other rapidly. In a liquid the water molecules are moving at medium speed bumping into each other with space between them. In a solid the water molecules are moving at a slow speed with little space between them. |

**Please write your answer for question 2 ON THE NEXT PAGE.**

**25. Shayna had a small bottle of Bromine gas. The bottle was closed with a cork. She tied a string to the cork, and then placed the bottle inside a larger jar. The large jar had air in it. She sealed the large jar shut. (See Figure 1.) Next, Shayna opened the small bottle by pulling the string connected to the cork. Figure 2 shows what happened after the cork of the small bottle was opened.**

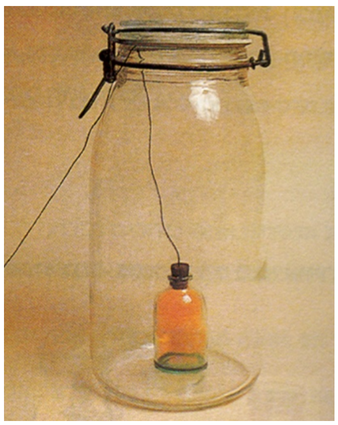
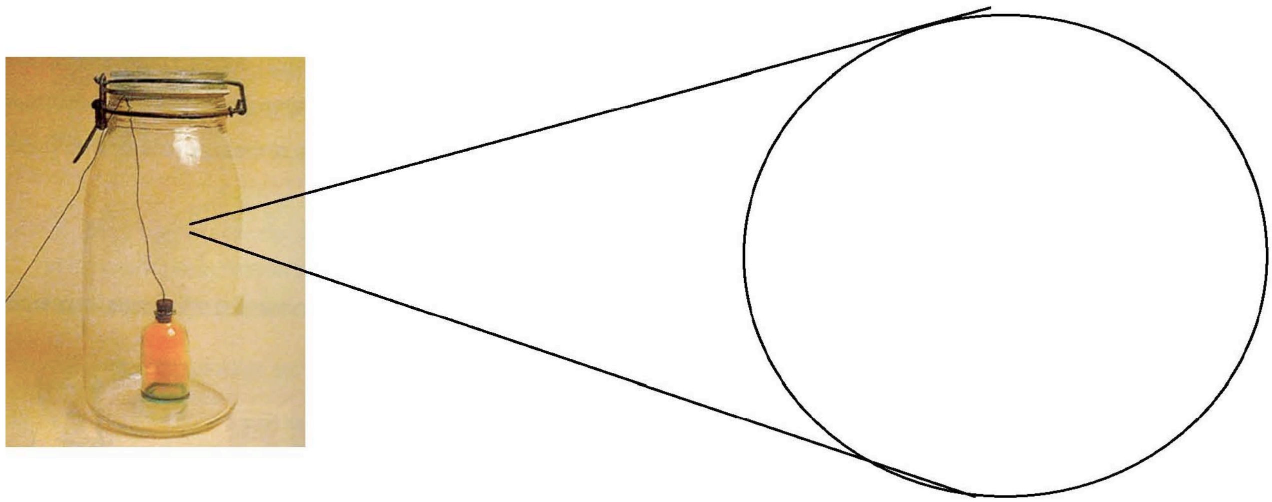
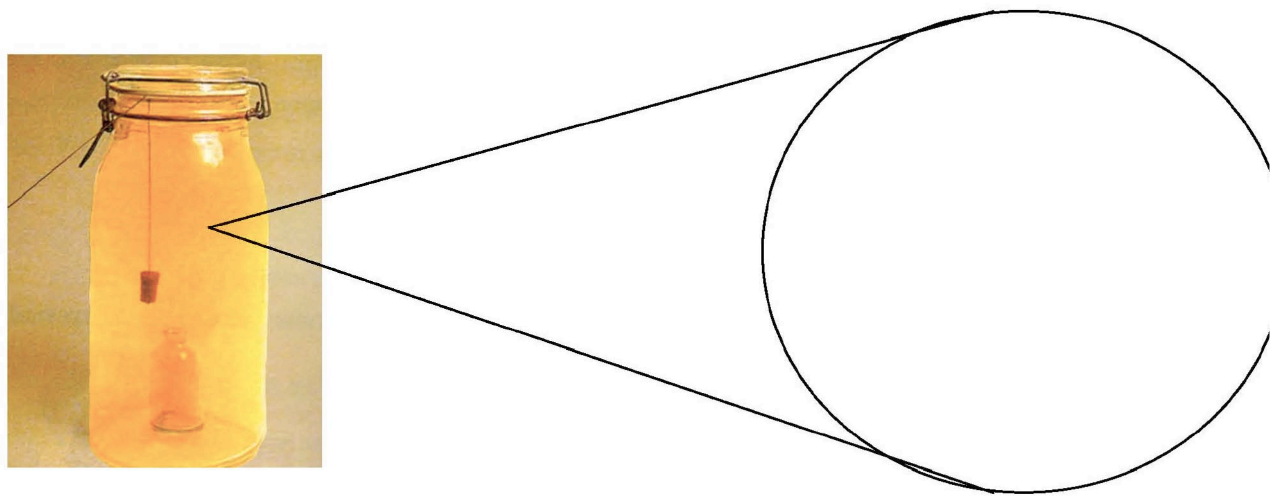


**Figure 1**

**Figure 2**

A. Imagine that you have a very powerful microscope that would allow you to zoom into a tiny spot in the large jar. In the circles ***ON THE NEXT PAGE***, draw a picture of what you think is in the large jar *before* and *after* opening the cork of the small bottle.





| Code | Part A |
|---|---|
| | No drawing |
| 0 | Descriptive – describes in words or macro symbols of bromine/air; smaller version of phenomena (draws bottle with bromine in it); describes drawing |
| 1 | Mixed Model<br>- Particles and descriptive<br>- Descriptive can also be including unnecessary macrolevel objects<br>- Movement may be included |
| 2 | Incomplete Particle Model<br>- particles<br>- Movement not included |
| 3 | Basic Particle Model<br>- air particles (both models); bromine particles (2nd model)<br>- Motion |
| 4 | Complete Particle Model<br>- Air particles (both models); bromine particles (2nd model)<br>- random motion (movement in all directions; collisions between particles) is required |

B. Use your models to write a statement about what happened to the bromine, when the cork of the small bottle was opened in Figure 2.

