JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 54, NO. 8, PP. 1024–1048 (2017)

Research Article

A Construct-Modeling Approach to Develop a Learning Progression of how Students Understand the Structure of Matter

Linda Morell,[1] Tina Collier,[1] Paul Black,[2] and Mark Wilson[1]

[1]*University of California, Berkeley, California*
[2]*King's College London, London, England*

Received 10 September 2015; Accepted 4 April 2017

Abstract: This paper builds on the current literature base about learning progressions in science to address the question, "What is the nature of the learning progression in the content domain of the structure of matter?" We introduce a learning progression in response to that question and illustrate a methodology, the Construct Modeling (Wilson, 2005) approach, for investigating the progression through a developmentally based iterative process. This study puts forth a progression of how students understand the structure of matter by empirically inter-relating constructs of different levels of sophistication using a sample of 1,087 middle grade students from a large diverse public school district in the western part of the United States. The study also shows that student thinking can be more complex than hypothesized as in the case of our discovery of a substructure of understanding in a single construct within the larger progression. Data were analyzed using a multidimensional Rasch model. Implications for teaching and learning are discussed—we suggest that the teacher's choice of instructional approach needs to be fashioned in terms of a model, grounded in evidence, of the paths through which learning might best proceed, working toward the desired targets by a pedagogy which also cultivates students' development as effective learners. This research sheds light on the need for assessment methods to be used as guides for formative work and as tools to ensure the learning goals have been achieved at the end of the learning period. The development and investigation of a learning progression of how students understand the structure of matter using the Construct Modeling approach makes an important contribution to the research on learning progressions and serves as a guide to the planning and implementation in the teaching of this topic. © 2017 Wiley Periodicals, Inc. J Res Sci Teach 54: 1024–1048, 2017

Keywords: learning progression; structure of matter; science assessment; construct modeling

Introduction

Learning progressions can help guide and align curriculum, instruction, and assessment (Duncan & Hmelo-Silver, 2009; National Research Council (NRC), 2005; Wilson, 2009), as they provide a means for laying out likely trajectories of student learning toward more sophisticated understanding. This is particularly important for meeting the high expectations set forth by current reform in the US—specifically, the *Framework for K-12 Science Education* (NRC, 2012) and the *Next Generation Science Standards* (NGSS) (NGSS Lead States, 2013), where emphasis is placed

---

Contract grant sponsor: Institute of Education Sciences; Contract grant number: R305A100692.

Correspondence to: L Morell; E-mail: lindamorell@berkeley.edu

DOI 10.1002/tea.21397

on depth of understanding, focusing on a small set of core ideas, rather than the breadth of materials covered. This emphasis is appropriate for learning progressions, as "learning progressions can provide the opportunity to examine how students' ideas evolve over time" (Merritt & Krajcik, 2013, pp. 11).

This paper provides an example of a hypothesized learning progression situated in the context of the structure of matter and a method used for the systematic investigation of the progression. Specifically, the research described here addresses the question, "What is the nature of the learning progression in the content domain of the structure of matter?" and uses a construct modeling approach, the BEAR Assessment System (BAS; Wilson, 2005), as the guiding framework to develop an account of the progression in student thinking and learning in the structure of matter domain. It seemed clear at the outset that an empirically grounded understanding of how students learn scientific content over time—that is, the pathways in which their learning may progress—would both make an important contribution to the research on learning progressions and serve as a guide to the planning and implementation in the teaching of this topic. It was also clear that, while many learning progressions have been developed, relatively few have been tested empirically.

This paper is organized into four sections: in the first, we describe previous work and how the Learning Progressions in Science (LPS) project was developed based on that previous research and by using the BAS. Then, in the second, we describe the methodology and processes involved in empirically testing the hypothesized complex learning progression, a progression that was developed from the literature. In the third, we present the empirical results for our learning progression on the structure of matter and discuss its implications for use. Specifically, we test (i) the dimensionality of the constructs in the learning progression and (ii) the ordered nature of each dimension compared to the hypothesized constructs, and thereby show how the constructs might be further refined and developed in the light of our findings. However, full details are given for only one example. In the fourth section, we review the findings and discuss their implications.

The Background of Previous Research

Among the many publications (Abraham, Grzybowski, Renner, & Marek, 1992; Andersson, 1990; Gomez, Benarroch, & Marin, 2006; Merritt & Krajcik, 2013 among others) on progressions in the learning of science, we have identified three key bodies of work based on evidence about students' learning of the concepts of matter.

Three Key Resources

The **first** study (Smith, Wiser, Anderson, & Krajcik, 2006; Smith, Wiser, Anderson, Krajcik, & Coppola, 2004) proposed, on the basis of several empirical studies, a learning progression aimed at the understanding of the particle model of matter. Their scheme was in six stages. It started with the classification of matter by its macroscopic properties at the first stage and extended to the capacity to distinguish between the properties and interactions of atoms and molecules, and the macroscopic phenomena which these properties can explain, at the last stage. This review describes the data in three sections, dealing in turn with the grade ranges K to 2nd, 3rd to 5th, and 6th to 8th, so that the scheme proposed was also evolutionary. Overall, the evidence on which this study was based comprised about 40 different publications, some of which were confined to one of the three grade ranges, while others spanned two or three of them. Each section discussed a set of assessment items, some of these were new proposals by the authors, while others were adapted from items used in studies reported in previous studies. The authors did not present any empirical data on the responses of students to these items, and the only evidence of their use in practice was the data in the publications from which some of them were adapted.

The authors were careful to make their reasoning "transparent" and one of their conclusions was "We hope that this transparency will provide evidence that the sample items we have included are representative of much larger pools of items that could be developed using similar methods and a broader sampling of the research base" (page 61). This review has been used as a starting point by many others: in particular, it has been quoted as a source in many of the studies reported below.

In the **second** of the three studies, Johnson and Tymms (2011) focused their work on the concept of substance. Building on earlier (Johnson, 1998) interview-based studies of students' understanding, this study analyzed the results of over 400 students aged 11–14 across 30 schools in England using fixed response items. Each of three school year groups (years 7, 8, and 9 with 10) attempted about 78 items, with overlap between the years for 55 of the items. They used Rasch analysis, with careful inter-calibration between years, to produce a two-dimensional map of item difficulty by conceptual content. Their map (p. 869) showed 52 groups of items with the conceptual content represented as five groups with focus on (i) properties and substances; (ii) mixtures; (iii) chemical change; (iv) the particle model and explanations; and (v) mass changes. One conclusion was that there was a significant increase in "substance ability" over the 3 years, but it was relatively small, that is, less than 3% of their total scale of item difficulty. While the map represents an overall learning progression for the concept of substance, its complexity is such that there is no clear line of progression.

In a later article, Johnson (2013) used data, selected from the same study and re-aligned the item difficulties to produce a map (p.61) with only 18 constructs drawn from the previous groups of items. These 18 were aligned to show three lines of progression. The first of these was about learning that particles are the substance, the second about particle motions in the three states of matter and transitions between them, and the third about identifying molecular structures by their atomic composition. The author linked these to the earlier study (Johnson, 1998) that described the basic changes in students' understanding to be between replacing the idea that particles are embedded in the substance by the idea that they are the substance, and from there to the idea that particles do not have the macroscopic characteristics of the substance. One feature of particular interest about this last idea was that students started to show acquaintance with the atomic structure of molecules before they had shown a clear understanding of that idea. In this study, while some student interviews were used to guide the development of the items, the only items used were multiple-choice items, reflecting the authors' previous comment that "Some free text responses were trialed, but answers were ambiguous and difficult to score" (2011, p.857). While 11 of the 18 items were used for all the year groups, 6 were designed for only two, and 21 for one. However, the mean scores of the three groups ranged from 50.1 9(±7.4) for the year 7 sample, to 54.9 (±8.6) for the years 9-with-10 sample, leading the authors to comment that, "It is perhaps surprising that different ages, schools and teachers did not produce a much greater variation in the responses. Even the patterns of choices amongst the distractor options were very consistent" (p.63). There is no mention in this article of any discussions with teachers about the pedagogic implications of the results.

In the **third** of these three, Hadenfeldt, Liu, and Neumann (2014) and later Hadenfeldt, Neumann, Bernholt, Liu, and Parchmann (2016), explore student learning of the structure of matter. In the former publication, Hadenfeldt et al. (2014) published a survey of studies about the structure of matter. Their survey lists 82 papers on the topic published since 1990, exploring the topic, with samples of between 11 and over 3,000, over various sets of grades between three and 13. Their summary of the models of progression described six different sets of levels variously proposing between five and two levels. They point out that some authors, such as Talanquer (2009), propose multiple dimensions without clearly defined levels. Their own broad scheme of five levels for understanding matter for kindergarten to grade 12, started with “naïve concepts” and ended with “systemic particle” concepts by way of “simple particle concepts.” Within this general scheme, they proposed four parallel yet overlapping constructs (which they called “big ideas”) covering (i) structure and composition; (ii) physical properties and change; (iii) chemical properties and change; and (iv) conservation, with a caution that research studies showed that understandings of these four were “highly inter-twined.”

In their 2016 paper, Hadenfeldt et al. (2016), using as a framework the four big ideas with five levels in each, explored the progression of students in grades six to 13 across five secondary schools in Germany. Their sample of 1,358 students was distributed across eight grades in the five schools: all these schools were at the highest level of the stratified national system, which meant that nearly all of these students would expect to progress to degree studies. In each school, some classes would be taking advanced physics and chemistry courses, some at a basic level only, while some would not be studying these subjects. The data were the results from 42 ordered multiple-choice items; for about a third of these, the classifications of the responses were checked in discussion with 11 students.

