

THE STRUCTURED CONSTRUCTS MODEL (SCM): A FAMILY OF STATISTICAL MODELS RELATED TO LEARNING PROGRESSIONS

This paper describes a family of statistical models fashioned to embody some of the underlying conceptualizations that have gone into the work of the BEAR Center in the development of learning progressions. The core of all of these developments has been the construct map, which is the first building block in the BEAR Assessment System (BAS). After introducing the concept of a learning progression, the paper summarizes the elements of the BAS, emphasizing the central concept of a construct map. The paper then focuses on one of the more complex ways to see the relationship between a set of construct maps and a learning progression (see Wilson (in press) for the full set). Here the paper uses an example based on the Atomic-Molecular Model in middle school science. This provides the context for the development of the Structured Constructs Model. A simple example of an SCM is given in some detail. The paper then discusses some strengths and limitations of this conceptualization, and suggests further elaborations of the ideas.

Mark Wilson, University of California, Berkeley

INTRODUCTION

The concept of a learning progression is one that is still developing at the current time. However, it is really just the latest step in the growth of a much older set of ideas—consistencies in the development of students as they grow in cognitive sophistication. Developing assessment to measure a person's growth in a learning progression is crucial. In this paper, a family of statistical models is described which is seen as fundamentally important to the delineation of the measurement basis for a learning progression. We contextualize this development within the scope of one particular approach to measurement, called the BEAR Assessment System (BAS; Wilson, 2005; Wilson & Sloane, 2000), is used as a lens through which to portray a perspective on the many possible ways that learning progressions could be conceived of and measured. In this paper, the manner in which the measurement approach supports the learning progression is referred to as the *assessment structure* for the learning progression. Of course, there are other measurement approaches that one could take—but the specifics of these are outside the scope of this current effort, although the particular family of models described could also be used in other approaches.

The paper begins with a brief introduction to the idea of a learning progression, and adds some notes about assessment perspectives on learning progressions. It then summarizes the elements of the BAS, emphasizing the central concept of a construct map, which is the focus of the rest of the paper. The paper then focuses on one of the more complex ways to see the relationship between a set of construct maps and a learning progression (see Wilson (in press) for the full set). Here the paper uses an example based on the Atomic-Molecular Model in middle school science. This then provides the context for the development of the Structured Constructs Model (SCM): A simple example of an SCM is given in some detail. Finally, the paper discusses some strengths and limitations of this conceptualization, and suggests further elaborations of the ideas.

Learning progressions: Links to assessment

At a recent meeting of researchers working on the topic of learning progressions, the following broad description of learning progressions was suggested by a group consensus:

Learning progressions are descriptions of the successively more sophisticated ways of thinking about an important domain of knowledge and practice that can follow one another as children learn about and investigate a topic over a broad span of time. They are crucially dependent on instructional practices if they are to occur. (CCII, 2009)

The description is deliberately encompassing, allowing a wide possibility of usage, but, at the same time, it is intended to reserve the term to mean something more than just an ordered set of ideas or curriculum pieces. As well, the group saw it as a *requirement* that the learning progression should indeed describe the “progress” through a series of levels of sophistication in the student’s thinking.

Although the idea of a learning progression has links to many older and venerable ideas in education, the history of the specific term “learning progression” in the context of science education is a relatively brief one (CCII, 2009), starting with the publication of an NRC report (NRC, 2006). That report was focused on assessment in K-12 education, and hence the connections to assessment have been there right from the start. Nevertheless, given the brief time-span since then, there is not a great deal of extant literature regarding the relationship between the two, although this may well change in the near future. A second NRC report (NRC, 2007) also featured the concept, and enlarged upon classroom applications. Several assessment initiatives and perspectives are discussed in these reports, including references to the seminal 2001 NRC report *Knowing What Students Know*. Among the assessment programs highlighted there, probably the most prominent is the work on *progress variables* by the Australian researcher Geoff Masters and his colleagues (e.g., Masters, Adams & Wilson, 1990; Masters & Forster, 1996), and the closely-related work on the somewhat more elaborated BEAR Assessment System (Wilson, 2005; Wilson & Sloane, 2000). In this paper I will draw on the latter as the core set of assessment perspectives and practices to relate to learning progressions.

The BEAR Assessment System (BAS)

The BEAR Assessment System is based on the idea that good assessment addresses the need for sound measurement through four principles: (1) a developmental perspective, (2) a match between instruction and assessment, (3) the generating of quality evidence, and (4) management by instructors to allow appropriate feedback, feed forward and follow-up. These four principles, plus four building blocks that embody them are shown in Figure 1. Below we take up each of these principles and building blocks in turn. See Wilson (2005) for a detailed account of an instrument development process that works through these steps.