| **Code** | **Part B** | |
|---|---|---|
| | No response | |
| **0** | Descriptive – describes substances exactly as they appear and/or gives incorrect explanation | When bromine was trapped inside the bottle and had a cork put over it, pressure started to build and when the large jar was closed, the cork was released, causing pressure in the bromine shooting it up and penetrating the air in the large jar. |
| **1** | Mixed Model<br>Student explains gases mixing or entering the larger bottle and taking up space on a macro level. Student may refer to atoms/molecules (i.e. there are atoms and molecules), but not as a means to explain what is happening. | The bromine spread throughout the jar and gas molecules were more in the jar than before. |
| **2** | Incomplete Particle Model<br>Student may identify particles as molecules. Student explains movement of particles out of the smaller bottle (leaving the smaller bottle, taking up space), but may only refer to mixing of bromine and air on a macro level). | The compressed gas escaped and filled the closed big jar. The atoms in bromine filled up spaces between the water gas atoms. OR When the cork was opened bromine molecules went inside of the jar causing more molecules in it. |
| **3** | Basic Particle Model<br>Student identifies particles as air and bromine molecules. Student is able to identify that air and bromine molecules are mixing. | Before the jar was only filled with air molecules and then the bromine particles mixed into the air that was already inside the jar. |
| **4** | Complete Particle Model<br>Student identifies particles as air and bromine molecules. Student is able to correctly explain that bromine particles are mixing with air particles. Student explains movement of particles (spreading/scattering/bouncing). | When the cork was pulled releasing the bromine gas, the bromine atoms scattered across the jar. The air and bromine atoms are next to each other, scattered, and traveling fast, and far apart. |

# Appendix C: Ability Estimates

*MLE Estimates -Highlighted students have high infits, or a lot of random responses.*

| Student ID | Raw | Max | Est. | Err. | infit | t | outfit | t |
|---|---|---|---|---|---|---|---|---|
| 1001 | 46 | 52 | 3.53 | 0.49 | 1.82 | 1.69 | 0.85 | 0.08 |
| 1003 | 29 | 52 | 0.83 | 0.34 | 1.36 | 1.09 | 1.15 | 0.46 |
| 1006 | 33 | 52 | 1.32 | 0.36 | 1.10 | 0.40 | 0.84 | -0.20 |
| 1007 | 32 | 51 | 1.28 | 0.36 | 1.59 | 1.53 | 1.13 | 0.41 |
| 1008 | 25 | 52 | 0.38 | 0.33 | 1.18 | 0.67 | 1.10 | 0.40 |
| 1009 | 35 | 52 | 1.59 | 0.38 | 0.47 | -1.75 | 0.35 | -1.48 |
| 1010 | 26 | 52 | 0.49 | 0.33 | 0.83 | -0.53 | 0.82 | -0.39 |
| 1011 | 21 | 42 | 0.22 | 0.38 | 1.19 | 0.72 | 1.16 | 0.54 |
| 1012 | 26 | 52 | 0.49 | 0.33 | 0.88 | -0.34 | 1.81 | 1.86 |
| 1013 | 28 | 52 | 0.71 | 0.34 | 0.61 | -1.44 | 0.64 | -0.97 |
| 1014 | 32 | 52 | 1.19 | 0.36 | 0.82 | -0.43 | 0.82 | -0.19 |
| 1015 | 37 | 52 | 1.89 | 0.39 | 0.72 | -0.67 | 0.62 | -0.48 |
| 1016 | 31 | 52 | 1.07 | 0.35 | 0.68 | -0.91 | 1.57 | 1.12 |
| 1017 | 26 | 52 | 0.49 | 0.33 | 1.05 | 0.26 | 1.63 | 1.54 |
| 1018 | 28 | 52 | 0.71 | 0.34 | 0.85 | -0.44 | 0.84 | -0.32 |
| 1019 | 17 | 52 | -0.54 | 0.35 | 0.62 | -1.61 | 0.90 | -0.28 |
| 1020 | 30 | 46 | 1.37 | 0.40 | 1.36 | 1.00 | 0.91 | 0.00 |
| 1021 | 27 | 52 | 0.60 | 0.34 | 0.71 | -1.06 | 0.59 | -1.24 |
| 1022 | 31 | 52 | 1.07 | 0.35 | 0.95 | -0.03 | 0.60 | -0.76 |
| 1023 | 28 | 52 | 0.71 | 0.34 | 0.90 | -0.27 | 0.62 | -1.04 |
| 1024 | 21 | 52 | -0.07 | 0.34 | 0.95 | -0.10 | 0.92 | -0.19 |
| 1025 | 34 | 52 | 1.46 | 0.37 | 0.97 | 0.05 | 0.66 | -0.50 |
| 1026 | 30 | 52 | 0.95 | 0.35 | 0.57 | -1.55 | 0.58 | -1.05 |
| 1027 | 23 | 52 | 0.15 | 0.33 | 0.98 | 0.01 | 0.99 | 0.08 |
| 1028 | 21 | 52 | -0.07 | 0.34 | 0.73 | -1.03 | 0.88 | -0.34 |
| 1029 | 29 | 52 | 0.83 | 0.34 | 0.90 | -0.19 | 0.78 | -0.33 |
| 1030 | 28 | 52 | 0.71 | 0.34 | 0.76 | -0.81 | 0.76 | -0.57 |
| 1031 | 26 | 52 | 0.49 | 0.33 | 1.01 | 0.13 | 0.75 | -0.59 |
| 1032 | 37 | 52 | 1.89 | 0.39 | 1.07 | 0.29 | 0.82 | -0.09 |
| 1033 | 25 | 45 | 0.90 | 0.37 | 1.25 | 0.77 | 0.84 | -0.18 |
| 1034 | 36 | 52 | 1.74 | 0.38 | 1.32 | 0.88 | 0.66 | -0.44 |
| 1035 | 27 | 52 | 0.60 | 0.34 | 0.63 | -1.39 | 0.68 | -0.91 |
| 1036 | 34 | 52 | 1.46 | 0.37 | 1.57 | 1.47 | 1.27 | 0.64 |
| 1037 | 22 | 52 | 0.04 | 0.34 | 0.80 | -0.72 | 0.76 | -0.86 |
| 1038 | 24 | 52 | 0.27 | 0.33 | 1.63 | 2.03 | 1.66 | 1.89 |
| 1039 | 30 | 52 | 0.95 | 0.35 | 1.68 | 1.92 | 1.99 | 1.99 |
| 1040 | 21 | 52 | -0.07 | 0.34 | 1.46 | 1.61 | 1.61 | 1.93 |
| 1041 | 27 | 52 | 0.60 | 0.34 | 1.10 | 0.44 | 0.79 | -0.54 |
| 1042 | 18 | 52 | -0.42 | 0.35 | 1.00 | 0.09 | 1.00 | 0.10 |
| 1043 | 37 | 52 | 1.89 | 0.39 | 0.64 | -0.92 | 0.70 | -0.33 |
| 1044 | 16 | 52 | -0.66 | 0.35 | 1.23 | 0.90 | 1.01 | 0.11 |
| 1045 | 15 | 49 | -0.75 | 0.37 | 1.06 | 0.32 | 1.09 | 0.41 |
| 1046 | 28 | 52 | 0.71 | 0.34 | 0.93 | -0.16 | 1.20 | 0.62 |
| 1047 | 38 | 52 | 2.04 | 0.40 | 0.87 | -0.19 | 0.41 | -0.86 |
| 1048 | 28 | 52 | 0.71 | 0.34 | 0.72 | -0.95 | 0.84 | -0.33 |
| 1049 | 29 | 52 | 0.83 | 0.34 | 0.92 | -0.12 | 1.24 | 0.63 |