These results were based solely on the responses of a sample drawn from the upper levels of student achievement and on about 40 ordered multiple choice (OMC) items. They do not discuss any comparison between the variations in mean response levels between school years and the spreads in the responses within any 1 school year. The authors commented that with only 10 items for each of their four parallel schemes one “cannot draw meaningful comparison in understanding across the four big ideas.” There was no mention of any discussions with teachers—either about the items selected or about the pedagogic implications of the results. The authors also concluded that more research is needed on “how construct-relevant yet disturbing variance in item difficulty . . . can be controlled” (p. 704).

Critique of Published Work

Each of these studies produced a model to chart or to promote learners’ development from novice to expert in the understanding of that model, and each was based on empirical studies. However, the progression schemes proposed by these three did not align closely or clearly with one another. In formulating the basis of the present study, as set out in our previous paper (Black, Wilson, & Yao, 2011), the Smith et al. (2004) study seemed to align most closely with our model, while it also reported a set of items which could be used as a basis for building up data in the study’s sample schools.[1] Three elements that prompted us to undertake this study include (i) none of them report data which made significant use of students own expressions of their thinking—the views of small samples of students’ were only used to check the wording of fixed response items; (ii) the empirical evidence of mean differences between school years was that these are small compared with the spread of performances within any 1 year group; and (iii) none of them reported follow-up studies to explore the application of the results, either by feedback to teachers or by further empirical studies of their use in pedagogy.

Issues Which Any New Work Should Consider—Cautionary Considerations

Top-Down and Bottom-up Approaches to Learning Progression Construction. Among these and the many other studies of learning progressions, there appear ideas, which should be borne in mind in any attempt to construct a model to guide progressions in any topic. The discussion below will first outline three of these which bear directly on the task of advising teachers about the optimum design of a teaching program, and will then mention three others which are of more general relevance. The first idea arose in a broader review, encompassing many topic areas, by

Duschl, Maeng, and Sezen (2011). They stressed that in any attempt to construct and validate a learning progression, there are two possible approaches. The first is the “top-down” or validation approach where the aim is to supplant the learner’s own initial conceptions (or misconceptions), while the second is the “bottom-up” or evolutionary approach which builds on learner’s initial conceptions. Both approaches draw on conceptual change research. However, a study may draw on both approaches. For example, by starting with a top-down scheme and then modifying that scheme, first by aligning it with assessment items drawn from research into students’ learning, and later by amending the scheme in the light of the findings when students attempt these items. Initially, Wiser and Smith (2008) and then Duschl et al. (2011) also suggested that progressions might involve “stepping stones,” that is, points where appropriate instructional intervention by the teacher may be needed which would interrupt any linear sequence. This idea will be considered below in the discussion of our results.

Prior Knowledge of Student and New Knowledge. The second idea is a caution pointed out by Shavelson and Kurpius (2012), who defined a learning progression as “a sequence of successively more complex ways of reasoning about a set of ideas,” (p. 15). However, they also warned that: “learning progressions are not developmentally inevitable but depend on instruction interacting with students’ prior knowledge and construction of new knowledge.” This idea is also found in Chapter 8 of *Taking Science to School: Learning and Teaching Science in Grades K-8* (NRC, 2007). Thus, a set of components, which together form the construct of the particulate model of matter, may only be seen as linked in a coherent pattern when a high level of competence has been reached in all of these components. For example, the Hadenfeldt et al. (2016) paper mentions that “. . . as a result of being taught about the particle nature of matter, students attempt to integrate the idea of particles into their explanatory model” which suggests that their sample were told about the particle model and then had to relate it to evidence and experiences.

Interpretation of Student Results. The third idea, from Krajcik (2012), makes a similar point in stressing the need for caution in the interpretation of the results of surveys of students’ responses to any set of questions. He argued that some curriculum materials may lead students to memorize the particle model of matter as a fact rather than as an evidence-based model that can explain phenomena, whereas learning progressions should provide the tools needed to build on students’ current understandings so that they form richer and more connected ideas over time. The latter approach would work from observing macro properties of matter to seeing these in terms of microscopic models, a progression which would require learners to revisit the same ideas, on several occasions, within new contexts and levels of understanding so that their progression would not be a simple linear one (Stevens, Shin, & Krajcik, 2009). It follows that the instructional components, which should help teachers guide such progression, may play a key role in the implementation of any proposed learning progression. For several of the studies reviewed above, little or no evidence is presented about the teaching programs experienced by those who formed the samples for the research study.

Drawing Conclusions From Data. Of more general relevance to the interpretation of conclusions drawn from any set of data about student knowledge and understanding were three points. The first was a caution (Tsaparlis & Sevian, 2013) which pointed out that students have to become accustomed to working with inter-relationships between conceptual changes and the epistemological changes involved in working with a construct of matter as an inter-related set of linking macroscopic and sub-microscopic phenomena. The second was a point, made by Akaygun and Jones (2013), that in most progression schemes learners have to move from visual observations to symbolic representations, a progression which may well be helped by computer simulations. The third arises from the work of Denvir and Brown (1987); they showed that, in reaching a full understanding of the concept of number, students followed many different sequences of steps between the many intermediate stages, which had been revealed by analysis of interviews. The obvious conclusion was that it cannot be assumed that all learners will arrive at understanding of a key concept by the same route. They also reported that in their sample, which spanned 4 successive junior school years, "there is little relationship between performance and school year" (p.101).

Need for Interaction Between the Theoretical and the Empirical. Overall, it appears that several different models have been proposed, some formulated from careful analysis of empirical results, and some leading to advice about how teachers might use the findings to guide their own instruction. Many other papers have argued the need for progression schemes with aims variously expressed as supporting ambitious teaching practices (Furtak, Thompsom, Braaten & Windschitl, 2012) or for the design of state standards (Foster & Wiser, 2012). However, what has been lacking in most subsequent work in which any such model has been applied has been the use of the results to check the extent to which student learning has developed in accord with the model. As Krajcik (2012) commented, "the community needs to prevent force-fitting data to preconceived notions about learning progressions" (p. 34).

Methods

We approached the investigation of the learning progression using a construct modeling approach as recommended in *Developing Assessments for the Next Generation Science Standards* (National Research Council, 2014). This approach allows for an integrated means of developing assessments that support and build on the goals for curriculum and intentions of instruction. We used the BEAR Assessment System (Wilson, 2005; Wilson & Sloane, 2000) to provide guiding principles for our development of the learning progression of how students understand the structure of matter through a project named "The Learning Progression of Middle School Science Instruction & Assessment" (LPS).

The Learning Progressions in Science (LPS) Project

The Learning Progressions in Middle School Science Instruction & Assessment (LPS) project is a multi-year project funded by the Institute of Education Sciences (IES) of the US Department of Education. The purpose of LPS was to investigate (i) a learning progression for the structure of matter; (ii) a learning progression for scientific argumentation; and (ii) the relationship between these two progressions. This paper will focus solely on the first research goal—the learning progression for the structure of matter.[2] Broadly, we asked the research question: "What is the nature of the learning progression in the content domain of the structure of matter?" This is split up into three components:

(1) What is the nature of the structure of matter learning progression resulting from the previous rounds of iteration of the BAS?
(2) Is the hypothesized dimensionality of the constructs supported by the data?
(3) Is the ordering of the levels, as hypothesized in the construct map for each of the dimensions, supported by the data?

Figure 1 illustrates an initial version of this learning progression, which was first presented in Black et al. (2011), and was hypothesized to include six "constructs" illustrated as boxes in the figure. Of these six constructs, those numbered 1, 4, 5, and 6 are the core constructs for our topic.

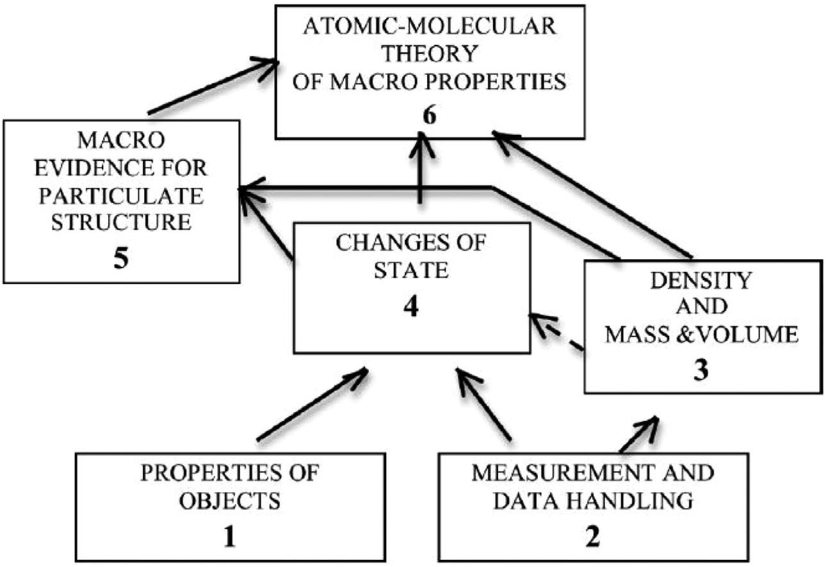


Figure 1. The hypothesized learning progression for the structure of matter. This was first presented in Black et al. (2011), pp. 83.

These constructs can be thought of as the important topics, or big ideas in the progression, such as understanding matter in terms of its macroscopic properties or understanding the atomic-molecular theory of macroscopic properties. In the figure, it is assumed that the progression moves upwards, with the simpler understandings at the bottom and the more sophisticated understandings at the top.

