A Hypothetical Learning Progression for Quantifying Phenomena in Science

Hui Jin[a], Cesar Delgado[b], Malcolm I. Bauer[a], E. Caroline Wylie[a], Dante Cisterna[a], and Kenneth F. Llort[a]

[a] Educational Testing Service, Princeton, NJ

[b] North Carolina State University, Raleigh, NC

Jin, H., Delgado, C., Bauer, M. I., Wylie, E. C., Cisterna, D. I., & Llort, K. F. (2019). A hypothetical learning progression for quantifying phenomena in science. *Science & Education.* https://doi.org/10.1007/s11191-019-00076-8.

Acknowledgement

This work was supported by the Institute of Education Sciences under Grant R305A160219.

A Hypothetical Learning Progression for Quantifying Phenomena in Science

Abstract

In this article, we report on a three-pronged effort to create a hypothetical learning progression for quantification in science. First, we drew from history and philosophy of science to define the quantification competency and develop hypothetical levels of the learning progression. More specifically, the quantification competency refers to the ability to analyze phenomena through (a) abstracting relevant measurable variables from phenomena and observations, (b) investigating the mathematical relationships among the variables, and (c) conceptualizing scientific ideas that explain the mathematical relationships. The quantification learning progression contains four levels of increasing sophistication: Level 1, holistic observation; Level 2, attributes; Level 3, measurable variables; and Level 4, relational complexity. Second, we analyzed the practices in the Next Generation Science Standards for current, largely tacit, assumptions about how quantification develops (or ought to develop) through K-12 education. While several pieces of evidence support the learning progression, we found that quantification was described inconsistently across practices. Third, we used empirical student data from a field test of items in physical and life sciences to illustrate qualitative differences in student thinking that align with levels in the hypothetical learning progression for quantification. By generating a hypothetical learning progression for quantification, we lay the groundwork for future standards development efforts to include this key practice and provide guidance for curriculum developers and instructors in helping students develop robust scientific understanding.

Keywords: assessment, learning progression, quantification

1 Introduction

There is a consensus that students should learn not only the products of science but also the process of doing science (NGSS Lead States, 2013; National Research Council [NRC], 1996, 2000, 2012; Osborne, 2010). In doing science, scientists spend significant amount of time and put considerable effort into coordinating theory and evidence (D. Kuhn & Pearsall, 2000). The coordination of theory and evidence involves many epistemic practices that are essential for science education at K-12 levels. These practices include dealing with variables (D. Kuhn & Dean, 2004); transforming, evaluating, and interpreting data (Duncan, Chinn, & Barzilai, 2018; McNeill & Berland, 2017); dealing with anomalies (Chinn & Brewer, 1993); using data to construct and revise models (Lehrer & Schauble, 2006); using data to support and evaluate arguments (McNeill & Krajcik, 2011); and so forth. Meaningful learning of these epistemic practices must focus on learning scientific thinking and reasoning rather than procedures or behaviors (Duschl, 2000). Researchers in the psychology of science have generated important findings about reasoning patterns that are required for the coordination of theory and evidence (see Dunbar & Fugelsang, 2005; Zimmerman, 2000). Some of those reasoning patterns are hypothetico-deductive reasoning (Lawson, 2004), causal reasoning (Cheng, 1997), and analogical reasoning (Dunbar, 2001).

In this article, we focus on one reasoning pattern used in the coordination of theory and evidence—mathematization or quantification of science. This reasoning pattern is essential for the generation of many important scientific concepts, theories, and ideas. We use the terms *mathematization* and *quantification* interchangeably because both terms are used in the literature. The term *mathematization* is commonly used in literature on the history and philosophy of science, whereas *quantification* is more commonly used in science education. Although only a few researchers have studied mathematization or quantification in science education (e.g., Lehrer & Schauble, 2006), its importance has been well recognized in the literature of philosophy and history of science. A consensus is that mathematical descriptions allow precise prediction and provide relatively objective bases for scientific argumentation and discussion (Holton & Brush, 2006; Kline, 1980; Osborne, Rafanelli, & Kind, 2018).

Thomas Kuhn's (1962) pioneering work in history and philosophy of science provides further information about the role of mathematics in scientific investigations. According to Kuhn, when a scientist approaches a new field, he or she must first determine "what aspects of the complex phenomenon" are relevant (p. 4). The main work of scientists is to study "fundamental entities of which the universe is composed" and how those entities "interact with each other and with the senses" (pp. 4–5). As such, identifying variables and exploring the relationships among variables are crucial in the development of scientific ideas in the history of science. Therefore, we define *quantification competency* as the ability to analyze phenomena through (a) abstracting relevant measurable variables from phenomena and observations, (b) investigating the mathematical relationships among the variables, and (c) conceptualizing scientific ideas that explain the mathematical relationships.

Given the importance of quantification to science and to science learning, it is critical to study how students might gradually learn to quantify and mathematize in science. We use a learning progression (LP) approach to study this issue. LPs are "descriptions of the successively more sophisticated ways of thinking about a topic that can follow one another as children learn about and investigate a topic over a broad span of time" (NRC, 2007, p. 219). They can lead to improved standards, curricula, instruction, and assessments, as well as better student outcomes (Corcoran, Mosher, & Rogat, 2009). We report on a three-pronged effort to create this LP. First, we developed a hypothetical LP for quantification based on literature from the history and philosophy of science. Second, we analyzed the Next Generation Science Standards (NGSS; NGSS Lead States, 2013) for current, largely tacit, assumptions about how quantification develops (or ought to develop) through K-12 schooling. We compared the developmental trends described in NGSS with the hypothetical LP for quantification. This work provides evidence that supports the LP levels, but it also indicates an inconsistency in the way quantification is described in NGSS. Third, we used empirical student data from written assessment items around topics in physics (energy) and the life sciences (ecosystems) to illustrate LP levels. By generating an LP for quantification, we lay the groundwork for future standards development efforts to include this key practice and provide guidance for curriculum developers and instructors responsible for guiding students to robust scientific understanding.

It is important to state clearly our conception of LPs at the outset. We recognize that not all students develop competencies following the same path and that students' thinking is context-dependent and emergent in many cases. Students may well think in advanced ways in one context but not another, and progress is not always linear. Our goal in developing an LP is thus to characterize qualitatively different ways of reasoning used in quantification that can be ordered in the degree of sophistication and similarity to accepted scientific thinking. This set of levels can be used to guide teachers in recognizing student ideas, to help curriculum developers determine instructional approaches, to inform grade band goals in standards, and to develop assessments, without being prescriptive about individual students' trajectories. In developing progressions with these characteristics and for these purposes, one must manage tensions between (a) identifying meaningful patterns in learning and (b) supporting students' learning while not over-constraining it. A good progression identifies meaningful conceptual shifts, enrichment, and integration that take place in students slowly and incrementally over weeks, months, or even years (Jin, Mikeska, Hokayem, & Mavronikolas, 2019). Knowing the kinds of understanding students currently have can affect the nature of learning not just with respect to the specific concept, but also may provide a lens into how students view and learn other concepts.

Hammer and Sikorski (2015) provided a critique of LPs. They pointed out that LPs cannot capture the fragmentation, contextualization, and dynamics of learning. More specifically, students may hold many fragmented pieces of knowledge and conceptions. Different pieces could be activated in different contexts. As such, learning is messy and dynamic; it is not linear. Similar to Lesh, Lamon, Gong, and Post's (1992) notion of a learning progress map, Hammer and Sikorski viewed performance as the result of a dynamic process that may react very differently to small changes in the environment. We agree that learning about science is complex. However, we believe patterns in learning and development can be articulated at relatively coarse grain sizes. We propose that while LPs are not sufficient in explaining all complex, emergent behavior during learning, if expressed at an appropriate grain size, the approach can identify patterns of understanding and behavior that are instructionally helpful. We view this as a design challenge to find grain sizes for the conceptual shifts that, though they may be emergent and manifest in different ways under different conditions, are persistent and can be affected over time by learning and instruction. This notion of LPs as expressing significant shifts in understanding at a coarse grain-size that are useful for instruction is consistent with many prior definitions and discussions of LPs including Black, Wilson, and Yao (2011), Corcoran et al. (2009), and Heritage (2008).

2 Development of the Hypothetical Learning Progression

Existing research of quantification can be divided into two groups. In one group, researchers treat quantification as a domain-general competency (Adamson et al., 2003; Lawson, 1983; Vass, Schiller, & Nappi, 2000). They study how students apply mathematical concepts such as proportion, probability, and correlation to science contexts, but the application does not require conceptual understanding of scientific knowledge. In the other group, researchers treat quantification as a domain-specific competency that is intertwined with scientific knowledge. As our definition of the quantification competency emphasizes how mathematical concepts and thinking are used in conceptualization of scientific ideas, we focused our review on the literature in the latter group.

In that group, an important finding is the different ways of thinking that experts and novice employ at three phases of problem solving (Chi, Feltovich, & Glaser, 1981; Kuo, Hull, Gupta, & Elby, 2013; Niss, 2017; Tuminaro & Redish, 2007). At the beginning, experts establish a conceptual story of a phenomenon and translate that conceptual story into mathematical forms; this step often involves identifying the underlying fundamental principle involved. Next, in mathematical processing, experts perform mathematical operations that are meaningful in science. Toward the end, experts generate scientific interpretations of the mathematical results. Unlike experts, novices seldom carry out the first step of establishing a conceptual understanding of the phenomenon. Instead, they start directly at the second step—mathematical processing. In this process, novices identify relevant mathematical symbols and equations based on surface features of the problem; they plug numbers into equations to calculate the target quantities. As a result, novices often do not apply the appropriate equations. Novices seldom carry out the third step of expert reasoning, which involves constructing a conceptual interpretation of the mathematical results. Another important finding comes from research into using graphs in science. Most of these studies were conducted in kinematics. Researchers found that students tend to confuse graphs with the real world. For example, students tend to use graphs about motion (distance-time graph, velocity-time graph, acceleration-time graph, etc.) as the picture of the motion (Kozhevnikov, Motes, & Hegarty, 2007; McDermott, Rosenquist, & van Zee, 1987). Understanding the scientific meanings of the variables in the graph is also challenging for students. For example, students often do not know the scientific meaning of slope and confuse slope with the height of a graph (Planinic, Milin-Sipus, Katic, Susac, & Ivanjek, 2012). In general, this body of literature suggests that students have difficulty in identifying variables in real phenomena and in graphs; they often do not understand scientific meanings of variables and the relationships amongst them. However, it does not provide enough information for us to hypothesize the developmental trend. More specifically, what are the qualitatively different achievement levels that students may experience?