| Student ID | Raw | Max | Est. | Err. | infit | t | outfit | t |
|---|---|---|---|---|---|---|---|---|
| 1050 | 29 | 51 | 0.92 | 0.35 | 0.73 | -0.85 | 0.59 | -0.99 |
| 1051 | 28 | 50 | 0.88 | 0.35 | 1.09 | 0.37 | 1.60 | 1.26 |
| 1052 | 31 | 52 | 1.07 | 0.35 | 1.12 | 0.44 | 0.57 | -0.84 |
| 1053 | 38 | 52 | 2.04 | 0.40 | 0.77 | -0.47 | 0.80 | -0.08 |
| 1054 | 31 | 52 | 1.07 | 0.35 | 1.58 | 1.52 | 1.07 | 0.31 |
| 1055 | 32 | 52 | 1.19 | 0.36 | 1.22 | 0.71 | 1.31 | 0.71 |
| 1056 | 31 | 52 | 1.07 | 0.35 | 1.05 | 0.26 | 1.01 | 0.19 |
| 1057 | 40 | 52 | 2.37 | 0.41 | 0.73 | -0.61 | 0.55 | -0.63 |
| 1058 | 25 | 52 | 0.38 | 0.33 | 1.31 | 1.06 | 2.13 | 2.47 |
| 1059 | 30 | 52 | 0.95 | 0.35 | 1.12 | 0.49 | 0.82 | -0.31 |
| 1060 | 32 | 52 | 1.19 | 0.36 | 0.40 | -2.12 | 0.29 | -1.81 |
| 1061 | 29 | 52 | 0.83 | 0.34 | 0.53 | -1.60 | 0.68 | -0.59 |
| 1062 | 34 | 52 | 1.46 | 0.37 | 1.34 | 0.98 | 1.00 | 0.19 |
| 1063 | 31 | 52 | 1.07 | 0.35 | 1.02 | 0.17 | 1.96 | 1.65 |
| 1064 | 30 | 52 | 0.95 | 0.35 | 0.97 | 0.01 | 0.59 | -1.01 |
| 1065 | 27 | 52 | 0.60 | 0.34 | 0.93 | -0.17 | 1.16 | 0.56 |
| 1066 | 24 | 51 | 0.33 | 0.34 | 0.81 | -0.61 | 0.57 | -1.27 |
| 1067 | 25 | 52 | 0.38 | 0.33 | 1.25 | 0.91 | 1.15 | 0.51 |
| 1068 | 33 | 52 | 1.32 | 0.36 | 1.35 | 1.06 | 2.36 | 2.21 |
| 1069 | 18 | 48 | -0.32 | 0.35 | 0.84 | -0.55 | 0.99 | 0.06 |
| 1070 | 30 | 52 | 0.95 | 0.35 | 0.62 | -1.32 | 0.61 | -0.95 |
| 1071 | 34 | 52 | 1.46 | 0.37 | 0.85 | -0.32 | 1.02 | 0.22 |
| 1072 | 28 | 52 | 0.71 | 0.34 | 1.20 | 0.74 | 0.99 | 0.09 |
| 1073 | 32 | 52 | 1.19 | 0.36 | 1.09 | 0.35 | 0.89 | -0.04 |
| 1074 | 28 | 51 | 0.77 | 0.35 | 1.06 | 0.30 | 0.90 | -0.09 |
| 1075 | 38 | 52 | 2.04 | 0.40 | 1.14 | 0.47 | 0.80 | -0.07 |
| 1076 | 29 | 52 | 0.83 | 0.34 | 1.07 | 0.31 | 0.67 | -0.61 |
| 1077 | 28 | 52 | 0.71 | 0.34 | 0.97 | -0.02 | 0.69 | -0.82 |
| 1078 | 30 | 52 | 0.95 | 0.35 | 0.89 | -0.28 | 0.70 | -0.67 |
| 1079 | 33 | 52 | 1.32 | 0.36 | 1.16 | 0.57 | 0.88 | -0.09 |
| 1080 | 26 | 52 | 0.49 | 0.33 | 0.99 | 0.05 | 0.80 | -0.43 |
| 1081 | 24 | 52 | 0.27 | 0.33 | 0.87 | -0.42 | 0.71 | -0.93 |
| 1082 | 25 | 51 | 0.48 | 0.34 | 0.87 | -0.38 | 1.57 | 1.61 |
| 1083 | 18 | 51 | -0.34 | 0.35 | 1.21 | 0.81 | 1.00 | 0.08 |
| 1084 | 24 | 52 | 0.27 | 0.33 | 0.99 | 0.04 | 1.23 | 0.79 |
| 1085 | 22 | 52 | 0.04 | 0.34 | 1.56 | 1.89 | 1.65 | 2.05 |
| 1086 | 23 | 52 | 0.15 | 0.33 | 1.29 | 1.02 | 2.02 | 2.39 |
| 1087 | 37 | 52 | 1.89 | 0.39 | 0.80 | -0.40 | 0.42 | -0.97 |
| 1088 | 28 | 52 | 0.71 | 0.34 | 0.69 | -1.09 | 0.66 | -0.91 |
| 1089 | 25 | 52 | 0.38 | 0.33 | 0.81 | -0.61 | 1.38 | 1.06 |
| 1090 | 27 | 48 | 0.75 | 0.35 | 1.13 | 0.50 | 1.74 | 1.59 |
| 1091 | 20 | 52 | -0.18 | 0.34 | 0.90 | -0.30 | 1.16 | 0.65 |
| 1092 | 29 | 52 | 0.83 | 0.34 | 1.36 | 1.08 | 1.35 | 0.81 |