At the time of the initial development, the details of each construct were as yet unspecified. Then, based on further review of the literature, internal research team meetings, and suggestions by middle school science teachers, detailed descriptions of students' increasingly sophisticated thinking for each big idea were added. Further information on individual constructs can be found in Black et al. (2011) and Wilson, Osborne, et al. (2013). Each construct was individually investigated following a construct modeling approach (Wilson, 2005), called the BEAR Assessment System.

Assessment Development

The BEAR Assessment System (BAS) is based on the ideas of developmental assessment (Masters & Forster, 1996; Masters, Adams, & Wilson, 1990; Wilson, 2005). The elements of BAS are based on four principles, described in detail in Wilson and Sloane (2000), and listed here:

(1) A developmental perspective on student learning.
(2) A match between instruction and assessment.
(3) Management by teachers.
(4) Assessments that uphold high-quality standards of reliability and validity.

A key feature is that the system is centered on *constructs*—the "big ideas" around which a curriculum is structured. A construct is an achievement continuum defined operationally by the assessment tasks to which students respond, and that can be used to track student progress over time (Masters et al., 1990). These constructs each represent an important set of the learning goals of the curriculum, something which is repeatedly assessed and for which teachers will wish to have summary information at critical points during the school year. It is this construct modeling approach to assessment that promotes the developmental perspective. Information drawn from these assessments can be used formatively to inform decisions about student progress and about the next steps in instruction. Constructs are based on comprehensive reviews of the theoretical and research background, and their usefulness is verified through empirical analyses. Part of this work to extend and combine constructs into learning progressions is to provide a coherent and deeper view of students' understandings of science conceptions.

The BAS assumes that one is developing multiple tasks to chronicle students' learning over time using the following four building blocks: construct map, items design, outcome spaces, and the measurement model (Wilson, 2005). These four building blocks of BAS are shown in Figure 2.

A **construct map** "is a more precise concept than a construct" (Wilson, 2005, p. 6). The construct map outlines a developmental perspective to determine *how students are progressing* from less to greater expertise in the domain of interest, rather than using assessment only to measure generalized correctness after learning activities are completed. This initial step is usually accomplished through domain analysis that considers the extant literature, the particular goals of related curricula, teachers' expertise, input from other experts in the domain, and the theory guiding the larger learning progression. An essential tension when choosing construct maps is the tradeoff between coverage, which drives the creation of many construct maps representing every learning goal, and usability, which limits the number of construct maps that can realistically be learned by students and implemented by teachers. An example of this tension, discussed in our introductory literature review, is the comparison between the map produced by Johnson and Tymms (2011) and the more limited map published by Johnson (2013). It is, therefore, paramount to identify the most important learning trajectories to represent as construct maps.

After the construct map is defined, it is operationalized by utilizing tasks that prompt each student to provide evidence of where that student is located on the construct of interest. This is achieved through the **items design**, the systematic design of tasks to elicit the specific types of evidence about the levels of student knowledge as described in one or more construct maps. Each item is designed to engage a student and tap his/her knowledge or understanding of not only the construct(s) of interest but a particular level within the construct map. The items go through an iterative development and quality control process (Wilson, 2005) to ensure high quality and adequate coverage of the construct map.

These student responses to items are then mapped onto the **outcome space**, which defines the qualitatively different levels of responses (of the construct map) relative to a particular prompt or stimulus. Essentially, this is where a value is placed on student work. Generally, the scores will have a several-to-one relationship to the construct map levels, although sometimes it will be one-to-one. Scoring guides or rubrics are often used, along with exemplars of student performance at developmentally important levels. Scoring guides are hierarchical in nature. A higher score

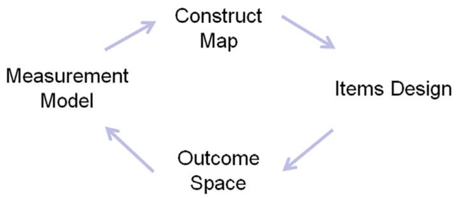
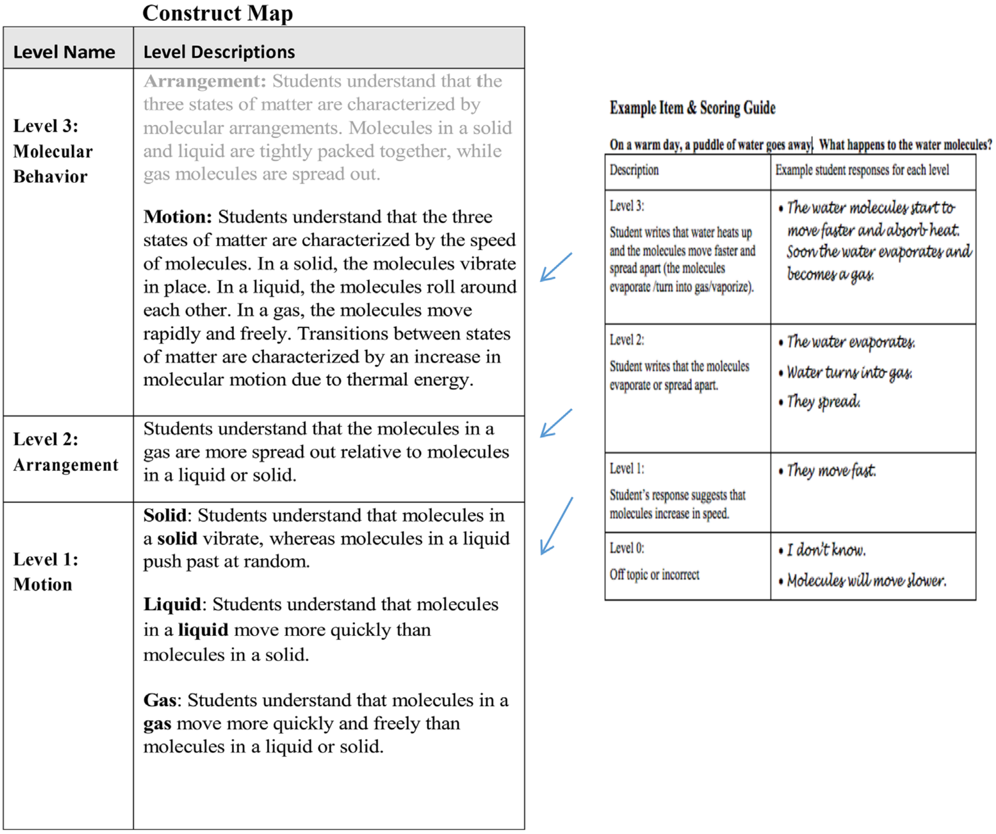


Figure 2. The BEAR Assessment System (BAS) from Wilson (2005) and Wilson and Sloane (2000). [Color figure can be viewed at wileyonlinelibrary.com].

represents a qualitatively better performance: not just more factual knowledge, but a deeper understanding. This also reflects the developmental perspective of the assessment system. This building block operationalizes the principle that teachers are to be the primary managers of assessment in the classroom. To accomplish this, they must have not only collected the data needed to assess student learning, but they also need to master the skill set required to use these data effectively. This implies a data-driven approach to assessment and teaching, in which teachers use assessment evidence to draw inferences about student knowledge and understanding. Figure 3 shows an example construct map plus an example item showing how typical item responses can be mapped to the levels of the construct map.

The **measurement model** defines how inferences about student understandings are drawn from the evaluated (scored) work. In other words, this is where the values derived from the outcome space are translated back to the construct map. There are many models available for analyzing the evaluated work, such as item response models, latent class models, or factor analysis. We will illustrate how this works for a Rasch analysis below.

Note that the BAS takes multiple iterations to implement. The results from the measurement model often provide important guidance for improving the descriptions of the construct map, and thus, one must cycle through BAS again and make adjustments to any of the four building blocks, as needed.



Construct Map

Level Name	Level Descriptions
Level 3: Molecular Behavior	**Arrangement:** Students understand that the three states of matter are characterized by molecular arrangements. Molecules in a solid and liquid are tightly packed together, while gas molecules are spread out. **Motion:** Students understand that the three states of matter are characterized by the speed of molecules. In a solid, the molecules vibrate in place. In a liquid, the molecules roll around each other. In a gas, the molecules move rapidly and freely. Transitions between states of matter are characterized by an increase in molecular motion due to thermal energy.
Level 2: Arrangement	Students understand that the molecules in a gas are more spread out relative to molecules in a liquid or solid.
Level 1: Motion	**Solid**: Students understand that molecules in a **solid** vibrate, whereas molecules in a liquid push past at random. **Liquid**: Students understand that molecules in a **liquid** move more quickly than molecules in a solid. **Gas**: Students understand that molecules in a **gas** move more quickly and freely than molecules in a liquid or solid.

Example Item & Scoring Guide

On a warm day, a puddle of water goes away. What happens to the water molecules?

Description	Example student responses for each level
Level 3: Student writes that water heats up and the molecules move faster and spread apart (the molecules evaporate /turn into gas/vaporize).	• The water molecules start to move faster and absorb heat. Soon the water evaporates and becomes a gas.
Level 2: Student writes that the molecules evaporate or spread apart.	• The water evaporates. • Water turns into gas. • They spread.
Level 1: Student's response suggests that molecules increase in speed.	• They move fast.
Level 0: Off topic or incorrect	• I don't know. • Molecules will move slower.

Figure 3. Sample construct map and associated sample item with scoring guide (Wilson, Black, et al., 2013). [Color figure can be viewed at wileyonlinelibrary.com].