The parallels between disciplinary and individual trajectories have been noted in the past (e.g., Ha & Nehm, 2014; Kuhn, 1962). Therefore, one research approach is to study the historical development of scientific ideas to shed light on students' development (McComas, Clough, & Almazroa, 1998; Wiser & Carey, 1983). We use this approach to begin the iterative development of the hypothetical LP for quantification. As T. Kuhn (1962) pointed out, long periods of "normal science" are interspersed by "scientific revolutions" that result in paradigm change; revolutions are spurred by anomalies that cannot be adequately explained by the existing paradigm's theory and methods. Events in the history of science suggest that mathematization plays an important role in both developing normal science and spurring scientific revolution (Kline, 1964). In this section, we examine quantification in five historical events across physics, biology, astronomy, and chemistry. We focus on how measurement and quantification enabled the generation of fundamental ideas in science disciplines. These fundamental ideas are also the core ideas in K-12 science curriculum (NRC, 2012). Among the five events, three focus on quantification in normal science. Examining these events allowed us to identify key features of mathematization or quantification in science. Two events are about quantification in scientific revolutions. Examining them allowed us to identify the conceptual shifts toward scientific quantification, which provides ideas for us to hypothesize the LP levels.

2.1 Quantification in Normal Science

Our first example of quantification in normal science is the development of the ideal gas law (Altig, 2014; Holton & Brush, 2006). In the 17th century, Boyle studied the compressibility of air quantitatively. In his experiment, a given mass of air was trapped in a J-shaped tube filled with mercury. The short arm of the tube was closed and contained the trapped air. The long arm of the tube was open, so that mercury could be poured into the tube. Boyle measured the volume of the trapped air (V) and the air pressure (P). He noted that, for a given mass of air trapped in the tube at a constant temperature, the product of V and P is a constant. About a hundred years later, scientists studied how the volume of different gases changed with temperature when pressure was held constant. Charles and Gay-Lussac found that, for different gases at constant pressure, the volume is proportional to temperature (Charles and Gay-Lussac's law). In the early 19th century, Avogadro made a hypothesis that equal volumes of different gases at the same temperature and pressure contain an equal number of gas particles. However, this hypothesis seemed inconsistent with Gay-Lussac's other observation that two volumes of hydrogen react with one volume of oxygen to form two volumes of water vapor. Assuming equal volumes, and thus equal numbers of particles, two volumes of hydrogen (particles: 2H) and one volume of oxygen (particles: 1O) should produce one volume rather than two volumes of water vapor ($2H + 1O = 1H_2O$). This inconsistency was resolved by making the assumption that the characteristic particles of gases are molecules, rather than atoms. Assuming hydrogen has the formula of H_2 and oxygen has the formula of O_2, Gay-Lussac's observation is consistent with Avogadro's hypothesis: Two volumes of H_2 and one volume of O_2 produce two volumes of H_2O ($2\ H_2 + 1\ O_2 = 2\ H_2O$). The ideal gas law ($PV = nRT$) was generated by combining these three crucial findings—Boyle's law, Charles and Gay-Lussac's law, and Avogadro's hypothesis. It describes the relationships among three variables of any given sample of gas—volume, pressure, and temperature—under ordinary conditions. The relationship among the variables additionally allowed for the definition of the idea gas constant, R. In this historical case, we see the progression from observing attributes (compressibility), to defining variables that provide measurements of attributes, to relationships among variables (Altig, 2014; Holton & Brush, 2006).

The second example is Mendel's discovery of the laws of hybridization (Allen, 2003; Gayon, 2016; Kampourakis, 2013). Mendel studied the hybridization of pea plants. He observed the hybridization patterns of seven pairs of physical characteristics in pea plants: plant height (tall vs. short), seed shape (round vs. wrinkled), flower color (purple vs. white), and so forth. Through self-pollination of the plants, Mendel obtained pure lines of pea plants for each characteristic. He then conducted a sequence of hybridization experiments on these pure line plants. Take flower color as an example. Mendel cross-pollinated pure line plants that produced purple flowers with those that produced white flowers. He found that the first generation all had purple flowers, while the second generation had an approximate 3:1 ratio of purple flowered plants to white flowered plants.[1] Additional cross-pollination experiments showed that the offspring of the white-flowered plants did not vary further; two thirds of the purple-flowered plants yielded an approximate 3:1 ratio of purple to white; and one third of the purple-flowered plants yielded purple-flowered offspring only. To explain these patterns, Mendel differentiated between two contrasting conditions: dominant and recessive. The character appeared in the first generation was dominant (e.g., purple flower), while the character that did not appear in the first generation was recessive (e.g., white flower). Although Mendel's intention was to explore patterns in hybridization rather than laws of heredity, his patterns, which were rediscovered by other scientists in 1900, suggested the existence of an entity controlling the expression of the characters; this entity was later conceptualized as gene (Gayon, 2016; Kampourakis, 2013). In this historical case, we again see the observation of characteristics (e.g., flower color), followed by the development of quantitative accounts of specific traits (ratios of purple to white flowers). Another important development here, the proposal of elements received from parents and the nature of these (dominant, recessive) are conceptual and mechanistic rather than having to do with quantification.

[1] The debate regarding the validity of Mendel's data is important but beyond the scope of this article.

A third example is the derivation of universal gravitation from Kepler's laws of planetary motion. Tycho Brahe, a Danish astronomer, collected what was at that time the most accurate and voluminous data on the positions and movements of stars, planets, and comets. To achieve accuracy, Brahe designed specialized instruments, built the instruments in an underground observatory, and performed calibration regularly in the process of data collection. Brahe's student, Kepler spent a lifetime analyzing these voluminous data sets and identified three mathematical laws about the planetary motion. The first law states that the orbits of planets are ellipses with the Sun at one focus. The second law states that a line connecting the Sun and the planet sweeps out equal areas in equal intervals of time. The third law states that the square of a planet's orbital period is proportional to the cube of its average distance from the Sun. Newton believed that these mathematical patterns must have a conceptual reason. He proposed the notion of universal gravitation to explain Kepler's laws. Holton and Brush (2006) described four crucial steps in Newton's conceptualization. In the first step, from Kepler's first law, Newton inferred that a net force must be exerted on the planet; otherwise, the planet would travel in a straight line rather than in an ellipse. In the second step, based on Kepler's second law, Newton constructed the mathematical proof that the force exerted on the planet must be a centripetal force. In the third step, from Kepler's third law, Newton derived that the centripetal force at any instant must be proportional to the inverse square of the distance between the planet and the Sun. In the final step, Newton searched for the origin of the centripetal force. He hypothesized that the centripetal force exerted on the planet is the gravitational attraction from the Sun. In other words, a universal gravitational force exists; the same type of attractive force exists between the Sun and its planets, between the Earth and the Moon, and between the Earth and a falling apple. While there were at that time other hypotheses about the nature of the centripetal force (e.g., magnetic attraction from the Sun; space being filled with invisible fluid), Newton proved that those hypotheses could not account for the mathematical patterns identified by Kepler. Newton further proved that Kepler's third law is the mathematical consequence of the gravitational force between the Sun and its planet. This historical case began with observation of attributes (regular motion), then definition of variables to measure (distance and time), then relationship among variables (Kepler's laws), and then another step to establish relationships among other variables explaining Kepler's Laws (Newton's universal gravitation).

These three examples suggest that scientific concepts, principles, and theories are generated to explain the quantitative descriptions of natural phenomena. The quantitative descriptions have three key features: relevance, measurability, and relational complexity. First, a phenomenon under investigation may have many aspects or attributes; it is important to identify and select relevant variables to investigate. This is a process of abstracting variables from messy phenomena. In the investigation of gas laws, scientists focused on temperature, volume, and pressure. Kepler focused on two variables of planetary motion—distance and time. Mendel focused on the number and ratios of plants with different traits.

Second, accurate measurement ensures that the mathematical patterns identified from the data are valid and reliable. Without Brahe's accurate and voluminous data, Kepler would not have been able to develop the mathematical description of planetary motion. Without the accurate measurement of volume, temperature, and pressure of gases, it would be impossible to uncover the proportional relationships among those variables. Scientists used different approaches to achieve accurate measurement. Brahe built specialized instruments in an underground observatory and conducted regular calibration. Mendel began a sequence of experiments with pure line plants, which allowed him to differentiate two types characters for the offspring: dominant and recessive. The strategies for accurate measurement were developed based upon the notion of measurability—variables have numerical values and units when measured. Although variables have numerical values when measured, we do not need to measure them or know their measures in order to reason about them (Thompson, 1993). Other examples in the history of science include the use of heartbeats to measure time (Rovelli, 2011) and the use of standard measures of length starting in the 18th century (Crosland, 1969).

Third, the conceptualization of scientific concepts, principles, and theories are intended to explain the mathematical patterns; such mathematical patterns are often described as complex relationships among the variables. Many phenomena are relationally complex, meaning that sophisticated understanding of those phenomena requires analysis that involves multiple variables and different types of variables (Thompson, 1993). In the development of the ideal gas law, the inconsistency between Gay-Lussac's observation and Avogadro's hypothesis emerged from the fine-grained description of the relationships among temperature, volume, pressure, and number of gas particles. The scientific idea that the characteristic particle of gases must be a molecule rather than an atom was generated to resolve this inconsistency. Mendel proposed the laws of hybridization to explain the complex relationships among the numbers of plants with contrasting characters in several generations. Kepler's laws describe the complex relationships between time and distance. To explain these complex relationships, Newton hypothesized that the force between the Sun and its planets is a type of gravitational force. This hypothesis allowed him to apply Newton's laws on terrestrial objects to celestial objects. Therefore, we focus on the development of the quantification competency—understanding the relevance, measurability, and relational complexity of variables. This competency provides a foundation for a later conceptualization of scientific concepts, principles, and theories. The above historical analysis also suggests that the mathematical description of phenomena is at the center of quantification. Therefore, quantification is not pure mathematical reasoning; it cannot be completely separated from understanding of science content.

2.2 Quantification in Scientific Revolutions

The examination of the events in the normal science uncover the nature of quantification—understanding the relevance, measurability, and relational complexity of variables. To hypothesize how this understanding develops over time, we refer to quantification in two scientific revolutions, because the conceptual change experienced by students can have parallels with the conceptual changes in the history of science (McComas et al., 1998; Wiser & Carey, 1983).