# Appendix D: Item Statistics

```
Item Statistics (MLE)
Number of Active Items = 29
Students = 89
--------------------------------------------------------------------
....................................................................

Item: 1      Item Set: base      Variable: PMM(by parameter)
Infit MNSQ = 0.91 t = -0.56 Outfit MNSQ = 0.90 t = -0.60

Categories             0      1      2 missing
Responses              0      1      2
Count                  8     41     39       1
Percent (%)         9.09  46.59  44.32
Pt-Biserial        -0.23  -0.21   0.35
Mean Ability        0.32   0.66   1.09      NA
SD Abilities        0.34   0.35   0.36      NA
Step Difficulties         -1.05   0.90
Thresholds            NA  -1.16   1.02
Error                 NA   0.15   0.40

====================================================================

Item: 2      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.90 t = -0.63 Outfit MNSQ = 0.89 t = -0.69

Categories             0      1 missing
Responses              0      1
Count                 49     40       0
Percent (%)        55.06  44.94
Pt-Biserial        -0.43   0.43
Mean Ability        0.55   1.15      NA
SD Abilities        0.34   0.36      NA
Step Difficulties          1.04
Thresholds            NA     NA
Error                 NA     NA

====================================================================

Item: 3      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.96 t = -0.22 Outfit MNSQ = 0.95 t = -0.29

Categories             0      1 missing
Responses              0      1
Count                 19     70       0
Percent (%)        21.35  78.65
Pt-Biserial        -0.41   0.41
Mean Ability        0.28   0.97      NA
SD Abilities        0.34   0.36      NA
Step Difficulties         -0.57
Thresholds            NA     NA
Error                 NA     NA

====================================================================
```

```
Item: 4      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.95 t = -0.25 Outfit MNSQ = 0.93 t = -0.42

Categories              0      1 missing
Responses               0      1
Count                   4     85       0
Percent (%)          4.49  95.51
Pt-Biserial         -0.20   0.20
Mean Ability         0.24   0.85      NA
SD Abilities         0.34   0.35      NA
Step Difficulties          -2.43
Thresholds             NA     NA
Error                  NA     NA

====================================================================

Item: 5      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.99 t = -0.03 Outfit MNSQ = 0.97 t = -0.16

Categories              0      1 missing
Responses               0      1
Count                  10     79       0
Percent (%)         11.24  88.76
Pt-Biserial         -0.13   0.13
Mean Ability         0.57   0.85      NA
SD Abilities         0.34   0.35      NA
Step Difficulties          -1.41
Thresholds             NA     NA
Error                  NA     NA

====================================================================

Item: 6      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.04 t = 0.31 Outfit MNSQ = 1.05 t = 0.38

Categories              0      1 missing
Responses               0      1
Count                  31     58       0
Percent (%)         34.83  65.17
Pt-Biserial         -0.18   0.18
Mean Ability         0.65   0.91      NA
SD Abilities         0.35   0.36      NA
Step Difficulties           0.12
Thresholds             NA     NA
Error                  NA     NA

====================================================================
```

```
Item: 7      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.14 t = 0.95 Outfit MNSQ = 1.15 t = 0.98

Categories              0      1 missing
Responses               0      1
Count                  45     43       1
Percent (%)         51.14  48.86
Pt-Biserial         -0.14   0.14
Mean Ability         0.72   0.92      NA
SD Abilities         0.35   0.36      NA
Step Difficulties           0.84
Thresholds             NA     NA
Error                  NA     NA

====================================================================

Item: 8      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.01 t = 0.14 Outfit MNSQ = 1.09 t = 0.63

Categories              0      1 missing
Responses               0      1
Count                   9     80       0
Percent (%)         10.11  89.89
Pt-Biserial         -0.29   0.29
Mean Ability         0.22   0.89      NA
SD Abilities         0.34   0.35      NA
Step Difficulties          -1.56
Thresholds             NA     NA
Error                  NA     NA

====================================================================

Item: 9      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.10 t = 0.67 Outfit MNSQ = 1.23 t = 1.44

Categories              0      1 missing
Responses               0      1
Count                   7     82       0
Percent (%)          7.87  92.13
Pt-Biserial         -0.14   0.14
Mean Ability         0.47   0.85      NA
SD Abilities         0.34   0.35      NA
Step Difficulties          -1.87
Thresholds             NA     NA
Error                  NA     NA

====================================================================
```

```
Item: 10      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.00 t = 0.08 Outfit MNSQ = 0.99 t = 0.02

Categories              0      1 missing
Responses               0      1
Count                  23     66       0
Percent (%)         25.84  74.16
Pt-Biserial         -0.11   0.11
Mean Ability         0.70   0.86      NA
SD Abilities         0.35   0.35      NA
Step Difficulties          -0.35
Thresholds             NA     NA
Error                  NA     NA

====================================================================

Item: 11      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.00 t = 0.06 Outfit MNSQ = 1.00 t = 0.02

Categories              0      1 missing
Responses               0      1
Count                  46     43       0
Percent (%)         51.69  48.31
Pt-Biserial         -0.56   0.56
Mean Ability         0.45   1.22      NA
SD Abilities         0.34   0.36      NA
Step Difficulties           0.85
Thresholds             NA     NA
Error                  NA     NA

====================================================================

Item: 12      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.84 t = -1.08 Outfit MNSQ = 0.82 t = -1.18

Categories              0      1 missing
Responses               0      1
Count                  21     68       0
Percent (%)         23.60  76.40
Pt-Biserial         -0.36   0.36
Mean Ability         0.37   0.96      NA
SD Abilities         0.35   0.35      NA
Step Difficulties          -0.48
Thresholds             NA     NA
Error                  NA     NA

====================================================================
```

```
Item: 13      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.06 t = 0.40 Outfit MNSQ = 1.04 t = 0.32

Categories             0      1 missing
Responses              0      1
Count                 14     75       0
Percent (%)        15.73  84.27
Pt-Biserial        -0.18   0.18
Mean Ability        0.54   0.87      NA
SD Abilities        0.34   0.35      NA
Step Difficulties         -1.01
Thresholds            NA     NA
Error                 NA     NA

=====================================================================

Item: 14      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.92 t = -0.46 Outfit MNSQ = 0.97 t = -0.12

Categories             0      1 missing
Responses              0      1
Count                  5     84       0
Percent (%)         5.62  94.38
Pt-Biserial        -0.08   0.08
Mean Ability        0.59   0.83      NA
SD Abilities        0.35   0.35      NA
Step Difficulties         -2.22
Thresholds            NA     NA
Error                 NA     NA

=====================================================================

Item: 15      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.95 t = -0.25 Outfit MNSQ = 0.94 t = -0.38

Categories             0      1 missing
Responses              0      1
Count                 12     76       1
Percent (%)        13.64  86.36
Pt-Biserial        -0.31   0.31
Mean Ability        0.32   0.92      NA
SD Abilities        0.34   0.35      NA
Step Difficulties         -1.16
Thresholds            NA     NA
Error                 NA     NA

=====================================================================
```

```
Item: 16      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.91 t = -0.60 Outfit MNSQ = 0.92 t = -0.52

Categories              0      1 missing
Responses               0      1
Count                  26     61       2
Percent (%)         29.89  70.11
Pt-Biserial         -0.32   0.32
Mean Ability         0.47   0.97      NA
SD Abilities         0.34   0.36      NA
Step Difficulties          -0.11
Thresholds             NA     NA
Error                  NA     NA

==================================================================

Item: 17      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.10 t = 0.70 Outfit MNSQ = 1.09 t = 0.62

Categories              0      1      2 missing
Responses               0      1      3
Count                  28     59      1       1
Percent (%)         31.82  67.05   1.14
Pt-Biserial         -0.14   0.13   0.05
Mean Ability         0.67   0.89   1.07      NA
SD Abilities         0.35   0.35   0.35      NA
Step Difficulties           0.01   5.29
Thresholds             NA   0.01   5.30
Error                  NA   0.13   0.48

===================================================================

Item: 18      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.00 t = 0.08 Outfit MNSQ = 0.99 t = -0.01

Categories              0      1 missing
Responses               0      1
Count                  14     74       1
Percent (%)         15.91  84.09
Pt-Biserial         -0.11   0.11
Mean Ability         0.61   0.86      NA
SD Abilities         0.34   0.35      NA
Step Difficulties          -0.94
Thresholds             NA     NA
Error                  NA     NA

===================================================================
```

```
Item: 19      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.04 t = 0.31 Outfit MNSQ = 1.03 t = 0.25

Categories              0      1 missing
Responses               0      1
Count                  40     47       2
Percent (%)         45.98  54.02
Pt-Biserial         -0.03   0.03
Mean Ability         0.80   0.83      NA
SD Abilities         0.35   0.35      NA
Step Difficulties           0.64
Thresholds             NA     NA
Error                  NA     NA

=====================================================================

Item: 20      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.90 t = -0.66 Outfit MNSQ = 0.91 t = -0.55

Categories              0      1 missing
Responses               0      1
Count                  22     65       2
Percent (%)         25.29  74.71
Pt-Biserial         -0.45   0.45
Mean Ability         0.27   1.00      NA
SD Abilities         0.34   0.36      NA
Step Difficulties          -0.33
Thresholds             NA     NA
Error                  NA     NA

=====================================================================

Item: 21      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.97 t = -0.10 Outfit MNSQ = 0.95 t = -0.25

Categories              0      1 missing
Responses               0      1
Count                   5     82       2
Percent (%)          5.75  94.25
Pt-Biserial         -0.24   0.24
Mean Ability         0.17   0.86      NA
SD Abilities         0.34   0.35      NA
Step Difficulties          -2.03
Thresholds             NA     NA
Error                  NA     NA

=====================================================================
```