Research Question 1: Structure of the Learning Progression

The assessments went through a rigorous development process, cycling through BAS for multiple iterations where researchers began with theoretical constructs and changed them or considered changing them based on empirical evidence. Through these iterations, researchers developed and refined the constructs, items, and scoring guides associated with the structure of matter learning progression working closely with scientists, science education professionals, and elementary and secondary school educators.

Research Question 2: Dimensionality of the Four Core Constructs

We used the multidimensional random coefficients multinomial logit model (MRCMLM) (Wang, Wilson, & Adams, 1997) to investigate the constructs together as a representation of the learning progression. The MRCMLM was chosen because of its flexibility in modeling a wide range of Rasch models, including the multidimensional extension to the partial credit model. Each construct was treated as its own distinct dimension and the MRCMLM allows for these dimensions to be non-orthogonal (i.e., oblique). More specifically, to test the constructs of the learning progression simultaneously, a between-item multidimensional model was used (Wang et al., 1997), since each item on this assessment was assumed to measure only one latent dimension. For model identification, the student ability mean for each dimension was constrained to 0.00.[3] Because of this constraint, direct comparisons across dimensions cannot be made, since it is unreasonable to assume that all dimensions have the same origin.

Research Question 3: Ordering of the Construct Maps for the Four Core Constructs

Beyond the dimensionality of the four core constructs, we also investigated whether the levels in the construct maps were empirically ordered according to the hypothesized construct maps. Delta dimensional alignment (DDA; Schwartz & Ayers, 2011) is one technique where the item parameters from a multidimensional model are transformed to the same logit metric. After transforming item parameters, these can then be used as item anchors to estimate the student ability distributions so that these are also on the same metric. After the transformation, both the student ability and item difficulty distributions can be directly compared. All analyses were done in ConQuest 4.0 (Adams, Wu, and Wilson, 2015).

Sample

Research was conducted with middle school students from a large urban school district in the western part of the United States. At the time of this research, there were 10,131 students enrolled in middle school with 3,766 being in eighth grade. The diversity of the student population is reflected in the 44 different languages spoken within the district and the 26.5% of students who speak English as a second language. Demographic information shows that there are nearly equal numbers of boys and girls, and that the student body consists of African-Americans (11%), Chinese (28%), Filipino (6%), Latino (26%), White (11%), and other (18%). Fifteen percent of students are English Language Learners, 13% have an individual education plan (IEP), and 33% are in the gifted and talented education (GATE) program. Fifty-eight percent receive free or reduced lunch.

The data and results reported in this paper represent the third iteration of data collection. A total of 16 students, identified by eight teachers, were also interviewed as a part of the pilot study. Two teachers helped us trial materials before the pilot study using a sample of 97 students. Lastly, 11 teachers from eight schools helped us to conduct a general administration with their students for a total of 1,087 eighth grade students. Students completed one of four test forms. A total of 128 content items were represented on the test forms. In addition, scientific argumentation items also appeared on three of these forms, but were not included in this particular analysis, as they were not part of the hypothesized structure of matter learning progression.

Students in grade 8 were chosen because the *Framework* (NRC, 2012, Box 5.1 on page 105) and *NGSS* (NGSS Lead States, 2013, Disciplinary Core Idea—PS1.A and PS1.B) identify the structure of matter as a core idea in science and students in that grade received specific instruction in the structure of matter (e.g., the nature of matter, states of matter, atoms and bonding, chemical reactions, etc.) throughout the school year from their teachers. In other iterations of this project, sixth graders and tenth graders were also sampled, though the overall sample sizes were smaller. While our iteration for the results reported here only contained eighth graders, we found that—due to the large and diverse sample—the range of performance was wide enough to provide substantive insight into the progression, particularly toward the latter ends of our progression, as that content coincides with what students are learning in the classroom.

Results

Research Question 1: Structure of the Learning Progression

Cycling through BAS for multiple iterations, researchers developed and refined the constructs, items, and scoring guides associated with the structure of matter learning progression. The cycle of development began by consulting the literature (see the background section above), and, based on that review, choosing individual constructs to study separately. For each construct chosen, the research team, made up of experts in science, science education, and/or measurement, and a group of middle school science teachers then drafted construct maps. The construct maps included hypothesized hierarchical and qualitatively distinct levels (also based on the literature review).

Pilot Phase. The initial constructs were principally informed and inspired by the work of Smith et al. (2004). We developed or adapted tasks and began to consider how student responses might align with the specific levels on the construct map. We used both open-ended (e.g., written response and drawing) and fixed-choice items (e.g., multiple choice). An example of the kind of open-ended item is shown in Figure 3. The research team participated in a peer review of the materials. In these meetings along with meetings with teachers, we refined the assessment materials, which included the construct maps, items, and scoring guides. After making adjustments as needed, we then conducted think-aloud interviews with middle school students to ensure that the items were being understood as intended.

Through this process many issues arose ranging from minor wording choices to the reconsideration of the hypothetical structure of a construct. Sometimes the research caused teachers to reflect on their teaching and how their own students would respond to particular items. For example, the puddle item (see the right side of Figure 3) initially asked: “On a warm day, a puddle of water disappears. What happens to the water?” By the time it was administered to a large group of students it had been changed to ask: “On a warm day, a puddle of water goes away. What happens to the water molecules?” At first glance, these versions appear to be similar. However, the absence of the word “molecules” in the initial version proved problematic. Researchers were interested in eliciting student responses that communicate that water molecules move faster and spread apart when heated. However, because the question asked about “water” and not “water molecules” researchers found that most students repeated the prompt to say that the “water goes away.” Participating teachers decided to use this scenario in class to see students’ initial ideas about how matter changes states—whether students talked at the macroscopic or microscopic levels shaped how the teachers delivered the subsequent lesson.

Trial Phase. After refining the materials further, a trial was administered to a small group of students under normal test-like conditions. This was done to ensure adequate time for test-taking, identify any additional flaws, gather preliminary student responses, and (again) ensure that students were perceiving and responding to items as anticipated by the research team. After the trial and further revisions were completed, a general administration of items was prepared.

General Administration Phase. The general administration is the main data collection phase and consists of a substantially larger sample of students than in the previous phases. Scoring guide moderation, scoring, and data cleaning were then completed. Finally, the ConQuest software (Adams et al., 2015) was used to estimate student and item parameters for the data. Specifically, the partial credit model (Masters, 1982), a Rasch-based item response model that could handle assessments with both dichotomously and polytomously scored items, was used to investigate the technical qualities of each construct individually.

Rasch models provide convenient and rich ways to model person proficiency and item difficulty measures using the same scale. In addition, items from different types of assessments can be scaled together so that student gains can be evaluated in a straightforward way without requiring students to take the same pre- and post-test. This approach can improve the interpretability of student work and help teachers focus on the specific needs of their students in the context of a curriculum's central learning progression, while ensuring (i) that appropriate evidence is produced to draw reliable inferences about the student proficiencies of interest and (ii) that those inferences can be interpreted in a straightforward way, meaningful for teachers, students, and other stakeholders. The partial credit analyses for each construct revealed how well the empirical results fit with the original construct map. This included examining the item locations (e.g., were the items linked to lower levels actually easier for the students?) and item fit —as determined by the weighted mean square fit statistics (Adams et al., 2015). Items were flagged and further investigated qualitatively if any of these appeared problematic. Certain item features, such as potentially misleading pictures, confusing sentence structures, or complicated words, were examined closely. We relied on many different resources to help investigate these flagged items, including additional interviews with students and teachers, as well as internal team meetings. Following these meetings, some items were adjusted and tested again in the next cycle, while others were set aside if they were deemed too problematic by our research team's judgment. Adjustments to the construct map were also made following these analyses and another cycle through BAS began, following analogous procedures as described above. The empirical results presented here for our learning progression results from three years of cycling through BAS. While a description of the specific changes to all of the constructs is beyond the scope of this paper, an example for one construct is provided later in this paper.

Over the course of the research project, construct maps articulating the constructs were named, defined, renamed, and redefined, following advice from the project advisory board,[4] further literature review, and the results from the partial credit analyses. Four construct maps named: Macro Properties (MAC), Changes of State and other Physical Changes (PHS), Particulate Explanations of Physical Changes (EPC), and Particulate Explanations of Chemical Changes (ECC) were identified as core to the learning of the structure of matter and two [Measurement and Data Handling (MDH) and Density and Mass & Volume (DMV)] were identified as auxiliary. The focus of this paper will be on the four core constructs.

The structure of matter learning progression begins with MAC. Students, at this point, identify and classify matter using macro properties. Building on these initial ideas, students at the PHS level are those who understand the conservation of mass and volume during phase changes and other physical changes. Finally, students wrestle with understanding the particulate nature of physical changes (EPC) and also chemical changes (ECC). The constructs and their hypothesized relationships are shown in Figure 4, which illustrates the large-scale vision of the learning progression before the final large-scale data collection and analysis. Figure 4 represents the framework on the structure of the learning progression.

The four core constructs in Figure 4 show an increasingly sophisticated understanding of the structure of matter and mirror the grade 8 curriculum. Students are introduced to physical and chemical changes at the beginning of the school year by their teachers through a variety of approaches. Ideas central to understanding the structure of matter are discussed repeatedly throughout the school year, beginning with simple and familiar examples and activities the progressing through the school year to more complex and less familiar ideas. Initially, they are (re)introduced to matter and how to distinguish matter from non-matter, which reflects the contents of the MAC construct. The students, then, learn about and experiment with mixtures with a focus on the states of matter (PHS). They transition into describing how matter can change from one state to another and back again during a physical change which bridges ideas from the PHS construct map to the EPC construct map. Finally, students are taught how to distinguish between physical changes and chemical changes (spanning EPC and the last construct, ECC) with a focus at the atomic level so they can determine the number of atoms of each element, distinguish between reactants and products, and learn about conservation in a closed system.