The first event is the chemical revolution—the paradigm shift from the phlogiston theory to the oxygen theory of combustion (Bynum, 2013; Thagard, 1992). The phlogiston theory was once a popular theory that explained phenomena such as burning and rusting. The word, *phlogiston*, comes from ancient Greek, meaning fire principle. According to the phlogiston theory, when a material burns in air, its phlogiston is transferred into the air. When losing its phlogiston, the material becomes ashes and weighs much less. Materials ceased burning in an enclosed space because the air in that space is saturated with phlogiston. Saturated air does not support burning. The phlogiston theory attempts to conserve materials qualitatively—a substance lost weight after combustion, so phlogiston must be released into the air. There was no attempt to quantify conversation such as measuring the phlogiston or the mass gained or lost in materials.

During the 18th century, many scientists were conducting experiments of burning, calcination, and breathing. However, Lavoisier was the first to conduct these experiments in closed systems and with accurate measurements of mass. From his experiments, Lavoisier found phenomena that could not be explained by the phlogiston theory: sulfur gained weight after combustion; when metals changed into calxes (powder), the latter weighed more than the original metals. During that time, Priestley found a mysterious “new air” by heating red calx (mercury oxide). The new air seemed to support breathing and burning. Priestley introduced the new air to Lavoisier. Lavoisier later named this new air oxygen and considered oxygen’s role in burning. Lavoisier studied combustion and calcination of different materials in a closed vessel system. By doing so, he was able to shift the focus from the mass of the material to the mass of the whole system and to consider gas’ contribution to mass change. This focus is reflected in the following quote from Lavoisier’s book, *Elements of Chemistry*. In this quote, Lavoisier described the result of burning iron wire in a closed vessel (Holton & Brush, 2006, p. 205):

> If the experiment has succeeded well, from 100 grains [5.3 grams] of iron will be obtained 135 or 136 grains of ethiops [oxide of iron], which is an augmentation [of mass or of weight] by 35 percent. … … Having therefore burnt 100 grains of iron, which has required on additional weight of 35 grains, the diminution of air will be found exactly 70 cubical inches; and it will be found, in the sequel, that the weight of vital air [oxygen] is pretty nearly had a grain for each cubical inch; so that, in effect, the augmentation of weight in the one exactly coincides with the loss of it in the other.

Based on the mathematical patterns identified from the data, Lavoisier proposed new ideas about air and the combustion process (Holton & Brush, 2006). He concluded that air has two elements; while oxygen supported combustion and breathing, fixed air (i.e., carbon dioxide) did not. With the consideration of oxygen’s role in combustion, he was able to develop an oxygen theory of combustion and claim that the total mass is conserved in combustion.

The second example of quantification in scientific revolution comes from forces and motion. The quantification of motion has three important stages: Aristotelian conceptualization of motion, early efforts to quantify motion, and Newtonian quantification of motion (Damerow, Freudenthal, McLaughlin, & Renn, 1991; Paty, 2003). Aristotle differentiated between two types of motion—natural motion and violent motion. In natural motion, bodies always move towards their natural position, which is usually caused by the combination of fire, water, soil, and air. In violent motion, an external force pushes the body. Aristotle thus began by identifying a putatively important attribute of motion: its cause. A precursor of quantification can be found in Aristotle’s lengthy discussion of *quicker*: The quicker of two bodies traverses more space in the same time and the same space in less time (Damerow et al., 1991). In Aristotle, attributes of motions (natural vs. violent; quick vs. slow) are identified and compared. However, these attributes are not quantified, because whether a body moves quicker than another is determined based on perceptions rather than measurement.

Two ideas are important in early efforts to quantify motion. First, Buridan developed the impetus theory to explain motion. He defined impetus as being proportional to the amount of matter and the speed (Stinner, 1994). As such, impetus is a compound quantity—a quantity resulting from operation on other quantities (Brahmia, Boudreaux, & Kanim, 2016). Although this definition of impetus resembles the modern concept of momentum, impetus is treated as an internal property of moving objects and the cause of motion rather than the effect of motion. Second, Oresme developed a tool, the doctrine of the configuration of qualities, to quantify a wide variety of physical and moral qualities such as whiteness and charity (Damerow et al., 1991). Using this tool, the intensity of a dimension of a phenomenon is expressed by degrees; and the quantity of that dimension is conceived as dependent on both the intensity and the size of the substance. In application of this tool to motion (e.g., by Descartes and Buridan), the intensity of motion is depicted as velocity, acceleration, or impetus, and the extension of the motion is described as that intensity being accumulated during a time span. In these early efforts, we see a shift from qualitative attributes to measurable variables. While Aristotle's analysis is one-dimensional and qualitative, Buridan and Oresme treated motion as a relationally complex phenomenon and used multiple variables to analyze motion: Buridan used a compound quantity (involving mass and velocity) to define impetus; and Oresme and Descartes quantified motion by differentiating between two types of variables—intensive and extensive variables.

In classical mechanics, established based on the work of Newton and Galileo, motion is interpreted using multiple variables, including displacement, time, speed, velocity, acceleration, momentum, and force. These variables are clearly defined and distinguished from each other. The relationships among the variables are also clarified. Newtonian quantification differs from the earlier quantification in that it treats force as interaction and associates force with acceleration rather than velocity.

The two scientific revolutions discussed above suggest that the fundamental change of theories was enabled by two conceptual shifts. The first shift is about the nature of variables; it is a shift from a qualitative perspective to a quantitative perspective. While the qualitative perspective focuses on *attributes* of phenomena (e.g., less material, quicker, more time), the quantitative perspective focuses on measurable variables. As elaborated above, two important features of scientific quantification are the identification of relevant variables and the recognition of the measurability of the variables. These two related features are missing in the qualitative perspective. More importantly, when interpreting and analyzing phenomena in terms of attributes, it is unnecessary to measure the numeric values of the variables. For example, the phlogiston theory assumes the existence of phlogiston but made no attempt to measure it. Aristotle described motion as natural versus violent, which was based on perception rather than measurement.

The second shift is focused on relationships among variables; it is a shift from understanding of simple relationships among variables to the understanding of relational complexity. Newton's laws of forces and motion provide a clear differentiation among velocity, acceleration, and the force—acceleration is the change in velocity; force is associated with acceleration not velocity. However, in the impetus theory, the relationships between velocity/acceleration and force is vague. Lavoisier's explanation of how mass changed in burning iron suggests that he considered the relationships among several variables, including the mass of the original iron, the mass of the ethiops (oxide of iron), mass of the oxygen, and mass of all substances in the closed system. However, the phlogiston theory only considered whether the material would change its mass after burning.

2.3 A Hypothetical Learning Progression Based on the Historical Examination

In parallel with the conceptual shifts that happened in the history of science, we hypothesize that four levels of an LP for quantification could exist in student learning. The levels on the LP are:

- *Level 1. Holistic observation*: Students treat phenomena as a whole and do not identify or distinguish attributes or aspects of the phenomena.

- *Level 2. Attributes*: Students describe attributes of a phenomenon in light of their everyday concepts, staying at the level of observation. At this level, students identify attributes and characteristics of a phenomenon, but do not quantify them as measurable quantities or variables.
- *Level 3. Measurable Variables*: Students analyze a phenomenon in terms of measurable quantities—the quantity or variable should and can be measured in terms of numeric values. They understand simple relationships among quantities but not the scientific meaning of the complex relationships (e.g., compound quantities, relationships between change and rate of change, distinctions between extensive and intensive variables). They may identify some but not all relevant variables that are required to describe the mathematical patterns. Students at this level demonstrate a beginning understanding of graphs to help them examine relationships. They understand the scientific meaning of the points on the graph. They may also identify the mathematical relations, patterns, and trends in the graph. However, they do not understand the scientific meanings of those relations, patterns and trends.
- *Level 4. Relational Complexity*: Students distinguish among the different types of variables and understand the complex relationships among variables in terms of their scientific meanings (e.g., compound quantities, relationships between change and rate of change, distinctions between extensive and intensive variables, proportional relationship between a quantity and a square of a quantity). They also develop a sophisticated understanding of the scientific meanings of the relations, patterns, and trends in the graphs.

This LP for quantification provides a general view of successively more sophisticated ways of thinking about phenomena, from experiencing them holistically, to identifying attributes, to developing quantifiable/measurable variables to capture the attributes, to ultimately being able to understand the scientific meaning of the complex relationships among variables and/or types of variables. We hypothesize that this grain size of LP will be useful to teachers, researchers, and developers. In contrast, some previous work in this area by Mayes and colleagues is complex and multifaceted, encapsulating three progress variables, each with four elements, which in turn have four achievement levels each (for 48 distinctions total) (e.g., Mayes, Forrester, Christus, Peterson, & Walker, 2014; Mayes, Peterson, & Bonilla, 2013).

While Mayes and colleagues' approach provides details, our approach is more parsimonious and therefore provides a big picture for teachers to understand student learning. More specifically, the grain size of our LP is intended to (a) capture meaningful patterns of conceptual shifts that occur incrementally, and (b) be instructionally relevant but not overly constraining. In particular, the *shifts* we have outlined are likely to be of instructional significance, suggesting instructional activities to spur development along the LP. The *levels* should be a powerful conceptual tool for teachers to recognize and respond appropriately to students' ideas and approaches to quantification. Another important characteristic of this LP is the integration of mathematical thinking and scientific thinking. As shown in the historical analyses, scientists conceptualized mathematical relationships, patterns, and trend into scientific theories. In the LP, this conceptualization of mathematics begins at Level 3 and is fully developed at Level 4. We developed the hypothetical LP based on five historical events that brought significant advances in scientific knowledge. It is important to note that although mathematization plays a crucial role in knowledge advancement, it is not the only important aspect of science.

3 Evidence to Support the Quantification Learning Progression in the NGSS

3.1 Overview

Having examined multiple examples from the history of science for evidence that supports the quantification in science LP, we turn now to evidence from contemporary sources. In this section, we analyze the NGSS and associated documents for descriptions of the development of quantification through K-12 schooling. By doing so, we search for evidence about the levels of the hypothetical LP for quantification in science. As mentioned earlier, by generating an LP for quantification, we also lay the groundwork for future standards development efforts to include quantification as a stand-alone practice.