Item: 22b Item Set: base Variable: PMM (by parameter)
Infit MNSQ = 1.17 t = 1.24 Outfit MNSQ = 1.19 t = 1.20

| Categories | 0 | 1 | 2 | 3 | missing |
|---|---|---|---|---|---|
| Responses | 0 | 1 | 2 | 3 | |
| Count | 25 | 20 | 38 | 5 | 1 |
| Percent (%) | 28.41 | 22.73 | 43.18 | 5.68 | |
| Pt-Biserial | -0.40 | -0.01 | 0.21 | 0.36 | |
| Mean Ability | 0.42 | 0.80 | 0.99 | 1.95 | NA |
| SD Abilities | 0.34 | 0.35 | 0.36 | 0.40 | NA |
| Step Difficulties | | 0.86 | 0.23 | 3.19 | |
| Thresholds | NA | 0.18 | 0.84 | 3.24 | |
| Error | NA | 0.22 | 0.33 | 0.43 | |

====================================================================

Item: 22c Item Set: base Variable: PMM (by parameter)
Infit MNSQ = 1.00 t = 0.05 Outfit MNSQ = 1.01 t = 0.14

| Categories | 0 | 1 | 2 | 3 | 4 | missing |
|---|---|---|---|---|---|---|
| Responses | 0 | 1 | 2 | 3 | 4 | |
| Count | 29 | 20 | 28 | 10 | 1 | 1 |
| Percent (%) | 32.95 | 22.73 | 31.82 | 11.36 | 1.14 | |
| Pt-Biserial | -0.37 | 0.07 | 0.02 | 0.31 | 0.33 | |
| Mean Ability | 0.46 | 0.91 | 0.83 | 1.43 | 3.53 | NA |
| SD Abilities | 0.35 | 0.35 | 0.35 | 0.37 | 0.49 | NA |
| Step Difficulties | | 1.01 | 0.54 | 2.15 | 3.79 | |
| Thresholds | NA | 0.34 | 0.97 | 2.19 | 3.95 | |
| Error | NA | 0.22 | 0.34 | 0.45 | 0.60 | |

====================================================================

Item: 23a Item Set: base Variable: PMM (by parameter)
Infit MNSQ = 0.98 t = -0.10 Outfit MNSQ = 0.97 t = -0.13

| Categories | 0 | 1 | 2 | 3 | missing |
|---|---|---|---|---|---|
| Responses | 0 | 1 | 2 | 3 | |
| Count | 12 | 40 | 24 | 11 | 2 |
| Percent (%) | 13.79 | 45.98 | 27.59 | 12.64 | |
| Pt-Biserial | -0.43 | -0.15 | 0.16 | 0.46 | |
| Mean Ability | 0.10 | 0.70 | 0.98 | 1.71 | NA |
| SD Abilities | 0.34 | 0.34 | 0.35 | 0.38 | NA |
| Step Difficulties | | -0.61 | 1.35 | 1.89 | |
| Thresholds | NA | -0.74 | 1.12 | 2.25 | |
| Error | NA | 0.14 | 0.43 | 0.46 | |

====================================================================

```
Item: 23b      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.85 t = -1.12 Outfit MNSQ = 0.83 t = -1.09

Categories              0      1      2      3 missing
Responses               0      1      2      3
Count                  18     30     35      4       2
Percent (%)         20.69  34.48  40.23   4.60
Pt-Biserial         -0.39  -0.17   0.34   0.35
Mean Ability         0.30   0.64   1.11   1.98      NA
SD Abilities         0.34   0.35   0.36   0.39      NA
Step Difficulties           0.10   0.71   3.34
Thresholds             NA  -0.23   0.97   3.41
Error                  NA   0.19   0.39   0.45

==================================================================

Item: 24a      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.99 t = -0.03 Outfit MNSQ = 0.97 t = -0.18

Categories              0      1      2      3      4 missing
Responses               0      1      2      3      4
Count                   5     23     44     15      2       0
Percent (%)          5.62  25.84  49.44  16.85   2.25
Pt-Biserial         -0.43  -0.04  -0.09   0.32   0.28
Mean Ability        -0.33   0.75   0.76   1.30   2.36      NA
SD Abilities         0.35   0.35   0.35   0.36   0.42      NA
Step Difficulties          -1.05   0.07   2.06   3.32
Thresholds             NA  -1.29   0.18   1.97   3.53
Error                  NA   0.22   0.57   0.49   0.53

==================================================================

Item: 24b      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.96 t = -0.26 Outfit MNSQ = 0.95 t = -0.27

Categories              0      1      2      3      4 missing
Responses               0      1      2      3      4
Count                  13     11     46     17      2       0
Percent (%)         14.61  12.36  51.69  19.10   2.25
Pt-Biserial         -0.37  -0.25   0.14   0.30   0.17
Mean Ability         0.21   0.37   0.90   1.26   1.60      NA
SD Abilities         0.34   0.34   0.35   0.37   0.38      NA
Step Difficulties           0.65  -0.70   1.97   3.43
Thresholds             NA  -0.31   0.13   1.88   3.61
Error                  NA   0.28   0.37   0.47   0.51

==================================================================
```

Item: 25a      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 0.98 t = -0.06 Outfit MNSQ = 1.00 t = 0.04

| Categories | 0 | 1 | 2 | 3 | 4 | missing |
|---|---|---|---|---|---|---|
| Responses | 0 | 1 | 2 | 3 | 4 | |
| Count | 1 | 11 | 69 | 7 | 1 | 0 |
| Percent (%) | 1.12 | 12.36 | 77.53 | 7.87 | 1.12 | |
| Pt-Biserial | -0.24 | -0.19 | -0.01 | 0.20 | 0.33 | |
| Mean Ability | -0.66 | 0.46 | 0.82 | 1.25 | 3.53 | NA |
| SD Abilities | 0.35 | 0.35 | 0.35 | 0.37 | 0.49 | NA |
| Step Difficulties | | -1.94 | -1.16 | 3.26 | 3.29 | |
| Thresholds | NA | -2.23 | -0.88 | 2.80 | 3.76 | |
| Error | NA | 0.38 | 0.96 | 0.63 | 0.69 | |