After the development and refinement of the four core constructs through BAS, it was determined that they were each reasonably well-defined and the remaining items had reasonable item properties (e.g., good item fit, difficulties seemed reasonable in relation to other items). The next step involved investigating the relationship of these constructs to each other so that evidence for the learning progression as a whole could be gathered.

Research Question 2: Dimensionality of the Four Core Constructs

To empirically test our research question about the nature of a learning progression of the structure of matter, we began with a dimensional analysis of the four core constructs. This analysis allowed us to see whether the internal structure of the empirical results matched our hypothesized structure, shown in Figure 4. Two models were compared: a four-dimensional model, where each construct was treated as its own dimension, and a unidimensional model where only one latent dimension was assumed. Because these two models are nested, a likelihood ratio test was used to

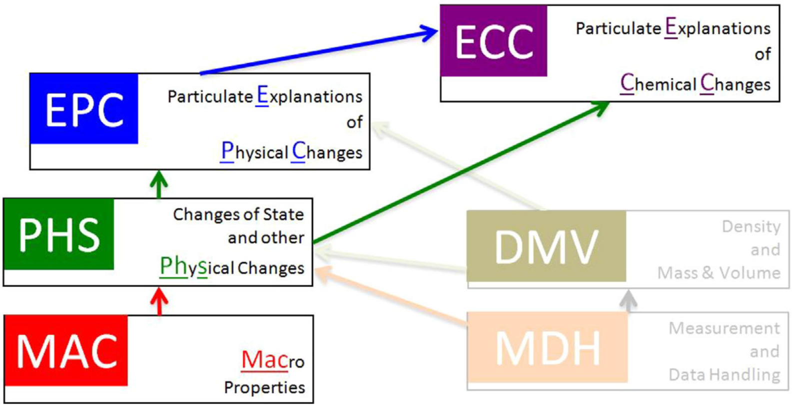


Figure 4. Revised learning progression for the structure of matter. [Color figure can be viewed at wileyonlinelibrary.com].

test the goodness-of-fit. The results were statistically significant[5] ($\chi^2 = 687,\ df = 9, p < 0.001$) indicating that the multidimensional model fits the data statistically better than the unidimensional model.

In addition to the likelihood ratio test, two additional fit criteria were used to compare the models, the Akaike Information Criterion (AIC; Akaike, 1974) and the Bayesian Information Criterion (BIC; Schwarz, 1978). Both are derived from the differences in the log-likelihood while also including a penalty for the number of parameters. Additionally, the BIC also accounts for the sample size. For both criteria, lower values indicate a better fit. Results, shown in Table 1, confirmed that the multidimensional model fits the data statistically significantly better than the unidimensional model using these indices, providing support that the distinctions between the four constructs are useful to acknowledge.

Because direct comparisons of the item and student distributions cannot be made without a transformation to the parameters for the multidimensional model, delta dimensional alignment (DDA; Schwartz & Ayers, 2011) was applied to the parameters. Tables 2 and 3 show the disattenuated correlations and the descriptive results, respectively. The number of items per construct ranged from 15 for PHS to 43 for ECC. We considered the effect size of the multidimensional models expressed here via the correlation coefficients shown in Table 2.

The correlations are quite high for some of the constructs. This finding was anticipated, as the constructs are part of the same learning progression. Nevertheless, we consider that, in the main, the correlations are in a range that warrants maintaining the dimensions for educational purposes. The highest correlation is between MAC and PHS, with a correlation of 0.93 and the lowest is between PHS and EPC, with a correlation of 0.81. Nevertheless, we decided to investigate two additional models, to see if combining dimensions did, indeed, make the results clearer. First, we tested a three-dimensional model, with MAC and PHS combined to form one dimension, while EPC and ECC remained separate. We also tested a two-dimensional model, with MAC and PHS combined and EPC and ECC combined, because these represent the highest correlations. The results are illustrated in Table 3.

Of these four models, the four-dimensional model still fit the data the best, as it has the lowest AIC and BIC values. Because the four-dimensional model was the best-fitting and since it follows the structure of matter learning progression, it will be discussed further. As Table 4 shows, the average means for students on MAC, PHS, EPC, and ECC were −1.86, −0.71, −0.31, and −0.19 logits, respectively. The variances for the constructs range from 0.77 for ECC to 2.26 for MAC. Although some of the constructs have only a moderate number of items, the reliabilities are quite reasonable, due to the advantages of using the multidimensional estimation, which uses all of every students' responses for each of the construct estimates: The EAP/PV reliabilities were 0.72, 0.75, 0.83, and 0.78, for MAC, PHS, EPC, and ECC, respectively.

The Wright Map from the multidimensional partial credit analysis, after DDA was applied, is shown in Figure 5. The Wright Map is a useful tool that illustrates both the distributions of the person abilities and the item difficulties on the same figure, using the same scale. The first column is the scale for the map, which applies to all four constructs, and ranges from approximately

Table 1
Comparisoin between multidimentional and undimensional models

Model	Number of Parameters	Final Deviance	AIC	BIC
Unidimensional	138	72693	72969	73112
Multidimensional	147	72006	72300	72452

Table 2
Structure of matter- construct correlations

	MAC	PHS	EPC
MAC	1.00	–	–
PHS	0.93	1.00	–
EPC	0.82	0.81	1.00
ECC	0.84	0.84	0.88

−6 to +4 logit. For each construct, there are two columns of information: the first is the distribution of the student abilities and the second is the distribution of the item difficulties. Each "X" in the student ability distributions represents approximately ten students. Higher logit values (being higher on the map) indicate higher ability levels on the construct for the students and more difficult items. In addition, the map provides a clear picture for how the group performed on each set of items. When a person has the same estimated ability level as an item, then the person has a 50% probability of answering that item correctly: We interpret this as indicating the point of "most active learning" for that student. If an item has a higher logit value, then the person has less than a 50% chance of answering the item correctly. Lastly, if the item has a lower logit value, then the person would have more than a 50% chance of answering that item correctly.

In our results (see Wright Map, Figure 5), we can see that an eighth grade sample was reasonably well-mapped to the (entire) set of items across the four constructs, and hence, able to provide us with valuable insights into the learning progression. For instance, for both the MAC and PHS person distributions (i.e., the two left-most histograms), there are some students who are matched with items in these constructs. While MAC had about seven items where students all have over a 50% probability of answering correctly, there are still a substantial number of items where students have a 50% probability or less of answering correctly. Thus, while most eighth graders found MAC and PHS items generally easy, there is still a sizeable number of students who may still experience some difficulties.

For both the EPC and ECC constructs, the distribution of the items matches the distribution of the students more evenly than both MAC and PHS. This was anticipated since the students are learning about particulate explanations for both physical and chemical changes in eighth grade. For these two constructs, there is considerable overlap of both the distributions of student abilities and of item difficulties, suggesting that they have similar levels of difficulty. However, the similarity of the item difficulties for these two constructs was a surprise, as it was initially hypothesized that the items relating to physical changes (EPC) would be easier than items on chemical changes (ECC). This pattern has also occurred in a previous iteration (Yao, Wilson, & Black, 2013). This finding is explored in the "Discussion and Implications" section.

Table 3
Comaparision between undimentional, two-, three-, and four-dimentional models

Model	Number of Parameters	Final Deviance	AIC	BIC
Undimentional	138	72693	72969	73112
Two-Dimentions (MAC/PHS/EPC/ECC)	140	72295	72575	72720
Three-Dimensions (MAC/PHS, EPC, ECC)	143	72078	72364	72512
Four-Dimensions (MAC,PHS,EPC, ECC)	147	72006	72300	72452

Table 4
Descriptive results for the four dimentions

	MAC	PHS	EPC	ECC
Number of items	39	15	31	43
Person mean	-1.86	-0.71	-0.31	-0.19
Variance	2.26	1.63	1.32	0.77
EAP/PV reliability	0.72	0.75	0.83	0.78

Research Question 3: Ordering of the Construct Maps for the Four Core Constructs

As the ordering of the construct levels is also an important part of the empirical testing of the hypothesized construct maps, we also examined this question, using more detailed versions of the Wright maps. We found that the levels within the dimensions did not align as well as we had expected they would with the hypothesized levels in the construct map. Hence, we explored the possibility of a sub-structure within the item sequences in each construct. We started with a re-examination of the results from the partial credit analyses and also included a qualitative examination of the questions within the construct and considered whether they might be grouped in relation to similarities within the items of each group. The qualitative work was done through our internal research meetings with our research team and taken to teachers for their feedback. The outcome of this type of exploration is illustrated for the ECC construct in Figure 6. Similar efforts have been made for the EPC and PHS constructs, and discussed in other publications (Black et al., 2011; Wilson, Black, et al., 2013).