The first step in our analysis consisted of a high-level examination of the NGSS (The NGSS Lead States, 2013) and *Framework* (NRC, 2012). The *Framework* that guided the development of the NGSS has a chapter on practices that presents an overview of each practice, including general descriptions, progressions, and grade 12 goals. However, there is no detail on how the development of each practice progresses by grade band, so the *Framework* was not found useful for LP development. The NGSS has two sections where quantification is contemplated: the scientific and engineering practices and the crosscutting concepts (CCCs). Appendix F of the NGSS is devoted specifically to the practices and "[d]escribes the progression of the practices across K-12, detailing the specific elements of each practice that are targets for students at each grade band." Appendix G provides a similar description for the CCCs. These two appendices provide rich and compact descriptions of the skills that students are expected to develop, by grade level, including quantitative competencies and reasoning.

The NGSS contemplate eight scientific and engineering practices, none of which explicitly focuses on quantification. Six of them *include* aspects of quantification (see Table 1). Two practices, *Engaging in Argument from Evidence* (Practice 7) and *Obtaining, Evaluating, and Communicating Information* (Practice 8), have no relevant text on quantification in Appendix F. Our next step in the analysis of the science standards documents was to extract the quantitative aspects of each practice at each grade level from Appendix F and relate them to our posited levels. Text was reviewed from each grade band description of each practice for examples or expectations of quantification and a determination made about which level of the quantification LP was most relevant.

To ensure a deeper analysis of the quantitative aspects of the practices, we next examined the full NGSS standards organized by disciplinary core ideas (DCI), henceforth termed NGSS/DCI, to capture any relevant information that did not appear in the more condensed Appendix F. Specifically, we identified DCIs that list the *Analyzing and Interpreting Data or Mathematical and Computational Thinking* practices in each disciplinary area (physical sciences, life sciences, and earth/space sciences) at grade 2, grade 5, middle school, and high school. We focused on these two practices as we considered these to be the richest in quantification. We then analyzed the text for those DCIs listing the selected practices for language on quantification and added any novel ideas to the list developed from Appendix F.

We then followed an analogous process for CCCs, first by using Appendix G and then by examining the NGSS/DCI for the following CCCs: *Scale, Proportion and Quantity*; *Patterns*; and *Systems and System Models*. Finally, we examined the linked *Common Core Mathematics Standards* that the selected DCIs analyzed in the previous step listed and again compared these to the LP levels.

3.2 NGSS Practices

In this section, we present our findings concerning quantification in the NGSS/DCI and Appendix F in the NGSS.

- *Practice 1, Asking Questions and Defining Problems* (Appendix F), contains aspects of quantification in the description of each grade band's recommendations. Students in the grade band of K-2 are to build on prior experiences, which in our interpretation they likely experienced holistically. By developing descriptive questions, they are being urged to focus on particular attributes, consistent with our shift from Level 1 (holistic) to Level 2 (attributes). The description of 3-5 grade band involves first qualitative relationships, which in our interpretation involve attributes and subsequently measurement, consistent with the shift from Level 2 (attributes) to Level 3 (definition of variables). The descriptions for grade bands 6-8 and 9-12 are about understanding of the relationships among variables and types of variables, which is captured in Level 4 (relational complexity). In summary, Practice 1 follows the *order* of our LP's levels and furthermore proposes grade bands suitable for each level and shift.
- *Practice 2, Developing and Using Models* (Appendix F), takes a very different approach to quantification. Already in the K-2 grade band, Level 3 understandings are included: "Develop and/or use a model to represent amounts, relationships, relative scales (bigger, smaller), and/or patterns in the natural world." (p. 6) We interpret "amounts" to be the results of measurement and thus to involve variables (Level 3). By grades 3-5 students should develop or revise models to show relationships among variables (Level 4), The grade bands 6-8 and 9-12 likewise propose that students think about the relationship among variables (Level 4).
- *Practice 3, Planning and Carrying out Investigations* (Appendix F), places measurement and thus variables already in the K-2 grade band (Level 3), with control variables (thus introducing types of variable – Level 4) in the 3-5 grade band, and additional types of variables in the 6-8 grade band (independent, dependent) and the 9-12 grade band (confounding variables), again consistent with Level 4. As observations are included for K-2, it seems that Levels 2-4 and possibly 1 are included in this practice; however, as with Practice 2, Level 3 understandings are already included at K-2.
- *Practice 4, Analyzing and Interpreting Data,* places attributes and perhaps holistic phenomena (Levels 2 and 1) at the K-2 grade band through the collection of observations, in Appendix F. The NGSS/DCI further elaborate: Students collect, record, and share observations, which we interpret as focusing on attributes (2-PS1, K-2-ETS1), and "analyze data from tests of an object or tool to determine if it works as intended" (p. 9)—which again does not refer to measurement or quantitative data but instead suggests a focus on attributes. Quantitative measurements and thus variables are introduced in grades 3-5 (Level 3), per Appendix F. The NGSS/DCI concurs, noting that students use "quantitative approaches to collecting data," including representation of data in graphs (5-ESS1). By grades 6-8, variables in complex, nonlinear relationships (Level 4) are proposed in Appendix F. The NGSS/DCI likewise mentions quantitative analysis, distinguishing between causation and correlation, error analysis (MS-PS1), and identifying linear and nonlinear relationships (MS-PS3, MS-LS2, MS-LS4, MS-ESS1, MS-ESS2, MS-ESS3, MS-ETS1). Quantification is not included in the 9-12 grade band for Practice 4 in Appendix F. The NGSS/DCI mentions

statistical analyses, use of models (HS-PS2, HS-LS3, HS-ESS2, HS-ESS3) and curve fitting (HS-LS3, HS-LS4), consistent with Level 4.

- *Practice 5, Using Mathematics and Computational Thinking*, has K-2 students already measuring quantitative attributes, for example, using variables (Level 3), as well as deciding the appropriateness of qualitative versus quantitative data for given scenario, in Appendix F. The NGSS/DCI for grade 2 does not include links to this practice. By grades 3-5, students measure various physical properties including area, volume, weight, and time, per Appendix F. The NGSS/DCI elaborates: Students "[extend] quantitative measurements to a variety of physical properties" (p. 10) and use computation and mathematics to analyze data (5-PS1, 5-ESS2)—implying the purposeful use of variables, and mentioning weight, area, and volume explicitly. There is no information for the 6-8 grade band in Appendix F, and the NGSS/DCI text only discusses identifying patterns and using mathematical concepts and representations (MS-PS4, MS-LS4). For the high school grade band, Appendix F has no relevant information, but the NGSS/DCI mentions a range of linear and nonlinear functions to model data mathematically, consistent with our Level 4 (HS-PS1, HS-PS2, HS-PS3, HS-PS4, HS-LS2, HS-LS4, HS-ESS1, HS-ETS1). In contrast to the term *variable* used in the other practices, this practice uses *quantitative attribute*, *quantity*, and "quantitative measurement [of] a variety of physical properties" (Appendix F, p. 10), and data modeling or mathematical or computational representations of data (NGSS/DCI)
- *Practice 6, Constructing Explanations and Designing Solutions* (Appendix F), places observations of natural phenomena, which we interpret to mean attributes or possibly holistic phenomena at K-2 (Levels 2 and 1, respectively); use of variables and measurement by grades 3-5 (Level 3); and progresses to quantitative relationships among variables by grades 6-8 and types of variables (dependent, independent) by 9-12 (Level 4).
- *Practice 7, Engaging in Argument From Evidence*, had no relevant text on quantification in Appendix F. However, the NGSS/DCI has some potentially relevant fragments, mentioning supporting an argument with data in grade 5 (5-PS2) or empirical evidence in middle school (MS-PS2, MS-PS3, MS-LS1, MS-LS2, MS-ESS3) and use of evidence in high school (HS-PS4, HS-LS2, HS-LS3, HS-LS4, HS-ESS1, HS-ESS2), data (HS-ESS2) and empirical evidence (HS-ESS3). While evidence may include attributes, measurement, and variables, these are not mentioned explicitly. It is very important for NGSS to provide a clear definition of evidence, because the term evidence "is used to denote a variety of different kinds of information including personal experience, empirical data, simulation-derived data, science reports in popular media, and so on (Duncan, Chinn, & Barzilai, 2018, p.911)". As the types of data or evidence are not elaborated further, we consider that Practice 7 does not meaningfully delve into quantification.
- *Practice 8, Obtaining, Evaluating, and Communicating Information,* did not include any relevant text on quantification in Appendix F. The NGSS/DCI further mentions observations at grade 2 (2-ESS2), but at grade 5 (5-ESS3) and middle school (MS-PS1, MS-PS4, MS-LS1, MS-LS4) solely discusses obtaining information from texts and media, and at high school does not mention quantification-related concepts. It is not clear what constitutes information.

[Insert Table 1 about here]

Having identified how the NGSS/DCI describe quantification across the practices and grade bands, next, we analyzed the treatment of quantification among the different scientific and engineering practices in the NGSS by the levels of our proposed LP (see Table 2).

[Insert Table 2 about here]

There is notable variation in the age bands proposed for each Level of the LP, as well as differences in terminology. For instance, K-2 students are expected to be at Levels 1 or 2 by Practices 1, 4, and 6, but at Level 3 by Practices 3 and 5, and at Level 4 by Practice 2. Practice 5 uses very different terms for *variable*, while Practices 1-4 and 6 use only the term variable. Practices 7 and 8 do not mention anything explicitly related to quantification. Clearly, future standards should take explicit account of quantification and ensure that there is an explicit and coordinated progression in this important topic.

3.3 NGSS Crosscutting Concepts (CCCs)

In this section, we present findings concerning quantification in the NGSS/DCI and Appendix G.

3.3.1 Patterns

Appendix G includes some consideration of quantification in the CCC of patterns and links it explicitly to Practice 4, *Analyzing and Interpreting Data*, and Practice 5, *Using Mathematics and Computational Thinking*. Examples in the introductory text on this CCC include geographical patterns (probably dealing with attributes), plotting data values on a graph (involving the measurement of variables), and visual inspection of organisms of minerals (attributes). By grade bands, Appendix G refers to observations and description for K-2 (holistic phenomena, Level 1, and attributes, Level 2); sorting and classifying (attributes) and using rates and cycles related to time (measurement, Level 3) at grades 3-5; rates of change and other numerical relationships (Level 3 and potentially Level 4) at middle school; and using mathematical representations to identify patterns (Level 3 and probably Level 4) at high school.