====================================================================

Item: 25b      Item Set: base      Variable: PMM (by parameter)
Infit MNSQ = 1.01 t = -19.54 Outfit MNSQ = 1.01 t = -34.02

| Categories | 0 | 1 | 2 | 3 | 4 | missing |
|---|---|---|---|---|---|---|
| Responses | 0 | 1 | 2 | 3 | 4 | |
| Count | 8 | 47 | 28 | 3 | 1 | 2 |
| Percent (%) | 9.20 | 54.02 | 32.18 | 3.45 | 1.15 | |
| Pt-Biserial | -0.31 | -0.02 | 0.09 | 0.18 | 0.22 | |
| Mean Ability | 0.19 | 0.81 | 0.94 | 1.50 | 2.37 | NA |
| SD Abilities | 0.34 | 0.35 | 0.35 | 0.37 | 0.41 | NA |
| Step Difficulties | | -1.17 | 1.38 | 3.37 | 2.59 | |
| Thresholds | NA | -1.24 | 1.31 | 2.79 | 3.36 | |
| Error | NA | 0.04 | 0.49 | 0.55 | 0.70 | |

====================================================================

The following are raw score statistics.
Missing responses are treated as scores of 0.

| | Raw | Percent |
|---|---|---|
| Mean test score | 28.65 | 55.09% |
| Standard deviation | 5.58 | 10.73% |
| Student Count | 89 | |
| Cronbach's Alpha | 0.7 | |
| Missing Data Percentage | 0.81% | |

====================================================================

The following statistics include complete cases only.

| | |
|---|---|
| Cronbach's Alpha | 0.71 |
| Student Count | 76 |

# Appendix E: Embedded Assessments and Scoring Rubric

| Name | Date |
|---|---|
| | |

# Activity 1.1

## Can you Smell What I Smell?

### PURPOSE

In class, your teacher opened a jar that had an object in it. The object in the jar had an odor to it, and the odor moved across the room. In this activity, you will record your ideas about how an odor can go from an object to your nose.

### DATA COLLECTION/OBSERVATION

1. Write the odor you smell

______________________________________________

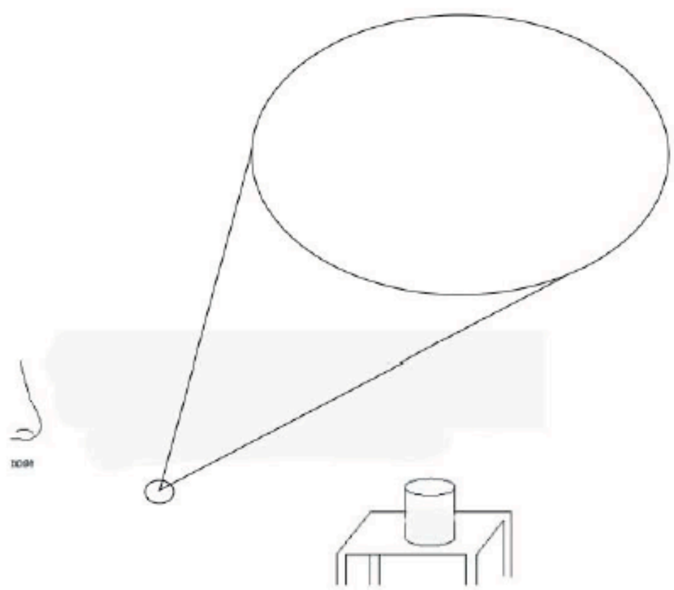
2. Describe what happens when your teacher opens the jar.

______________________________________________

______________________________________________

______________________________________________

3. Imagine that you have a very powerful microscope that would allow you to see the odor up really, really close. What would you see? Create a drawing that shows how the odor got from the source to your nose. The large circle in the drawing below represents the magnified part of the air between the jar and your nose. In the circle, draw a picture of what you think the odor looks like between the jar and your nose.




4. Label what the parts in your drawing represent.

Name Date

# Activity 1.1

5. Now, imagine that a friend of yours was looking at your drawing. In the space below, describe for your friend how your drawing helps to explain that odors travel from the source to your nose.

Name Date

## Activity 5.2

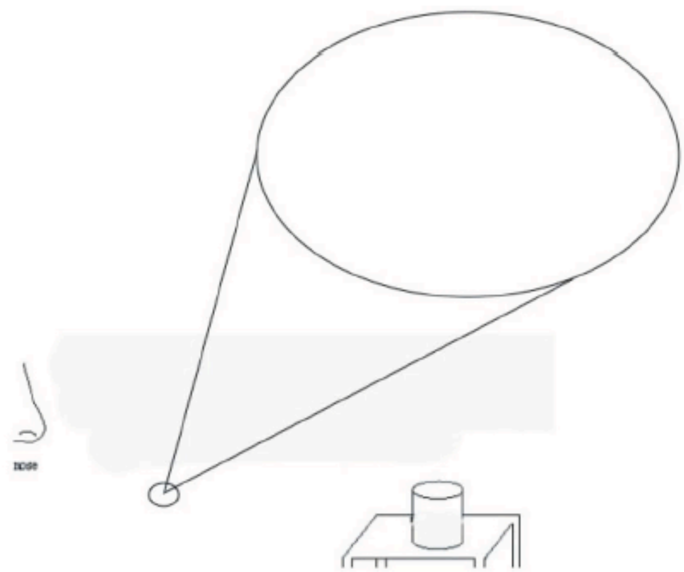
# How Does the Odor Get to My Nose?

### PURPOSE

In class, you observed indicator paper changing color without dipping it into the liquids in the flasks. Then, you watched a computer simulation of odor particles traveling. The purpose of this activity is to create new models or revise your model that explain how an odor moves across the room.

### PROCEDURE/CREATING MODELS

Imagine you have a very powerful microscope that allows you to see the odor and air up really, really close. What would you see? Create a model that can help to explai how you can smell an odor from a distance. The large circle in the drawing below represents the magnified part of the air between the source and your nose. In the circle, draw a picture of what you think the odor looks like between the jar and your nose.




In this box write what the symbols in your model represent.

Key:

Name Date

## Activity 5.2, *continued*

1. What does the model represent?

2. Imagine a friend of yours was looking at your model. In the space below, describe for your friend how your model can be used to explain what happened.