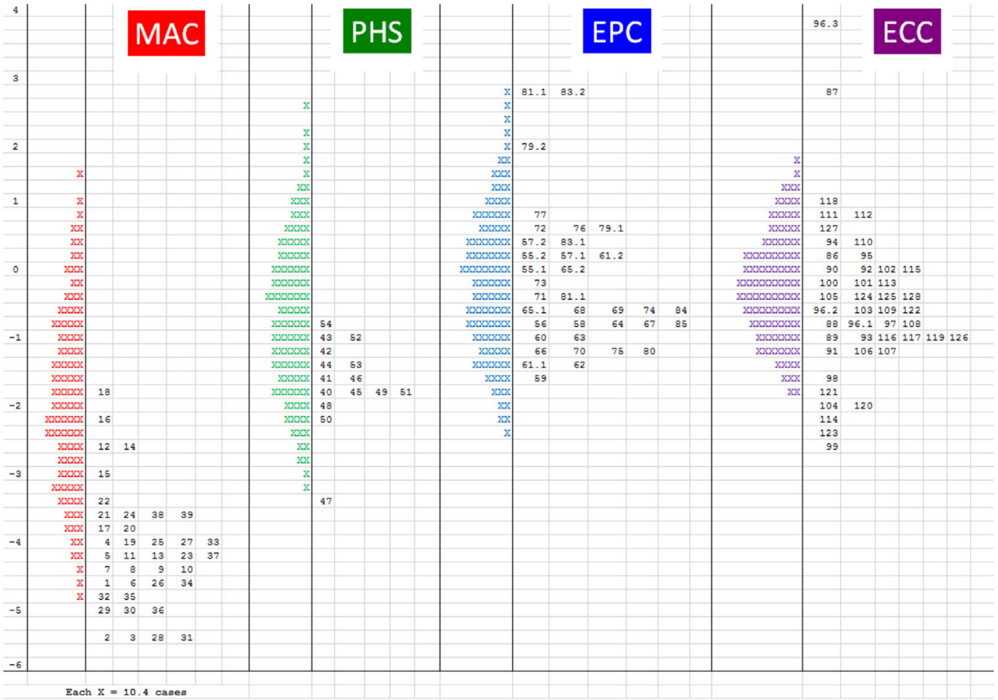


Figure 5. Wright Map of the four main constructs in the structure of matter learning progression. [Color figure can be viewed at wileyonlinelibrary.com].

The initial hypothesis of the Particulate Explanations of Chemical Changes construct map started at the lowest level (Level 0) with the student misconception that matter is not conserved during any type of substance change. Researchers hypothesized that student understanding moved through levels of increasing understanding until Level 5, which states that students understand that during a chemical change, the atoms will be unchanged, but may combine in new ways to form different molecules (new substances). The empirical results led the research team to re-conceptualize the progression. Researchers worked with teachers to interpret the data and develop an alternative construct map that captured a more comprehensive view of how learning develops. The revised construct map shows that students' understanding of chemical change develops along three lines (or strands).

The ECC items can be categorized into three groups, which are labeled as strands, with each strand being a distinct component of the main ECC construct. These three strands are described as follows:

- ECC-A: Chemical and physical changes in the inter-atomic combinations and in the arrangements of atoms and molecules.

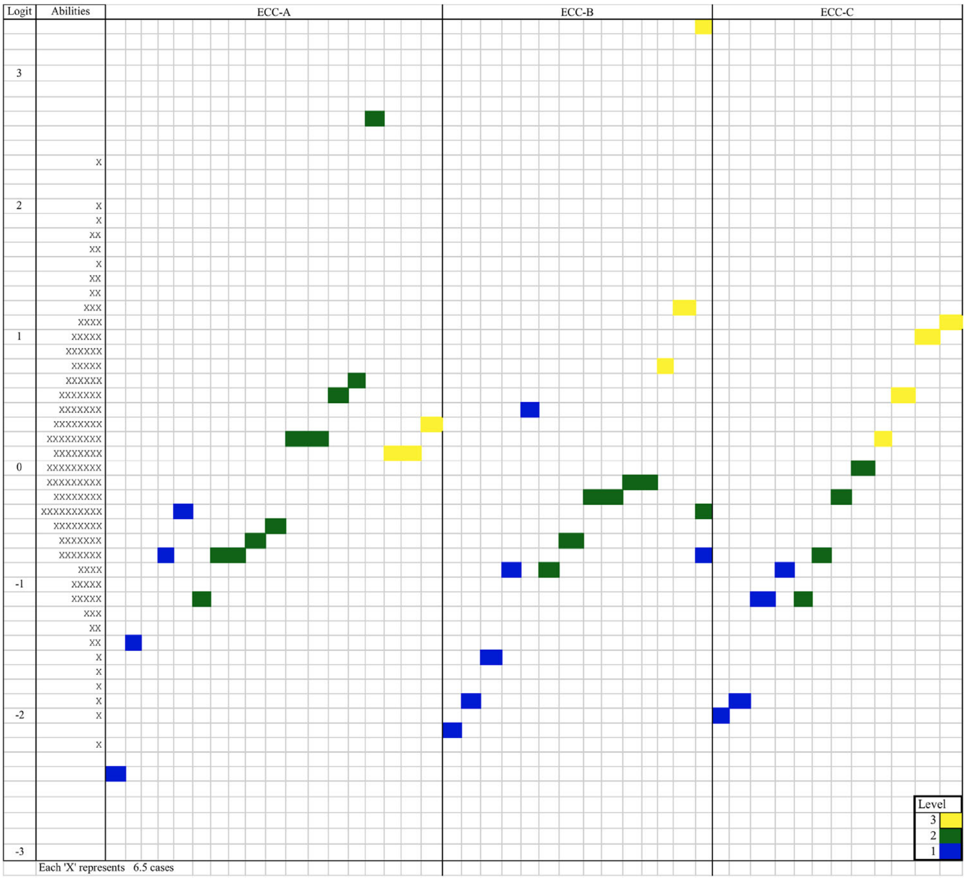


Figure 6. Wright Map of the strands of the ECC construct. [Color figure can be viewed at wileyonlinelibrary.com].

- ECC-B: Changes in macroscopic properties, which accompany chemical and/or physical changes.
- ECC-C: Representations of elements, compounds, and different phases, in terms of arrangements of atoms and molecules.

In ECC-A, students can explain chemical and physical changes in terms of the inter-atomic combinations and the arrangements of atoms and molecules. At the first level, students can recognize and explain molecular and atomic representations of physical and chemical changes. In the second level, students understand that in a chemical change, the atoms or molecules change the way they combine and form new materials. Students also know that there are empty spaces or a vacuum between atoms and molecules. Lastly, in the third level, students can construct diagrams to give molecular and atomic representations of both physical and chemical changes. They know that in a chemical change, the atoms stay the same, but in different molecules. The numbers of atoms involved remain the same for both physical and chemical changes.

In ECC-B, students can distinguish between chemical and physical changes through the observations of macroscopic properties. For the first level, students can distinguish between these changes for obvious, familiar cases, such as when a chunk of wood burns or when an ice cube melts. In the second level, students can do this for less obvious cases, such as when a candle is burned or when sugar is dissolved into tea. Students also understand that mass is conserved throughout these changes. In the last stage, students can recognize that the properties differ after a chemical change. They can also distinguish chemical from physical changes in unfamiliar cases, such as when solid iodine turns into a gas.

In ECC-C, students are able to represent elements, compounds, and different phases, in terms of the arrangements of atoms and molecules. At the first level, students can recognize representations of monatomic molecules. They are aware that atoms and molecules are not usually visible and may also have different sizes. In the second level, students can recognize representations of diatomic molecules. They also know that atoms and molecules have weight. At the topmost level, students can recognize and distinguish between diagrams of different polyatomic elements and compounds.

As just mentioned, within each of these three strands, the items could be divided into three levels, with the type of demand similar within each group, but different, both qualitatively and in expected level of demand, between the groups. The outcome of this analysis is shown as a Wright map in Figure 6. Here the three strands are shown separately, but with respect to the same overall student score scale for ECC; so this map is the same as the map for ECC in Figure 5 but with the additional division into three strands. The other feature added here is that the three groups with each strand are represented by the colors—blue, green, and yellow, representing the three levels from the lowest to the highest, respectively.

Figure 6 shows evidence of a progression sequence within each strand, but also shows that the overlaps between the three produced the apparent lack of such sequence when they are represented together, as in Figure 5. The colors of questions also indicate a progression sequence, in a clear way for ECC-C but not for ECC-A. Such apparent "discrepancies" will have to be explored by more detailed examinations of the ways in which the demands of individual questions are perceived by students. Further work will also be needed on the possible strands within the other constructs.

Discussion and Implications

Learning progressions can be a useful tool for designing high quality assessments that align to both curriculum and instruction, but because of their hypothetical nature, they must be tested for reliability and validity. For learning progressions to have a positive impact in the field of education, they must begin with a sound theoretical frame and then be verified through empirical testing. This paper illustrates one such method for hypothesizing a learning progression of how students come to understand the structure of matter and empirically testing the learning progression using a multidimensional model to test the relationships among constructs within the progression. The methodology involved following the BAS framework, where first, each individual construct within the learning progression was investigated. After each construct had been investigated once or more and determined to have sound psychometric properties, then it was possible to investigate the learning progression as a whole. While this procedure is both time-consuming and resource-intensive, efforts of this order are essential for gathering high quality empirical evidence. For instance, it is through these earlier analyses that four constructs were identified as the core constructs in the progression, whereas two of them were identified as auxiliary constructs. While these two auxiliary constructs should also be investigated alongside the core constructs for future research, this designation allowed us to focus our limited resources on what we considered the most important relationships in the learning progression.

This section begins by addressing our overall research question: What is the nature of the learning progression in the content domain of the structure of matter? First, we discuss the results regarding the structure and dimensionality of the learning progression. Next, we discuss the order of the construct maps and the discovery of an underlying substructure in the ECC data set. This discussion is followed by a review of the current study with respect to earlier research on student understanding of the structure of matter and of learning progressions in general.