The NGSS/DCI provides very little additional detail. At grade 2, the only description is that "Patterns…can be observed" (2-PS1, 2-ESS2). At grade 5, "Similarities and differences in patterns can be used to sort, classify, communicate, and analyze simple rates of change for natural phenomena" (5-ESS1). At middle school, there is more inclusion of patterns, yet with insufficient detail. The relationship between atomic/micro-level explanation of macro-level phenomena is included (MS-PS1), as is the usefulness of graphs to identify patterns in data (MS-PS4, MS-LS4, MS-ESS3); the latter involves variables (Level 3 and potentially Level 4). Additionally, the usefulness of patterns in identifying cause and effect relationships is presented (MS-LS2, MS-LS4, MS-ESS1)—which might involve attributes or variables (Levels 2-4), and rates of change and other numerical relationships (MS-ESS2; Level 3 and possibly Level 4).

3.3.2 Systems and System Models

Appendix G defines a system in terms of forces, as well as flows of matter and energy, which are variables. At the K-2 and 3-5 grade bands, students are to describe objects and organisms in terms of their parts, consistent with Level 2 (attributes). At middle and high school grade bands, input and outputs in terms of matter, energy, and information are discussed, consistent with Level 3 (variables). It is unclear whether understanding complex relationships among variables is expected (Level 4).

The NGSS/DCI is aligned with the previous descriptions from the Appendix G, with no inclusion of this CCC at grade 2; components and interactions at grade 5; and inputs, outputs, and flows of energy and matter at middle school. At high school, attention is drawn to initial conditions and boundaries, and the nature of models and modeling, which are not directly or explicitly related to quantification.

3.3.3 Scale, Proportion, and Quantity

Appendix G defines scale in terms of size, time, and energy, and links this CCC explicitly to Practice 4, *Analyzing and Interpreting Data*, and Practice 5, *Using Mathematics and Computational Thinking*. Both qualitative relationships and measurement of variables are discussed: "At a basic level, in order to identify something as bigger or smaller than something else—and how much bigger or smaller—a student must appreciate the units used to measure it and develop a feel for quantity." (p.???) Proportional comes into play through the ratios of simple quantities that result in new variables, such as speed or density.

Per Appendix G, at K-2 students use relative scale such as hotter/colder or faster/slower to describe objects, consistent with our Level 2, focusing on attributes. They begin to measure length (Level 3, variables). At grades 3-5, measurement extends to weight, time, temperature, and volume (Level 3). In middle school, proportional relationships result in variables such as time or density, and students use algebraic expressions to represent scientific relationships (Level 3 and possibly Level 4, depending on the types of relationship). In high school, they progress to thinking about orders of magnitude and nonlinear relationships including exponential (Level 4).

The NGSS/DCI are consistent with Appendix G's descriptions. There is no inclusion of this CCC at grade 2. At grade 5, the NGSS/DCI descriptions of this CCC include measurement of the same variables mentioned in Appendix G (5-PS1, 5-ESS2). At middle school the use of proportional relationships to generate rates or variables such as density is consistent with the Appendix G description (MS-PS3). Likewise, at high school, orders of magnitude (HS-LS2) and exponential relationships (HS-LS3, HS-ESS1) are presented in alignment with Appendix G.

Table 3 summarizes the key information gleaned from the examination of the NGSS/DCI and Appendix G for quantification-related concepts and location by grade band.

[Insert Table 3 about here]

3.4 Common Core Standards for Mathematics

We first identified the Common Core State Standards for Math (CCSS-M) standards linked to the DCIs that contained mentions of Practices 4 and 5 (*Analyzing and Interpreting Data*, and *Using Mathematics and Computational Thinking*, respectively). Next, we determined whether standards related in a meaningful, detailed way to quantification and removed those standards that did not from further consideration. Such standards included very general ones, such as MP.2 Reason abstractly and quantitatively, MP.4 Model with mathematics, MP.5 Use appropriate tools strategically, 5.NBT.A.1 Explain patterns in the number of zeros of the product when multiplying a number by powers of 10, explain patterns in the placement of the decimal point when a decimal is multiplied or divided by a power of 10, and use whole-number exponents to denote powers of 10. Other unrelated mathematics standards dealt with purely mathematical skills, such as HSA-CED.A.4 Rearrange formulas to highlight a quantity of interest, using the same reasoning as in solving equations. (For the full list of mathematics standards deemed unrelated, see online supplementary materials, Table S1.)

We then arranged the referenced CCSS-M standards by grade band of the NGSS/DCI referencing the mathematics standard. We found that the two standards documents lined up well, with science NGSS/DCIs referencing mathematics standards in the same grade band or earlier in every case. Finally, we related the relevant mathematics standards to our LP levels, as described next. For the K-2 grade band, the mathematics standards include data sets with up to four categories, for picture graphs and bar graphs (2.MD.D.10). Such graphs usually relate to counts of objects such as pets, meaning that the object involved is treated holistically, consistent with our Level 1. For the 3-5 grade band, students are to graph using the coordinate plane and interpret values in context (5.G.A.2); converting among measurement units within a single system (5.MD.A.1); and understand (5.MD.C.3) and carry out (5.MD.C.4) volume measurement, all of which imply the use of measurable variables (Level 3). Additionally, foundational understanding of powers of 10 are mentioned (5.NBT.A.1), laying the basis for later using orders of magnitude and exponential relationships. Middle school mathematics standards linked to NGSS/DCIs explicitly refer to variables: understanding that they represent an unknown number (6.EE.B.6) and can be used to solve real-world problems (7.EE.B.4); using two variables to represent quantities that co-vary and conceptualize variables as dependent and independent (6.EE.C.9); use ratios (6.RP.A.1) and rates (6.RP.A.2) to solve real-world problems (6.RP.A.3); recognize proportional relationships (7.RP.A.2); and model linear equations and give examples of nonlinear functions (8.F.A.3). These mathematics standards imply Level 3 understanding of variables, along with Level 4 understanding of types of variables (dependent, independent) and nonlinear relationships. The high school mathematics standards include solving problems involving variables (HSA-CED.A.1); using equations and constructing graphs with two or more variables (HSA-CED.A.2); represent data on a number line (HSS-ID.A.1) or scatter plot (HSS-ID-B.6); use units as tool to understand and solve problems (HSN.Q.A.1); and define quantities for descriptive modeling (HSN-Q.A.2). These standards rise to Level 4, given the treatment of multiple variables and the relationships among them.

Clearly, the CCSS-M standards referenced in the selected NGSS/DCIs align well with our LP level, with higher levels corresponding to higher grade bands. The most significant difference between CCSS-M standards and our LP concerns holistic observation (Level 1) and attributes (Level 2), as these are mainly absent in the CCSS-M. The only mention of attributes in the mathematics standards examined is for volume as an attribute of solid figures. Additionally, the mathematics standards refer to many valuable skills that are routinely used in science, such as rearranging formulas, graphing functions, or developing probability models that fall beyond the focus of our LP.

3.5 Key Learnings from Review of Standards

In summary, our analysis reveals that while quantification is present in most of the NGSS practices and CCCs, the treatment of quantification is often tacit and the terminology and timeline for development of quantification are frequently inconsistent across practices and/or CCCs. Given the crucial role of quantification in science and science learning and its tacit presence in the NGSS's practices, our LP for quantification can help strengthen and make more consistent the NGSS's vision of scientific practices. This effort is consistent with Osborne et al.'s (2018) proposal of using mathematical deduction (mathematization) as crosscutting theme to achieve curricular coherence. According to them mathematical deduction is one of the styles of reasoning (mathematical deduction, experimental evaluation, hypothetical modeling, etc.) that scientists used to answer fundamental ontic, causal, and epistemic questions in scientific inquiry, and therefore they should be used as crosscutting themes across all science disciplines, and by doing so, promote coherent and in-depth understanding of science.

4 Empirical Evidence for the Levels of the Hypothetical Learning Progression

A third source of evidence for the levels of the hypothetical LP is students' responses to items designed to elicit quantification in science. We studied quantification in the following topics: energy in physical sciences and carbon cycle in life sciences (Jin & Anderson, 2012a, 2012b; Jin, Zhan, & Anderson, 2013). We applied the hypothetical LP for quantification to student responses to examine whether the levels could be identified. This process allows a proof of existence as well as providing rich illustrations of each level. This application occurred in three steps. In the first step, we conducted interviews to explore how a variety of scenarios and questions can be used to elicit students' reasoning patterns in quantification. The interview participants were 44 students from urban and suburban high schools in the New Jersey and New York City area. This first step was mainly a learning process for us to understand how to design scenarios and questions to assess quantification. Based on this understanding, as a second step we developed a pool of written assessment items. We conducted think-aloud interviews (Ericsson & Simon, 1993) with eight high school students to obtain validity evidence for the response process that students used to answer these items. We revised the items based on the think-aloud data. In the third step, we administered the items to high school students from different states. All students had completed learning of the relevant science topics before taking the test. We are currently analyzing the written responses from more than 5,000 students to revise and validate the LP.

In this section, we use eleven responses to two written assessment items to illustrate the levels of the LP. Item 1 assesses students' ability to engage in quantification in the context of energy in physical sciences (PS) and Item 2 in the context of carbon cycle in the life sciences (LS). For each item, we first present the item and the responses at Levels 2, 3, and 4. Then, we discuss how the responses illustrate the reasoning patterns at each of these three levels. As our participants are all high school students, Level 1 responses were expected to be rare in this sample. We did not find representative responses for whole phenomena reasoning at this stage. Item 1 and responses at Levels 2, 3, and 4 are presented below in Figure 1 and Table 4.

[Insert Figure 1 about here]

[Insert Table 4 about here]

At Level 4, students develop an understanding of relational complexity. They are able to identify all relevant variables, generate quantitative description of the complex relationships among those variables, and understand the scientific meaning of those quantitative descriptions. Response PS1 and Response PS2 are provided as examples for Level 4. Both responses recognize that both the amount and temperature should be considered, when determining the effect—the temperature of the mixture. Response PS2 provides an equation for the complex relationships among the variables and provide a conceptual reason for the equation—the heat lost from the hot water is gained by the cold water. Response PS1 does not provide an equation, but it does explain how the effect (the temperature of the mixture) is determined by the relative influence from both the amount and temperature of the hot/cold water. Therefore, both responses suggest an understanding of relational complexity.