3. What part(s) of the phenomena is the model unable to show?

### FOLLOW-UP QUESTIONS

1. How is this model different from the first model you created in lesson 1?

2. What new ideas have you included in this model?

GO

Name ____________ Date ____________

## Activity 15.1, *continued*

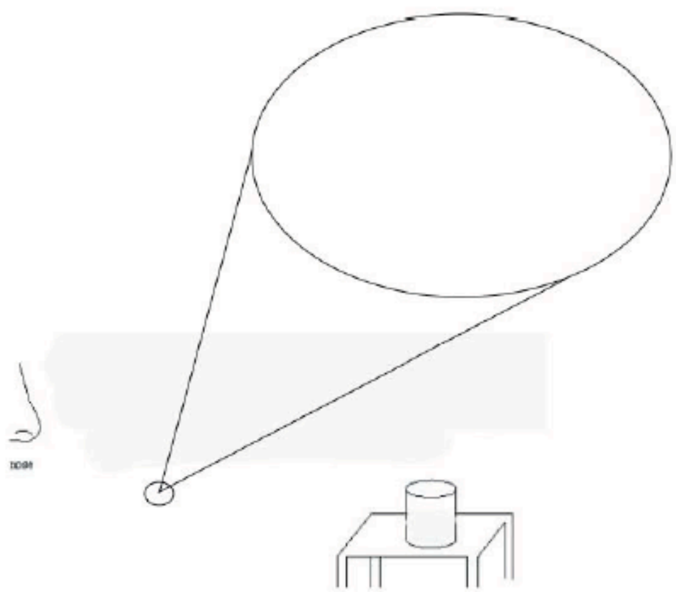
# What Did We Explain with Our Model?

### CONNECTION

In lesson one you constructed a model to represent an odor traveling across the classroom from a jar to your nose. Throughout this unit, you have observed many phenomena to learn more about the particle nature of matter. In this lesson you will construct another model of how odors travel using all of the information you learned in this unit.

### A. CONSTRUCTING INDIVIDUAL MODELS

1. Imagine that you have a very powerful microscope that would allow you to see the odor up really, really close. What would you see? Create a model that can help to explain how you can smell an odor from a distance. The large circle in the drawing below represents the magnified part of the air between the source and your nose. In the circle, draw a picture of what you think the odor looks like between the jar and your nose.



2. Label the parts in your drawing.

Name Date

## Activity 15.1, *continued*

3. Now, imagine that a friend of yours from a different science class was looking at your drawing. Describe for your friend how your model helps to explain that odors travel from the source to other places.

4. Compare the model you constructed today to the models you constructed during lesson one and lesson 5. How has your model changed? Why?

| Code | Drawing portion of model |
|---|---|
| | No drawing |
| **0** | Descriptive – describes in words or macro symbols of gas/air; smaller version of phenomena |
| **1** | Mixed Model<br>- Particles and Descriptive (i.e. air particles, but waves of odor)<br>- Descriptive can also be including unnecessary macrolevel objects<br>- Movement may be included |
| **2** | Basic Particle Model<br>- air particles and/or odor particles<br>- Motion may be included |
| **3** | Complete Particle Model<br>- Air particles<br>- odor particles<br>- random motion (movement in all directions; collisions between particles) is required |

| Code | Explanation part of model | |
|---|---|---|
| | No response | |
| **0** | Descriptive – describes model OR describes what happened in class OR gives completely incorrect explanation OR uses prior experience to explain what is happening. | It shows how odor traval through air. OR The lines were ammonia and little circles are air particals and arrows were movement. |
| **1** | Mixed Model<br>Student tries to explain odors traveling from the source to the nose, but uses the incorrect mechanism or focuses on a macrolevel. | The air and scent go f aster more heat and slower less heat. OR<br>The fan blows air into the air blowing over the tuna smell picking up the smell traveling in a straight path to the nose. OR<br>The odor molecules mix in the air and flow up the nose. |
| **2** | Basic Particle Model<br>Student is able to explain that odor molecules travel from the source to the nose. Student may explain how air helps in this process. | The oder is in a gaseous state. The air and odor molecules spread around the room. |
| **3** | Complete Particle Model<br>Student is able to correctly explain on a microscopic level the movement: odor particles travel in air, random collisions of odor and air molecules. Student may also describe sublimation/evaporation from the source on a microlevel. | First the particles gain enough energy to evaporate and turn to gaseous ammonia. Then it moves in a straight path tell it runs into something. Eventually it travels to a nose. |

# REFERENCES


American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

American Educational Research Association, American Psychological Association, National Council for Measurement in Education. (1999). *Standards for educational and psychologicaltesting*. Washington, DC: American Psychological Association.

Baker, S., & Talley, L. (1972). The relationship of visualization skills to achievement in freshman chemistry. Journal of Chemical Education, 49(11), 775-776.

Barak, M. & Dori, Y. (2001). Virtual and physical molecular modeling: fostering model perception and spatial understanding. *Educational Technology & Society*, 4(1), 61-74.

Ben-Zvi, R., Eylon, B., & Silberstein, J. (1987). Is an atom of copper malleable? *Journal of Chemical Education*, 63, 64-66.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*,18, 32-42.

Claesgens, J., Scalise, K. Wilson, M., & Stacy, A. (2008). Mapping student understanding in chemistry: The perspectives of chemists. *Science Education*, 93, 56-85.

Corcoran, T., Mosher,F.A., & Rogat, A. (2009). *Learning Progressions in Science: An evidence-based approach to reform*. Center on continuous Instructional Improvement. Teachers College, Columbia University.

Dagher, Z. & Cossman, G. (1992). Verbal explanations given by science teachers: Their nature and implications. *Journal of Research in Science Teaching*, 29, 361-374.

Davis, E.A., & Krajcik, J. (2005). Designing educative curriculm materials to promote teacher learning, *Educational Researcher*, 34(3), 3-14.

deVos, W. & Verdonk, A.H. (1996). The particulate nature of matter in science education and in science. *Journal of Research in Science Teaching*, 33, 557-664.

Driver, R., Guesne, E., & Tiberghein, A. (1985). *Children's ideas in science*. Philadelphia: Open University Press.

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

Duschl, R., Schweingruber, H. & Shouse, A. (Eds.) (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: The National Academies Press.

Ealy, J. (1999). A student evaluation of molecular modeling in first year college chemistry. *Journal of Science Education and Technology*, 8(4), 309-321).

Embretson, S. & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ferguson, R., & Bodner, G. (2006). Misconceptions held by chemistry majors while taking organic chemistry. Paper presented at the 2006 Annual Meeting of the National Association of Research in Science Teaching, San Francisco, CA.

Gilbert, S. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*. 28, 73-99.

Gilbert, J. & Boulter, C. (1998). Models in explanations, part 1: Horses for courses? *International Journal of Science Education*, 20, 83-97.

Gilbert, J. & Boulter, C. (1998). Models in explanations, part 2: Whose voice? Whose ears? *International Journal of Science Education*, 20, 187-203.

Grosslight, L., Unger, C., Jay, E., & Smith, C.L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching*, 29, 799-822.

Haertel, E. H. (1999). Performance assessment and education reform. *Phi Delta Kappan*, 80 (May), 662-666.

Hambleton, R. & Jones, R. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.

Harrison, A., & Treagust, D. (1996). Secondary students' mental models of atoms and molecules: implications for teaching chemistry. *Science Education*, 80, 509-534.

Harrison, A., & Treagust, D. (1998). Modelling in science lessons: are there better ways to learn with models? *School Science and Mathematics*, 98, 420-429.

Harrison, A., Treagust, D. (2000). Learning about atoms, molecules, and chemical bonds: a case study of multiple-model use in grade 11 chemistry. *Science Education*, 84, 352-381.