Structure and Dimensionality of the Learning Progression

The empirical multidimensional results displayed an overall pattern that supported our initial expectations of the structure and dimensionality of the learning progression. First, the four-dimensional model had the best fit over a three-, two- and one-dimensional model, providing empirical evidence to our four core-construct learning progression. In addition to the number of dimensions, the structure of the dimensions seemed to fit our progression as well. The MAC items were found to be the easiest, followed by the PHS items, and finally, the EPC and ECC items were the most difficult. However, there was one major exception to the structure—the two most difficult constructs were not ordered in difficulty as we hypothesized. Specifically, the items about physical changes were not found to be easier for students than the items about chemical changes. Rather, their difficulties appeared similar and this pattern is discussed in more detail in the next section.

Order of the Constructs

While the hypothesized progression holds generally, the items for the Particulate Explanations of Physical Changes (EPC) construct and the items for the Particulate Explanations of Chemical Changes (ECC) construct were not found to be ordered in difficulty in accordance with our hypothesis. We originally expected the items about chemical changes to be more difficult for students to answer correctly than items about physical changes. However, results indicated that the difficulties were in a similar range—a pattern we had also found in a previous iteration of this project (Yao et al., 2013).

A small-scale investigation of the items and constructs combined with discussions with teachers on how they teach these two topics helped shed some light on this finding. Resulting from these informal steps, we identified two points to consider: (i) the expectation of the order was influenced by experts' knowledge that further studies in chemistry would lead far beyond the simple physical phenomena involved in the ECC construct items; however the actual ECC items were not representative of that higher knowledge, as they are designed to address only the lower chemistry phenomena as specified in typical middle school science curricula; and (ii) many of the items that were included actually asked students to identify (implicitly) the difference between physical and chemical changes, and hence it makes sense that such items will be of similar difficulty whether they relate directly to EPC (i.e., because the correct response involves a physical phenomenon) or directly to ECC (i.e., because the correct response involves a chemical phenomenon). For future research, this finding and these speculations should be retested to see if the results hold true again, and whether the speculations are borne out.

Discovery of a Substructure Within a Construct

The discovery of a substructure, the three strands, associated with the ECC construct is a unique finding of this study. Initially in ECC, the overlapping item difficulties obscured the developmental sequence of progression at the construct level. After examining the content of the ECC items, we divided the ECC items into three strands with similar task demands. Upon further examination, each of the three strands were found to fall into three qualitatively distinct levels of understanding. Figure 6 shows the resulting three strands of ECC, along with the three levels within each strand. Once these strands and levels were identified, a developmental progression within ECC became clearer, though more so for some strands (i.e., ECC-B) than others (i.e., ECC-A). The discovery of the substructure was critical to learning how middle school students come to understand the particulate nature of chemical changes. While the substructure provides insight into the differences of task demands, for instance distinguishing between physical and chemical changes through macroscopic properties (ECC-B) and of explaining these changes in terms of the interatomic combinations of atoms and molecules (ECC-A), more research is required to explore the levels within each strand.

The Current Study Given Earlier Studies

This section discusses some of the key aims and methods of this study in the light of existing published work in order to highlight its specific contributions to the fields of learning progressions and science education.

By contrast with the three main published studies described in our background section, the present study restricted its scope to the particulate model of the structure of matter. In consequence, we have elicited a progression which has a more clear conceptual unity, while still covering a significant and central element of the science curriculum as a whole.

While the six big ideas identified by Smith et al. (2004, p. 11) inspired our learning progression (see Figure 4), we moved beyond that analysis by collecting empirical evidence to test the progression. Essentially, we responded to their invitation to investigate a learning progression of how students understand matter and molecular theory.

Many of the constructs produced in earlier studies can be related to the present findings, but the overall structures are significantly different. For example, while Johnson (2013) took a uni-dimensional approach to investigating his learning progression we took a multi-dimensional one, through which, on the basis of our empirical analyses of students' responses we have shown that their conceptions of the structure of matter are more validly represented by four distinct yet related constructs. Again, while Hadenfeldt et al. (2016) proposed four parallel overlapping constructs, they concluded that one "cannot draw meaningful comparison in understanding across the four big ideas."

A further new feature has emerged through our incorporation, in our data collections and analyses, of students' responses to open-ended test items in addition to their choices in multiple-choice items, whereas the published studies have relied exclusively on multiple-choice items. Hadenfeldt et al. (2016) and Johnson and Tymms (2011), have drawn attention to the problem that success with any multiple choice item may not be firm evidence of understanding. In these studies, the validity of the multiple choice items has been checked with students' explanations in either interview or writing, and such checks have been used to modify some multiple choice items. However, they have not incorporated open-ended explanation in their large-scale surveys. The use of both types of response in the present study has added information about the significant differences between the two in our levels of analyses, and thereby adds to our advice to teachers to not just ask students to choose an answer but to also to ask "why" questions to investigate the limits of student understanding. We provide samples of items in the Supplementary file. This expanded way to gather response data is beneficial to the interpretations of the items in their own assessment work.

However, proceeding to the design and implementation of these strategies in a curriculum and pedagogy program is beyond the scope of the present project, but we do see it as an important next step. Our items, as located in the Wright maps, meet the requirement of Krajcik (2012) by providing the tools needed to build on student's current understanding, and by inter-relating these within its broad framework; thus, they should help to meet his criterion that such tools should help build, over time, a set of connected understandings. In practice, as Shavelson & Kurpuis's (2012) emphasized, any individual student's learning progression will depend both on any new instruction and on his or her prior knowledge. In particular, a teacher may be working with a class who have previously been taught in a variety of ways, including top-down approaches, which could mean that they know the final "answer" for the issue to be explored for a construct but do not understand it. So, for example, in starting the study of a new topic within a classroom, the teacher could assess the knowledge and understanding of what students "bring to the table" by using either some of the basic level questions for the new topic, or some of the top-level questions from constructs which are a required basis for that topic, so chosen that success with them is required by the new topic. Such exploration of the nature of students' understanding is a key way of using the established benefits of formative assessment practices (Black and Wiliam, 2009). In addition, a selected set of items which match to several levels of progression in a topic could be used as a summative test, either to explore more comprehensively the prior knowledge and understanding of a topic, or in a comprehensive review of what has been achieved at the end of the study of a construct. In both cases, the aim is not to grade the students, but to use the results formatively (as in Black, Harrison, Lee, Marshall, & Wiliam, 2003, pp.53–57).

It is not claimed that the learning progression of a particular individual student will follow the general sequence established in the types of sequence shown in Figures 5 and 6. The results presented here may be seen as the combined effect of many different individual progressions of the type shown by Denvir and Brown (1987). However, in planning whole-class activities, a teacher needs to make use of the overall combined effects, while being as responsive as possible to the difficulties of some individual students.

The present study examined both the students' conceptual understanding and their ability to engage in argumentation about the concepts, by asking students, for example, to "give reasons for your answer." Thus we have responded to the argument of Tsaparlis and Sevian (2013) that any study should explore progression in both the conceptual and the epistemological aspects.

The approach taken in the work reported here has been built on the basis of an argument about the links between curriculum, pedagogy and assessment. The central aim of the work reported here was to check and, where necessary, modify a learning progression which teachers could use to guide their work, and to develop the assessment methods and tools that would support the use of the learning progression as a guide to developing students' learning.

Current reform efforts in science identify assessment as a critical support for instruction (NGSS Lead States, 2013). The NGSS describes specific goals for science learning in the form of performance expectations (statements about what students should know and be able to do at each grade level). This study provides an example of an empirically based learning progression of student understanding of the structure of matter—an example that addresses the progressive nature of learning by providing a continuum on which students can be placed. This research should inform the professional development of teachers as they implement new reform ideas in their classrooms to teach and assess student learning.

Summary

This study puts forth a progression of how students understand the structure of matter by empirically inter-relating constructs at different levels of sophistication to develop a unified learning progression. We also encountered through the study that student thinking can be more complex than hypothesized from earlier work, as in the case of our discovery of a substructure (or strands) of understanding within a single construct.

The paper was shaped by and contributed to other important papers about how student understanding of the structure of matter develops. Previous studies (Hadenfeldt et al., 2014, 2016; Johnson, 2013; Johnson & Tymms, 2011; Smith et al., 2004) each produced a model to see how student conceptions of matter develop. However, the progressions proposed by previous studies did not align closely or clearly with one another. In formulating the basis of the present study, the Smith et al. (2004) study aligned most closely with our hypothesized model, while it also provided an initial set of items. However, none of the previous studies report data on a large scale which made significant use of students' own expressions of their thinking in comparison to our study in allowing a large number of students to construct their responses via, mainly, the "please explain" questions. Other studies relied on forced choice items for their large data collection efforts. And, none of them reported follow-up to explore the application of the results, either by feedback to teachers or by further empirical studies of their use in pedagogy. The present study relied on teacher input throughout each iteration of the study.

The paper discusses cautionary considerations drawn from the larger landscape of learning progressions like Duschl's et al. (2011) top-down and bottom-up approaches and Shavelson & Kurpuis' (2012) warning about the importance of the interaction between instruction and students' prior knowledge. Researchers involved in the current study worked extensively with experienced teachers as they actively taught about the structure of matter and collaborated in the study.

This study is a significant resource to guide further work of researchers and educators. For researchers interested in empirically testing learning progressions, this study demonstrates that a construct modeling approach for investigating a progression through a developmentally based iterative process shows great promise. In particular, it has demonstrated the need for a thorough consideration of the core constructs and of their possible sub-structures which we have called strands. For educators, we recommend, on the basis of our results and study methods, that the choice of instructional approach needs to be fashioned in terms of a model, grounded in evidence, of the paths through which learning might best proceed, working toward the desired targets by a pedagogy which also fosters students' development as effective learners.