At Level 3, students recognize that variables are measurable in a general sense, but they often do not identify all relevant variables or do not understand the scientific meaning of the complex relationships among the relevant variables. Response PS3 and Response PS4 are examples for this Level 3 reasoning pattern. Response PS3 assumes that the temperature of the mixture is the average temperature of the hot water and the temperature of the cold water. Response PS4 considers the amount of water as the only factor that affects the temperature of the mixture. None of these two responses considers relative influence from both the amount and the temperature of the hot/cold water.

At Level 2, students focus on quality rather than the quantity. They describe the qualitative attributes of phenomena rather than measurable variables. Response PS5 and Response PS6 are two examples for this Level 2 reasoning. Response PS5 does not identify relevant variables. Response PS6 analyzes the situation in terms of qualitative attributes—"how hot the warm water is and how cold the cold water is." These responses are notable for not mentioning variables because the task itself introduces the concept of variables.

Item 2 assesses students' quantification in the topic of carbon cycle in life sciences. The item and its responses are presented below in Figure 2 and Table 5.

[Insert Figure 2 about here]

[Insert Table 5 about here]

At Level 4, students understand relational complexity. They recognize the mismatch between the numbers and search for the scientific meaning for it. The mismatch is between an increase of 120 ppm in the atmospheric carbon concentration and the increase in carbon emissions of 200 ppm due to fossil fuels. As scientists know, the reason for this mismatch is that the atmospheric carbon concentration is affected by both input (emission from burning fossil fuels, etc.) and output (sequestration into plants and sea water). The 200 ppm of carbon emissions are a carbon input into the atmosphere. However, there are also carbon outputs. When both input and output are considered, the total increase of atmospheric carbon of 120 ppm is not in conflict with the 200 ppm carbon input. In the example, the student appears to recognize the existence of other factors, although the student did not explicitly specify what those factors were. Thus, response LS1 suggests a beginning Level 4 reasoning.

At Level 3, students recognized the measurability of variables, but they did not understand the complex relationships among all relevant variables. Three responses are provided as examples to illustrate this Level 3 reasoning pattern. Student response LS2 equates the two quantities—the carbon emission from fossil fuels and the amount of atmospheric carbon dioxide. Response LS3 uses evidence to support a quantitative claim—the atmospheric carbon concentration must have increased a significant amount. However, this response does not connect the two numbers—carbon emission from burning fossil fuels and the increase of atmospheric carbon dioxide. Response LS4 identifies the mismatch of the two numbers but does not recognize that the mismatch is due to the carbon output—sequestration of carbon into plants and seawater. All these responses suggest that the students are reasoning about measurable variables. However, none of the response provides a correct description of the complex relationships that explain the increase in atmospheric carbon dioxide is determined by both carbon emission and carbon sequestration.

At Level 2, students reason about attributes rather than relevant and measurable variables for the phenomena. Response LS5 is an example for this Level 2 reasoning pattern. It describes the attributes—carbon emission is pollution and bad for humans. A hypothetical Level 1 response might be to talk about a relative's coal-burning stove.

5 Discussion

Quantification is crucial for science learning because the very extent to which we know about a phenomenon is limited by how precisely and accurately we can characterize, measure, model, or predict it. The history of science is full of cases in which phenomena were studied holistically, followed by the identification of relevant attributes, after which quantification and measurement of the attributes was undertaken—in many cases involving the development of new instrumentation. The measurability of attributes resulted in the conceptualization of variables, which afforded the generation of models in which the simple or complex relationships among variables are postulated.

In this article, we report on a three-pronged effort to generate a hypothetical LP for quantification in science and then explore its plausibility. First, based on a historical examination, we developed a hypothetical LP in terms of how understanding and misunderstandings of scientific concepts have evolved through quantification. Next, we examined the NGSS to determine whether and how the scientific and engineering practices and CCCs (including the connections to the Common Core mathematics standards) aligned with this LP. Finally, we used student response data from a large field test to illustrate the levels of the LP. We provided some evidence that the progression is at a grain size to characterize important conceptual shifts in student understanding. We are currently using this LP in conjunction with other LPs (Wylie, Bauer, & Arieli-Attali, 2015, April) to explore its instructional relevance with respect to formative assessment that combines science and mathematics concepts. In this section, we first discuss how we follow the criteria of LP (Anderson, 2008) to develop the LP for quantification of science. Then, we describe the implications of the LP for research, standards, and teaching.

5.1 Meeting the Criteria for Science LPs

Anderson (2008) proposed three criteria for science LPs: conceptual coherence, compatibility with current research, and empirical validation. Conceptual coherence means that "a learning progression should 'make sense,' in that it tells a comprehensible and reasonable story of how initially naïve students can develop mastery in a domain" (p. 3). Compatibility with current research refers to the need for an LP should build on existing findings about student learning, although existing research usually does not provide enough information for developing the specific achievement levels. Empirical validation means that an LP must be grounded in empirical data about real students.

At this stage of our research, we have obtained evidence showing that the LP meets the first two criteria. As described above, although existing research has uncovered difficulties in learning quantification of science, it does not provide enough information about the transitions that students may experience in developing the quantification competency. Therefore, to develop the specific achievement levels of the LP, we referred to literature in the history and philosophy of science. Based on this work, we identified two paradigm shifts in mathematization in the history of science and used the shifts to hypothesize the achievement levels. The LP tells a coherent story about students' development in quantification of phenomena. From Level 1 (holistic observation) to Level 2 (attributes), students make the transition from reasoning about phenomena to reasoning about qualitative relationships among entities identified based on surface features (e.g., fast vs. slow; hot vs. cold). At Level 3 (measurable variables), students develop the concept of measurability—they recognize that variables have numerical values. They also begin to think about the scientific meanings of variables and relationships among variables. However, they do not understand the scientific meaning of complex relationships among variables or distinction among different variable types (e.g., intensive variables versus extensive variables). At Level 4 (relational complexity), students understand the scientific meaning of different variable types and of complex relationships among variables. For example, students differentiate internal energy and temperature with the recognition that the former is an extensive variable that relies on quantity of the substance, while the latter is an intensive variable that does not depends on the quantity of the substance. This development story is compatible with existing findings that students encounter two major learning difficulties—identification of relevant variables in real phenomena and in graphs and understanding the scientific meanings of the variables and their relationships. More importantly, the story contains additional information about what exactly students do and know in relation to those learning difficulties.

Regarding the third criterion, empirical validation, we have been collecting validation evidence throughout the whole research program. As elaborated in another article about this project, a validation framework is used to guide the process of validation (Jin, van Rijn, et al., 2019). The framework was developed based on the testing standards (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014) and the work of Michael Kane (2013). It describes the validation activities to be conducted at different stages of the research: development, scoring, generalization, extrapolation, and use. Currently, we have collected validity evidence at the development stage. This evidence is qualitative, including the interview data and feedback from mathematics education experts in quantitative reasoning, science education experts in learning progressions, and science teachers. The think-aloud interview data provide information about the students' thought processes in completing the tasks. It shows that students understood the task questions to mean what we intend. We iteratively revised the LP based on input from the experts in our research group and expert panel. Following the validation framework, in the scoring stage, we will use an iterative process to develop and revise the scoring rubrics; in the validation stage, IRT (Item Response Theory) analysis will be performed, and Wright Maps will be developed to evaluate the order of and the differentiation among the LP levels. Evidence collected at these two stages may lead us to revise the LP levels, potentially adding sub-levels or merging levels (Shea & Duncan, 2013). At the extrapolation stage, we will study to what extent students' proficiency in quantification of science is linked to their performance in science courses. Finally, at the use stage, we will conduct a classroom study, where teachers will employ the LP with students and use the assessment results to inform their teaching. The data collected in the classroom study, including observation data, student pre- and post-tests, teacher surveys, and teacher interviews, will provide validity evidence showing to what extent the LP is useful for teachers to help students move toward higher levels on the LP (i.e., consequential validity).

5.2 Implications for Research, Standards, and Teaching

Our work provides two implications for research, standards, and teaching. Regarding research, one unique approach used our research is the historical analysis. The definition of quantification and the development of the quantification LP is based on examination of five events in the history of science. As conceptual change and conceptual development in the history of science often parallel students' development, this approach—proven fruitful here—can be used in other research on LPs. It is worth noting that this approach has been proven fruitful in the past, with the conceptual change current of constructivism (e.g., Posner, Strike, Hewson, & Gertzog, 1982) having been influenced by Kuhn's account of scientific revolutions (1962).

The NRC Framework (NRC, 2012) describe progressions in the learning of disciplinary core ideas, crosscutting concepts, and scientific and engineering practices. We examined NGSS to identify evidence for the levels of the LP. While some pieces of evidence support the order of the LP levels, our examination also suggests inconsistency in NGSS for different scientific practices, both in grade sequencing of the levels and in terminology. Future revision of NGSS could resolve this inconsistency. Future standards documents could also further develop Practices 7 and 8, and unpack the ideas of evidence and information thoroughly, linking these to quantification as well as precision and accuracy.

In a systematic review of LP literature, Jin et al. (2019) found that, although many LPs have been developed during the past decade, relatively fewer studies have been conducted to explore the use of LPs for instruction and teacher learning. As the ultimate goal of LP research is to promote teaching and learning in classrooms, more research efforts are needed to investigate teachers' learning and use of LPs. Given that LPs identify instructionally relevant patterns in students' understanding of a key concept, skill, or process, they can be used to support the development or deepening of teachers' content knowledge for teaching (Sztajn, Confrey, Holt Wilson & Edgington, 2012; Wilson, Sztajn, Edgington & Confrey, 2014). An understanding of the developmental levels of the quantification LP would help a teacher develop in-depth understanding of scientific knowledge and anticipate common student responses. In addition, the identified conceptual shifts can suggest instructional activities and prompts to propel students to advance along the LP. We will be working with teachers to connect the quantification LP and the associated assessments to their classroom practices. We expect to run a classroom study to begin to understand how the teachers use the LP and the assessment tasks, and how the use of both the LP and the tasks affect their content knowledge for teaching, their classroom practices, and student learning.

In helping the teachers understand and use the LP, existing research provides insightful ideas. Existing literature suggests major challenges for teachers: achieving the highest level of the LP; eliciting and interpreting student thinking described at different LP levels; and designing activities that use the LP levels as foundations for learning (Aschbacher & Alonzo, 2006; Furtak, 2012; Furtak & Heredia, 2014; Gunckel, Covitt, & Salinas, 2018; Jin, Johnson, & Yestness, 2015; Jin, Johnson, Shin, & Anderson, 2017; Jin, Shin, Johnson, Kim, & Anderson, 2015). Researchers have explored several useful strategies, including engaging teachers in analyzing videos of student learning (Aschbacher & Alonzo, 2006), guiding teachers in using the LP to develop formative assessment tasks, and providing "educative" materials (materials that support teacher learning – Beyer, Delgado, Davis, E. A., & Krajcik, 2009; Davis, E. A., & Krajcik, 2005) that describe the nature of LP and the use of LP for developing lesson plans (Gunckel et al., 2018). We will consider these strategies in preparing the participating teachers for the classroom study.