Harrison, A. & Treagust, D. (2002). The particulate nature of matter: Challenges in understanding the submicroscopic world. In J.K. Gilbert et al. (Eds.), *Chemical education: Towards research-based practice* (pp. 189-212). Boston: Kluwer Academic Publishers.

Hestenes, D. (1992). Modeling games in the Newtonian World. *American Journal of Physics*, 60(8), 732-748.

Holt, Rinehart, Winston (2006). Modern Chemistry. Austin, Texas: Holt, Rinehart and Winston.

Johnson, P. (1998). Progression in children's understanding of a 'basic' particle theory: a longitudinal study. *International Journal of Science Education*, 20(4), 393-412.

Johnson, P. & Papageorgiou (2010). Rethinking the introduction of particle theory: A substance-based framework. *Journal of Research in Science Teaching*, 47(2), 130-150.

Johnson, P. & Tymms, P. (2010). *Using Rasch modeling on a large cross-sectional data-set to test for a learning progression in chemistry suggested by a previous, small-scale, three year longitudinal study.* Paper presented at the Annual National Association for Research in Science Teaching in Philadelphia, PA.

Justi, R. & Gilbert, J. (2002). Models and modeling in chemical education. In J.K. Gilbert et al. (Eds.), *Chemical education: Towards research-based practice* (pp. 47-68). Boston: Kluwer Academic Publishers.

Kennedy, C., Brown, J., Draney, K., & Wilson, M. (2006). *Using progress variables and embedded assessment to improve teaching and learning.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Kenney, P.A. (2000). The seventh NAEP mathematics assessment: Overview. In E.A. Silver & P. A. Kenney (Eds.), *Results from the seventh mathematics assessment of the national assessment educational progress* (pp. 1-22). Reston, VA: National Council of Teachers of Mathematics.

Kozma, R., Chin, E., Russell, J. & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. The *Journal of the Learning Sciences*, 9(2), 105-143.

Krajcik, J., McNeill, K., & Reiser, B. (2008). Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education*

Krajcik, J.S. & Blumenfeld, P. (2006). Project-based learning. To appear in Sawyer, R. K. (Ed.), the *Cambridge Handbook of the Learning Sciences*. New York: Cambridge.

Liu, X., & Lesniak, K. (2006). Progression in childrens' understanding of the matter concept from elementary to high school. *Journal of Research in Science Teaching*, 43(3), 320-347.

MacKinnon, G. (2003). Why models sometimes fail. *Journal of College Science Teaching*, 32(7), 430-433.

Margel, H., Eylon, B., & Scherz, Z. (2008). A longitudinal study of junior high school students' conceptions of structure of materials. *Journal of Research in Science Teaching*, 45(1), 132-152.

Merritt, J., Rogat, A., & George, A. (2006). *Focus on Modeling: A Curriculum Approach to Learning the Particle Nature of Matter.* Paper presented at the Annual National Association for Research in Science Teaching in San Francisco, CA.

Merritt, J., Schwartz, Y. & Krajcik, J. (2007). *Middle School Students' Development of the Particle Model of Matter.* Paper presented at the Annual National Association for Research in Science Teaching in New Orleans, LA.

Merritt, J., Krajcik, J., & Schwartz, Y. (2008). *Development of a learning progression for the particle model of matter*. Proceedings from the 8th International Conference of the Learning Sciences in Utrecht, the Netherlands.

Metz, K. (1991). Development of explanation: Incremental and fundamental change in children's physics knowledge. *Journal of Research in Science Teaching*, 28, 785-797.

Mikelskis-Seifert, S., & Leisner, A. (2005). Investigation of effects and stability in teaching model competence. In K. Boersma et al. (Eds.), *Research and the quality of Science Education* (pp. 337-351). New York: Kluwer Academic Publishers.

Minstrell, J. (2001). Facets of students' thinking: Designing to cross the gap from research to standards-based practice. In K. Crowley, C.D. Shunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 415-443). Mahwah, NJ: Erlbaum.

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

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

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Niaz, M. Aguilera, D., Maza, A. & Liendo, G. (2002). Arguments, contradictions, resistances, and conceptual change in students' understanding of atomic structure. *Science Education*, 86, 505-525.

Novick, S., & Nussbaum, J. (1978). Junior high school pupils' understanding of the particulate nature of matter: An interview study. *Science Education*, 62, 273-281.

Nussbaum, J. & Novick, S. (1982). Alternative frameworks, conceptual conflict and accommodation: Toward a principled teaching strategy. *Instructional Science*, 11, 183-200.

Nussbaum, J. (1985). The particulate nature of matter. In R. Driver, E. Guesne, & A. Tiberghien (Eds.), *Children's ideas in science* (pp. 124-144). Philadelphia: Open University Press.

Osborne, R. & Freyberg, P. (1985). Learning in science. Portsmouth, NH: Heinemann Publishers.

Patton, M.Q. (2002). *Qualitative research and evaluation methods* (3rd edition). Thousand Oaks, CA: Sage.

Renstrom, L., Andersson, B., & Marton, F. (1990). Students' conceptions of matter. *Journal of Educational Psychology*, 82(3), 555-569.

Russ, R.S., Scherr, R.E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, 92(3), 499-525.

Saari, H., & Viiri, J. (2003). A research-based teaching sequence for teaching the concept of modeling to seventh-grade students. *International Journal of Science Education*, 25(11), 1333-1352.

Schwarz, C., Reiser, B., Davis, E. A., Kenyon, L., Acher, A., Fortus, D., Shwartz, Y., Hug, B. & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632-654.

Schwarz, C., & White, B. (2005). Metamodeling knowledge: developing students' understanding of scientific modeling. *Cognition and Instruction*, 23, 165-205.

Shin, N., Stevens, S.Y., Short, H., & Krajcik, J. (2009) Learning progressions to support coherence curricula in instructional material, instruction, and assessment design.

Paper presented at the Learning Progressions in Science Conference in Iowa City, Iowa.

Smith, C., Wiser, M., Anderson, C., & 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. *Measurement*, 14(1&2), 1-98.

Smith, J., diSessa, A., & Roschelle, J. (2003). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3(2), 115-163.

Songer, N., Kelcey, B., & Gotwals, A. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, 46(6), 610-631.

Stavy, R. (1991). Children's ideas about matter. *School Science and Curriculum*, 91, 240-244.

Stevens, S.Y., Delgado, C., & Krajcik, J. (2010). Developing a Hypothetical Multi-Dimensional Learning Progression for the Nature of Matter. *Journal of Research in Science Teaching*, 47(6), 687-715.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. A. Duschl & R. J. Hamilton (Eds.) *Philosophy of science, cognitive psychology, and educational theory and practice* (pp.147 – 176), Albany: State University of New York Press.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, 4, 45-69.

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: making science accessible to all students. *Cognition and Instruction*, 16, 3-118.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, 46(6), 716-730.

Wilson, M. (2005). *Constructing Measures: An item response modeling approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, M., & Sloan, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13, 181-208.

Wu, W.-K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, 88(3), 465-492.