Research should explore further the nature of the strands found in this study. Previous research has not produced such sub-structures. Future research should investigate what this division implies for instructional planning and how educators should think about any relevant strands and about the pedagogy required to address them.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A100692 to University of California, Berkeley. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Notes

[1] Most of the empirical data for the study presented here were collected in 2013 and 2014, so we were not able to use the Hadenfeldt et al. (2016) paper.

[2] Accounts of the results of investigations (b) and (c) have been communicated elsewhere (Osborne et al., 2016; Henderson, Osborne, Macpherson, & Szu, 2013; Osborne, Henderson, MacPherson, & Szu, 2013; Yao, 2013 and Yao, Wilson, Henderson, & Osborne, 2015).

[3] One can also constrain the item difficulties for each dimension to 0.00, as an alternative. This constraint is activated by setting one item difficulty, in each dimension, as the negative sum of all other item difficulties in that dimension.

[4] The advisory board included Alicia Alonzo, Paul Black, Douglas Clark, Richard Duschl, Joseph Krajcik, James Pellegrino, and Helen Quinn.

[5] The p-value was adjusted because it is at the boundary of the parameter space (Rabe-Hesketh & Skrondal, 2008, pp. 69).

References

Abraham, M. R., Grzybowski, E. B., Renner, J. W., & Marek, E. A. (1992). Understandings and misunderstandings of eighth graders of five chemistry concepts found in textbooks. Journal of Research in Science Teaching, 29(2), 105–120.

Adams, R. J., Wu, M., & Wilson, M. (2015). ConQuest 4.0 [computer program]. Hawthorn, Australia: ACER.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723.

Akaygun, S., & Jones, L. I. (2013). Dynamic visualizations: Tools for understanding the particulate nature of matter. In G. Tsaparlis, & H. Sevian (Eds.), Concepts of matter in science education (pp. 281–300). New York: Springer.

Andersson, B. (1990). Pupils' conceptions of matter and its transformations (age 12-16). Studies in Science Education, 18(1), 53–85.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21(1), 5–31.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). Assessment for learning—putting it into practice. Buckingham: Open University Press.

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(2–3), 71–123.

Denvir, B., & Brown, M. (1987). The feasibility of class administered diagnostic assessments in primary mathematics. Educational Research, 29(2), 95–107.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. Journal of Research in Science Teaching, 46(6), 716–730.

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. Studies in Science Education, 47(2), 123–182.

Foster, J, & Wiser, M (2012). The potential of learning progression research to inform the design of state science standards. In A. C. Alonzo, & A. W. Gotwals (Eds.), Learning progressions in science (pp. 435–460). Boston: Sense Publishers.

Furtak, E. M., Thompson, J., Braaten, M., & Windschitl, M (2012). Learning progressions to support ambitious teaching practices. In A. C. Alonzo, & A. W. Gotwals (Eds.), Learning progressions in science (pp. 405–434). Boston: Sense Publishers.

Gomez, E. J., Benarroch, A., & Marin, N. (2006). Evaluation of the degree of coherence found in students' conceptions concerning the particulate nature of matter. Journal of Research in Science Teaching, 43(6), 577–598.

Hadenfeldt, J. C., Liu, X., & Neumann, K. (2014). Framing students' progression in understanding matter: a review of previous research. Studies in Science Education, 50(2), 181–208.

Hadenfeldt, J. C., Neumann, K., Bernholt, S., & Liu, X. (2016). Students' progression in understanding the matter concept. Journal of Research in Science Teaching, 53(5), 683–708.

Henderson, B., Osborne, J. F., Macpherson, A., & Szu, E. (September 4, 2013). Validating a Learning Progression for Student Argumentation in Science. Paper presented at the European Science Education Research Association, Cyprus.

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

Johnson, P. (2013). How students' understanding of particle theory develops: A learning progression. In G. Tsaparlis, & H. Sevian (Eds.), Concepts of matter in science education (pp. 47–67). New York: Springer.

Johnson, P., & Tymms, P. (2011). The emergence of a learning progression in middle school chemistry. Journal of Research in Science Teaching, 48, 849–877.

Krajcik, J. S. (2012). The importance, cautions and future of learning progression research. In A. C. Alonzo, & A. W. Gotwals (Eds.), Learning Progressions in Science (pp. 27–36). Boston: Sense Publishers.

Masters, G. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149–174.

Masters, G. N., & Forster, M. (1996). Progress maps: Assessment resource kit (ARK). Camberwell, VIC: Australian Council for Educational Research.

Masters, G. N., Adams, R. A., & Wilson, M (1990). Charting student progress. In T. Husen, & T. N. Postlewaite (Eds.), International encyclopedia of education: Research and studies (Supplementary Volume 2, pp. 628–634). Oxford: Pergamon Press.

Merritt, J., & Krajcik, J. (2013). Learning progression developed to support students building a particle model of matter. In G. Tsaparlis, & H. Sevian (Eds.), Concepts of matter in science education (pp. 11–45). New York: Springer.

National Research Council (NRC). (2014). Developing assessments for the Next Generation Science Standards. Washington, D.C.: National Academies Press.

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

National Research Council (NRC). (2005). Systems for State Science Assessment. Washington, DC: The National Academies Press.

National Research Council (NRC). (2007). Taking Science to School: Learning and Teaching Science in Grades K-8. Washington, DC: The National Academies Press.

National Research Council (NRC). (2012). A framework for K-12 science education practices, crosscutting concepts, and core ideas. Washington, D.C.: National Academies Press.

Osborne, J., Henderson, B., MacPherson, A., & Szu, E. (April 25, 2013). Developing Assessment for a Learning Progression in Argumentation: Lessons Learned NARST symposium. In Assessing Scientific Argumentation: Challenges and Future Directions. National Association for Research in Science Teaching, Rio Grande, Puerto Rico.

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S.-Y. (2016). The development and validation of a learning progression for argumentation in science. Journal of Research in Science Teaching, 53, 821–846. https://doi.org/10.1002/tea.21316

Rabe-Hesketh, S., & Skrondal, A. (2008). Multilevel and longitudinal modeling using Stata. College Station, Texas: STATA press.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461–464.

Schwartz, R., & Ayers, E. (2011). Delta dimensional alignment: Comparing performances across dimensions of the Learning Progression for Assessing Data Modeling and Statistical Reasoning. Unpublished manuscript, University of California: Berkeley.

Shavelson, R. J., & Kurpius, A. (2012). Reflections on learning progressions. In A. C. Alonzo, & A. W. Gotwals (Eds.), Learning progressions in science (pp. 13–26). Rotterdam: Sense Publishers.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic molecular theory. Measurement: Interdisciplinary Research and Perspectives, 14, 1–98.

Smith, C. L., Wiser, M., Anderson, C. W., Krajcik, J., & Coppola, B. (2004). Implications of research on children's learning for assessment: Matter and atomic molecular theory. Paper commissioned by the committee on test design for K-12 science achievement. Washington, D.C: National Research Council.

Stevens, S. Y., Shin, N., & Krajcik, J. (2009). Towards a model for the development of an empirically tested learning progression. Paper presented at the Learning Progression in Science (LeaPS) Conference. Iowa City, IA.

Talanquer, V. (2009). On cognitive constraints and learning progressions: The case of "structure of matter". International Journal of Science Education, 31(15), 2123–2136.

Tsaparlis, G., & Sevian, H. (2013). Towards a scientifically sound understanding of concepts of matter. In G. Tsaparlis, & H Sevian (Eds.), Concepts of matter in science education (pp. 485–520). Dordrecht: Springer.

Wang, W. C., Wilson, M., & Adams, R. J. (1997). Rasch models for multidimensionality between items and within items. Objective Measurement: Theory into Practice, 4, 139–155.

Wilson, M. (2005). Constructing measures: An item response modeling approach. Mahwah, NJ: Erlbaum.

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

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. Applied Measurement in Education, 13(2), 181–208.

Wilson, M., Black, P., & Morell, L. (2013). A Learning Progression Approach to Understanding Students' Conceptions of the Structure of Matter. AERA symposium presentation. In J. Krajcik (Chair), Building Learning Progressions for Science and Math Learning. Symposium conducted at the meeting of the American Education Research Association, San Francisco, CA.

Wilson, M., Osborne, J., Morell, L., Donovan, S., & Daro, P. (2013). Learning Trajectories and Formative Assessments for Physical Science and Argumentation. Presented at the SERP Institute Sandpit – Presidio Trust, San Francisco, CA.

Wiser, M., & Smith, C. L (2008). Learning and teaching about matter in grades K-8: When should the atomic-molecular theory be introduced? In S. Vosniadou (Ed.), International handbook of research on conceptual change (pp. 205–239). Hillsdale: Erlbaum.

Yao, S.-Y. (2013). Investigating the validity of a scientific argumentation assessment using psychometric methods (Unpublished doctoral dissertation). University of California-Berkeley, Berkeley, CA.

Yao, S.-Y., Wilson, M., & Black, P. (April 25, 2013). Developing and Validating a Learning Progression-Based Assessment of the Structure of Matter. AERA symposium presentation. Symposium conducted at the meeting of the American Education Research Association, San Francisco, CA.

Yao, S.-Y., Wilson, M., Henderson, J. B., & Osborne, J. (2015). Investigating the function of content and argumentation items in a science test: A multidimensional approach. Journal of Applied Measurement, 16(2), 171–192.

Supporting Information

Additional Supporting Information may be found in the online version of this article.