References

Adamson, S. L., Banks, D., Burtch, M., Cox III, F., Judson, E., Turley, J. B., . . . Lawson, A. E. (2003). Reformed undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement. *Journal of Research in Science Teaching, 40*, 939–957.

Allen, G. E. (2003). Mendel and modern genetics: The legacy for today. *Endeavour, 27*, 63–68.

Altig, J. (2014). *The historical gas laws*. Retrieved from the Internet Archive Wayback Machine: https://web.archive.org/web/20171025064301/http://infohost.nmt.edu:80/~jaltig/HistoricalGasLaws.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, C. W. (2008). *Conceptual and empirical validation of learning progressions. Response to "*Learning progressions: Supporting instruction and formative assessment.". East Lansing, MI: Michigan State University.

Aschbacher, P., & Alonzo, A. C. (2006). Examing the utility of elementary science notebooks for formative assessment purposes. *Educational Assessment, 11*, 179–203.

Beyer, C., Delgado, C., Davis, E., & Krajcik, J. (2009). Investigating teacher learning supports in high school biology curricular programs to inform the design of educative curriculum materials. *Journal of Research in Science Teaching 46*(9), 977-998.

Black, P., Wilson, M., & Yao, S.-Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research & Perspective*, *9*, 71–123.

Brahmia, S., Boudreaux, A., & Kanim, S. E. (2016). *Obstacles to mathematization in introductory physics.* Retrieved from https://arxiv.org/abs/1601.01235

Bynum, W. (2013). *A little history of science*. London, UK: Yale University Press.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Chinn, C. A., & Brewer, W. F. (1993). The Role of Anomalous Data in Knowledge Acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*, 1–49.

Corcoran, T. B., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform* (Research Report No. RR-63). Philadelphia, PA: Consortium for Policy Research in Education.

Crosland, M. (1969). The Congress on Definitive Metric Standards, 1798–1799: The First International Scientific Conference? *Isis, 60*, 226–231.

Damerow, P., Freudenthal, G., McLaughlin, P., & Renn, J. (1991). *Exploring the limits of preclassical mechanics: A study of conceptual development in early modern science: Free fall and compounded motion in the work of Descartes, Galileo, and Beeckman*. New York, NY: Springer-Verlag.

Davis, B. (1997). Listening for difference: An evolving conception of mathematics teaching. *Journal for Research in Mathematics Education, 28*, 355–376.

Dunbar, K. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings, yet so difficult in the psychological laboratory. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 313–334). Cambridge, MA: MIT Press

Dunbar, K., & Fugelsang, J. (2005). Scientific thinking and reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 705–725). New York, NY: Cambridge University Press.

Duncan, R., Chinn, C., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding the next generation science standards' conceptualization of evidence. *Journal of Research in Science Teaching, 55*, 907–937.

Duschl, R. (2000). Making the nature of science explicit. In R. Millar, J. Leach, & J. Osborne (Eds.), *Improving science education: The contribution of research* (pp. 187–206). Philadelphia, PA: Open University Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised ed.). Cambridge, MA: MIT Press.

Furtak, E. M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. *Journal of Research in Science Teaching, 49*, 1181–1210.

Furtak, E. M., & Heredia, S. C. (2014). Exploring the influence of learning progressions in two teacher communities. *Journal of Research in Science Teaching, 51*, 982–1020.

Gayon, J. (2016). From Mendel to epigenetics: History of genetics. *Comptes Rendus Biologies, 339*, 225–230.

Gunckel, K. L., Covitt, B. A., & Salinas, I. (2018). Learning progressions as tools for supporting teacher content knowledge and pedagogical content knowledge about water in environmental systems. *Journal of Research in Science Teaching, 55*, 1339–1362.

Ha, M., & Nehm, R. H. (2014). Darwin's difficulties and students' struggles with trait loss: Cognitive-historical parallelisms in evolutionary explanation. *Science & Education, 23,* 1051–1074.

Hammer, D., & Sikorski, T.-R. (2015). Implications of complexity for research on learning progressions. *Science Education*, *99*, 424–431.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment.* Retrieved from the Center on Standards & Assessment Implementaiton website: https://www.csai-online.org/sites/default/files/Learning_Progressions_Supporting_2008.pdf

Holton, G., & Brush, S. G. (2006). *Physics, the human adventure: From Copernicus to Einstein and beyond*. New Brunswick, NJ: Rutgers University Press.

Jin, H., & Anderson, C. W. (2012a). Development of assessments for a learning progression on carbon cycling in social-ecological systems. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 151-182). Rotterdam, The Netherlands: Sense Publishers.

Jin, H., & Anderson, C. W. (2012b). A learning progression for energy in socio-ecological systems. *Journal of Research in Science Teaching, 49*(9), 1149-1180.

Jin, H., Johnson, M. E., & Yestness, R. N. (2015). A learning progression approach to incorporate climate sustainability into teacher education. In S. Stratton, R. Hagevik, A. Feldman, & M. Bloom (Eds.), *Educating science teachers for sustainability* (pp. 121-142). New York: Springer.

Jin, H., Johnson, M. E., Shin, H. J., & Anderson, C. W. (2017). Promoting student progressions in science classrooms: A video study. *Journal of Research in Science Teaching, 54*(7), 852-883.

Jin, H., Mikeska, J. N., Hokayem, H., & Mavronikolas, E. (2019). Toward coherence in curriculum, instruction, and assessment: A review of learning progression literature. *Science Education, 103*(5), 1206-1234.

Jin, H., Shin, H., Johnson, M. E., Kim, J., & Anderson, C. W. (2015). Developing learning progression-based teacher knowledge measures. *Journal of Research in Science Teaching, 52*(9), 1269-1295.

Jin, H., van Rijn, P., Moore, J. C., Bauer, M. I., Pressler, Y., & Yestness, N. (2019). A validation framework for science learning progression research. *International Journal of Science Education*.

Jin, H., Zhan, L., & Anderson, C. W. (2013). Developing a fine-grained learning progression framework for carbon-transforming processes. *International Journal of Science Education, 35*(10), 1663-1697.

Kampourakis, K. (2013). Mendel and the path to genetics: Portraying science as a social process. *Science & Education, 22*, 293–324.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73.

Kline, M. (1964). Mathematics in Western culture. Oxford, UK: Oxford University Press.

Kline, M. (1980). *Mathematics: The loss of certainty*. Oxford, UK: Oxford University Press.

Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science, 31*, 549–579.

Kuo, E., Hull, M. M., Gupta, A., & Elby, A. (2013). How students blend conceptual and formal mathematical reasoning in solving physics problems. *Science Education, 97*, 32–57.

Kuhn, D., & Dean, D. (2004). Metacognition A bridge between cognitive psychology and educational practice. *Theory into Practice, 43*, 268–273.

Kuhn, D., & Dean, D. (2004). Connecting scientific reasoning and causal inference. *Journal of Cognitive Development,* 5, 261–288.

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. Journal of Cognition and Development, 1, 113-129.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lawson, A. E. (1983). The acquisition of formal operational schemata during adolescence: The role of the biconditional *Journal of Research in Science Teaching, 20*, 347–356.

Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education, 2*, 307–338.

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R. K. Sawyer (Ed.), *The Cambridge handbook of: The learning sciences* (pp. 371–387). New York, NY: Cambridge University Press.

Lesh, R., Lamon, S. J., Gong, B., & Post, T. R. (1992). Using learning progress maps to improve instructional decision making. In R. Lesh & S. Lamon (Eds.), *Assessment of authentic performance in school mathematics* (pp. 343–365). Washington, DC: American Association for the Advancement of Science.

Mayes, R. L., Forrester, J., Christus, J. S., Peterson, F., & Walker, R. (2014). Quantitative reasoning learning progression: The matrix. *Numeracy: Advancing Education in Quantitative Literacy, 7*(2), 1–20.

Mayes, R. L., Peterson, F., & Bonilla, R. (2013). Quantitative reasoning learning progressions for environmental science: Developing a framework. *Numeracy: Advancing Education in Quantitative Literacy, 6*(1), 1–28.

McComas, W. F., Clough, M. P., & Almazroa, H. (1998). The role and character of the nature of science. In W. F. McComas (Ed.), *The nature of science in science education: Rationales and strategies* (pp. 3–39). Dordrecht, The Netherlands: Kluwer Academic Publishers.

McDermott, L. C., Rosenquist, M. L., & van Zee, E. H.(1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics, 55*, 503–513.

McNeill, K. L., & Berland, L. (2017). What is (or should be) scientific evidence use in k-12 classrooms? *Journal of Research in Science Teaching, 54*, 672–689.

McNeill K. L., & Krajcik, J. (2011). *Supporting grade 5–8 students in constructing explanations in science: The claim, evidence and reasoning framework for talk and writing*. New York, NY: Pearson.

National Research Council. (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning.* Washington, DC: National Academy Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: The National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states.* Washington, DC: The National Academies Press.

Niss, M. (2017). Obstacles related to structuring mathematization encountered by students when solving physics problems. *International Journal of Science and Mathematics Education, 15*, 1441–1462.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*(23)*,* 463-466.

Osborne, J., Rafanelli, S., & Kind, P. (2018). Toward a more coherent model for science education than the crosscutting concepts of the Next Generation Science Standards: The affordances of styles of reasoning. Journal of Research in Science Teaching, 55(7), 962–981.

Paty, M. (2003). The idea of quantity at the origin of the legitimacy of mathematization in physics. In C. Gould (Ed.), *Constructivism and practice: Towards a social and historical epistemology* (pp. 109–135). Baltimore, MD: Rowman & Littlefield.

Planinic, M., Milin-Sipus, Z., Katic, H., Susac, A., & Ivanjek, L. (2012). Comparison of student understanding of line graph slope in physics and mathematics. *International Journal of Science and Mathematics Education, 10*, 1393–1414.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*, 211-227.

Rovelli, C. (2011). "Forget Time." *Foundations of Physics, 41*, 1475–1490.

Shea, N. A., & Duncan, R. G. (2013). From theory to data: The process of refining learning progressions. *Journal of the Learning Sciences, 22*, 7–32.

Stinner, A. (1994). The story of force: From Aristotle to Einstein. *Physics Education, 29*, 77–85.

Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning trajectory based instruction: Toward a theory of teaching. *Educational Researcher, 41*(5), 147–156.

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

Thompson, P. W. (1993). Quantitative reasoning, complexity, and additive structures. *Educational Studies in Mathematics, 25*, 165–208.

Vass, E., Schiller, D., & Nappi, A. J. (2000). The effects of instructional intervention on improving proportional, probabilistic, and correlational reasoning skills among undergraduate education majors. *Journal of Research in Science Teaching, 37*, 981–995.

Wilson, P. H., Sztajn, P., Edgington, C., & Confrey,. (2014). Teachers' use of their mathematical knowledge for teaching in learning a mathematics learning trajectory. *Journal of Mathematics Teacher Education, 17*, 227–244.

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

Wylie, E. C., Bauer, M. I., & Arieli-Attali, M. (2015, April). *Validating and using learning progressions to support mathematics formative assessment.* Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

Table 1. Excerpts from NGSS Practices Appendix F by Grade Band Regarding Quantification

Practice	K-2	3-5	6-8	9-12
1 Asking questions and defining problems	"…builds on prior experiences and progresses to simple descriptive questions that can be tested."	"…progresses to specifying qualitative relationships"; "... what would happen if a variable is changed."	"…progresses to specifying relationships between variables"; "…determine relationships between the independent and dependent variables and relationships in models."	"Ask questions…to determine relationships, including quantitative relationships, between independent and dependent variables."
2 Developing and using models	"Develop and/or use a model to represent amounts, relationships, relative scales (bigger, smaller), and/or patterns…"	"Collaboratively develop and/or revise a model… that shows the relationships among variables…"	"Develop or modify a model… to match what happens if a variable…of a system is changed"; "Develop and/or revise a model to show the relationships among variables…"	"…progresses to using, synthesizing, and developing models to predict and show relationships among variables…"
3 Planning and carrying out investigations	"Evaluate different ways of observing and/or measuring a phenomenon…."; "Make observations… and/or measurements to collect data…."	"…progresses to include investigations that control variables…."; "Make observations and/or measurements…." (emphasis in the original)	"…progresses to include investigations that use multiple variables"; "…identify independent and dependent variables…." (emphasis in the original)	"…produce data…. Consider possible confounding variables"; "Make directional hypotheses that specify what happens to a dependent variable when an independent variable is manipulated"; "Manipulate variables and collect data…."
4 Analyzing and interpreting data	"…progresses to collecting, recording, and sharing observations"	"…progresses to introducing quantitative approaches to collecting data"; "…using…mathematics and/or computation"	"…progresses to extending quantitative analysis to investigations"; "…identify linear and nonlinear relationships."	No relevant information
5 Using mathematics and computational thinking	"Decide when to use qualitative vs. quantitative data."; "Describe, measure, and/or compare quantitative attributes…."	"…progresses to extending quantitative measurements to a variety of physical properties…."; "Describe, measure, estimate, and/or graph quantities (e.g., area, volume, weight, time)…."	No relevant information	"…progresses to using…a range of linear and nonlinear functions, exponentials and logarithms…to analyze, represent, and model data."; "…complicated measurement problems involving quantities with derived or compound units (such as mg/mL, kg/m3, acre-feet, etc.)."
6 Constructing explanations and designing solutions	"Make observations…to…account for natural phenomena."	"…progresses to the use of evidence in constructing explanations that specify variables…."; "Use evidence (e.g., measurements, observations…)"	"Construct an explanation that includes qualitative or quantitative relationships between variables…."	"Make a quantitative and/or qualitative claim regarding the relationship between dependent and independent variables."

Note. NGSS = Next Generation Science Standards.

Table 2. Treatment of Quantification in NGSS Scientific and Engineering Practices

Practice	Level 1	Level 2	Level 3	Level 4	Terminology for *variable*
1.	K-2	K-2	3-5	6-8 onward	variable
2.	K-2	K-2	K-2	3-5 onward	variable
3.	K-2	K-2	K-2 onward	6-8 onward	variable
4.	K-2	K-2	3-5	6-8 onward	variable
5.	K-2	K-2	K-2 onward	9-12	Quantitative measurement of physical property; quantity; quantitative attribute; data modeling; mathematical or computational representations of data
6.	K-2	K-2	3-5	6-8	variable
7.	N/A*	N/A	N/A	N/A	N/A
8.	N/A	N/A	N/A	N/A	N/A

Note. N/A = no mention of quantification-related concepts. NGSS = Next Generation Science Standards.

Table 3. Treatment of Quantification in NGSS Crosscutting Concepts (CCCs)

CCC	Level 1	Level 2	Level 3	Level 4
Patterns	K-2	K-2	3-5	6-8 onward
Systems and system models	--	K-2, 3-5	6-8 onward	6-8 onward
Scale, proportion, and quantity	--	K-2	K-2 (distance); 3-5 onward (other variables)	6-8 onward

Note. -- = no mention of Level 1 observation of holistic phenomena. NGSS = Next Generation Science Standards.

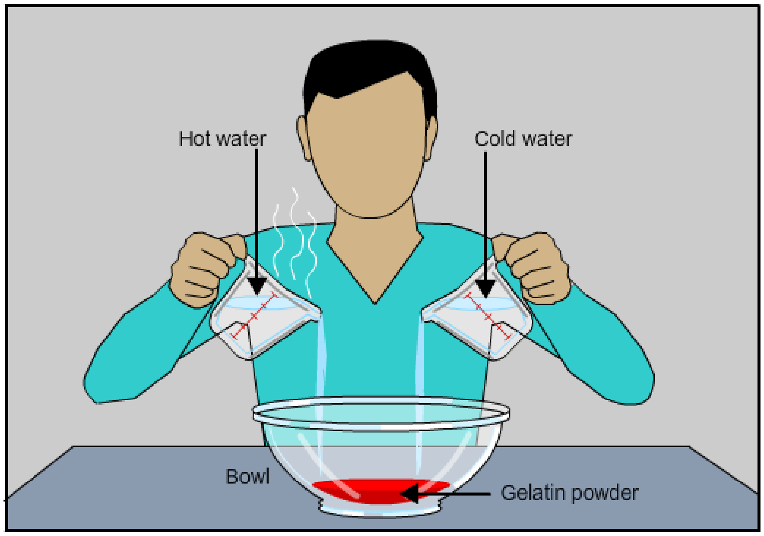
Table 4. Exemplar Responses to Item 1

Level	Example
4. Relational complexity	Response PS1: (Q1) The temperatures of the hot and cold water separately and how much water of each you have. (Q2) If the hot water is too hot and the cold water isn't cold enough or there isn't enough of the cold water, then mixing them will not result in a temperature of 40 degrees Celcius [Celsius]. Vice versa for the cold water being too cold. Response PS2: (Q1) The temperature of the hot and cold water, the volume of the hot and cold water. (Q2) If no heat is lost to the surroundings, then heat lost by the hot water must equal heat gained by the cold water. To do that, you need to make sure mass of the cold water multiplied by its change in temperature (to reach 40 degrees C) is equal to the mass of the hot water multiplied by its change in temperature (to reach 40 degrees C).
3. Measurable variables	Response PS3: (Q1) The cold water should be colder than 40 degrees celsius [Celsius] and the hot water should be warmer than 40 degrees Celsius. (Q2) This will ensure that the water added reaches 40 degrees celsius [Celsius], as the cold and hot will mix to find a middle temperature. Response PS4: (Q1) The amount of water should be recorded before they are mixed. (Q2) There should be an accurate measurement of each type of water to ensure that the mixture reaches 40 degrees Celsius. One could underestimate or overestimate the target without caution.
2. Attributes	Response PS5: (Q1) The size and type of bowl. (Q2) Because if the bowl is too big then there is a lot of space for the heat to stay and the type of bowl determines if the heat will be conserved or not. Response PS6: (Q1) how hot the warm water is and how cold the cold water is. (Q2) this will help determine the end result.

Table 5. Exemplar Responses to Item 2

Level	Example
4. Relational complexity	Response LS1 (Q1) A. (Q2) The data presnted [presented] shows that the concentration in the atmosphere increased by 120 ppm from 1750 to 2015. Therfore [Therefore] the study’s estimate of 200 ppm could be true because its greater than 120 ppm which is “caused by the fossil fuels”
3. Measurable variables	Response LS2. (Q1) C. (Q2) This could be true because fossil fuel burning does emit fairly large amounts of carbon dioxide into the atmosphere and even though the increase hasn’t reached 200ppm, its [it’s] pretty close to it.
	Response LS3. (Q1) A. (Q2) I think that the fossil fuels statement is true because since 1750 we have went through the industrial revolution and our entire world is powered by fossil fuels in which I already know have a large greenhouse gas impact on the enviroment [environment] so it only makes sense that the CO2 concentration in the atmosphere has increased by a signifigant [significant] amount over the years.
	Response LS4. (Q1) B. (Q2) If the overall increase was only 120ppm then fossil fuels could not have caused an increase of 200ppm.
2. Attributes	Response LS5. (Q1) A. (Q2) Burning fossil fuels creates pollution which is bad for humans, so is CO2 so most likely it is true.

Assessment Item 1: Gelatin is a substance used to make pudding. It dissolves in warm water that is at least 40°C in temperature. A person pours hot water and cold water into a big bowl that contains gelatin powder. Assume no heat is lost to the surrounding environment before the cold water and hot water are fully mixed.



Q1: To ensure the mixture reaches a temperature of 40°C, what variables should be measured before the hot water and the cold water are mixed?
Q2: Please explain the reasoning for your answer.

Figure 1. Illustrative physical sciences item.

Assessment Item 2: The CO_2 concentration in the atmosphere was about 280 ppm in Year 1750 and 400 ppm in Year 2015. These data are presented in the table below.

Year	CO_2 Concentration in the Atmosphere (ppm)
1750	280
2015	400
Increase between 1750 and 2015	120

Another study generated the following finding: Burning fossil fuels has caused the CO_2 concentration to increase about 200 ppm from 1750 to 2015.

Q1: Do you think the finding of the study is likely to be true?

A. Yes
B. No
C. Not enough information is provided for me to make a judgment.

Q2: Please explain why you chose that option.

Figure 2. Illustrative life science item.