=====
Insert Figure 1 about here
=====

Principle 1: A Developmental Perspective

A "developmental perspective" regarding student learning means assessing the development of student understanding of particular concepts and skills over time, as opposed to, for instance, making a single measurement at some final or supposedly significant time point. Establishing appropriate criteria for taking a developmental perspective has been a challenge educators for many years. What to assess and how to assess it, whether to focus on generalized learning goals or domain-specific knowledge, and the implications of a variety of teaching and learning theories all impact what approaches might best inform developmental assessment. Taxonomies such as Bloom's Taxonomy of Educational Objectives (Bloom, 1956), Haladyna's Cognitive Operations Dimensions (Haladyna, 1994) and the Structure of the Observed Learning Outcome (SOLO) Taxonomy (Biggs & Collis, 1982) are among many attempts to concretely identify generalizable frameworks. One issue is that as learning situations vary, and their goals and philosophical underpinnings take different forms, a "one-size-fits-all" development assessment approach rarely satisfies educational needs. Much of the strength of the BEAR Assessment System comes in providing tools to model *many different kinds of learning theories and learning domains*. What is to be measured and how it is to be valued in each BEAR assessment application is drawn from the expertise and learning theories of the teachers, the curriculum developers, and the assessment developers involved in the process of creating the assessments.

Building Block 1: Construct Maps. Construct maps (Wilson, 2005) embody this first of the four principles: that of a developmental perspective on assessment of student achievement and growth. A construct map is a well thought out and researched ordering of qualitatively different levels of performance focusing on one characteristic. Thus, a construct map defines what is to be measured or assessed in terms general enough to be interpretable within a curriculum and potentially across curricula, but specific enough to guide the development of the other components. When instructional practices are linked to the construct map, then the construct map also indicates the aims of the teaching. Construct maps are one model of how assessments can be integrated with instruction and accountability. They provide a way for large scale assessments to be linked in a principled way to what students are learning in classrooms, while at least having the potential to remain independent of the content of a specific curriculum.

This approach assumes that, within a given curriculum, student performance on curricular variables can be traced over the course of the curriculum, facilitating a more developmental perspective on student learning. Assessing the growth of students' understanding of particular concepts and skills requires a model of how student learning develops over a certain period of (instructional) time. A growth perspective helps one to move away from "one shot" testing situations, and away from cross sectional approaches to defining student performance, toward an approach that focuses on the process of learning and on an individual's progress through that process. Clear definitions of what students are expected to learn, and a theoretical framework of how that learning is expected to unfold as the student progresses through the instructional material (i.e., a in terms of learning performances), are necessary to establish the construct validity of an assessment system.

The idea of using construct maps as the basis for assessments offers the possibility of gaining significant *efficiency* in assessment: Although each new curriculum prides itself on bringing something new to the subject matter, in truth, most curricula are composed of a common stock of content. And, as the influence of national and state standards increases, this will become more true, and also easier to codify. Thus, we might expect innovative curricula to have one, or

perhaps even two variables that do not overlap with typical curricula, but the remainder will form a fairly stable set of variables that will be common across many curricula.

Construct maps are derived in part from research into the underlying cognitive structure of the domain and in part from professional judgments about what constitutes higher and lower levels of performance or competence, but are also informed by empirical research into how students respond to instruction or perform in practice (NRC, 2001). To more clearly understand what a progress variable is, consider the following example.

The example explored in this brief introduction is a test of science knowledge, focusing in particular on earth science knowledge in the area of “Earth and the Solar System” (ESS). The items in this test are distinctive, as they are Ordered Multiple Choice (OMC) items, which attempt to make use of the cognitive differences built into the options to make for more valid and reliable measurement (Briggs, Alonzo, Schwab & Wilson, 2006). The standards and benchmarks for “Earth in the Solar System” appear in Appendix A of the Briggs et al article (2006). According to these standards and the underlying research literature, by the 8th grade, students are expected to understand three different phenomena within the ESS domain: (1) the day/night cycle, (2) the phases of the Moon, and (3) the seasons -- in terms of the motion of objects in the Solar System. A complete scientific understanding of these three phenomena is the top level of our construct map. In order to define the lower levels of our construct map, the literature on student misconceptions with respect to ESS was reviewed by Briggs and his colleagues. Documented explanations of student misconceptions with respect to the day/night cycle, the phases of the Moon, and the seasons are displayed in Appendix A of the Briggs et al (2006) article.

The goal was to create a single continuum that could be used to describe typical students’ understanding of three phenomena within the ESS domain. In contrast, much of the existing literature documents students’ understandings about a particular ESS phenomena without connecting each understanding to their understandings about other related ESS phenomena. By examining student conceptions across the three phenomena and building on the progressions described by Vosniadou & Brewer (1994) and Baxter (1995), Briggs and his colleagues initially established a general outline of the construct map for student understanding of ESS. This general description helped them impose at least a partial order on the variety of student ideas represented in the literature. However, the levels were not fully defined until typical student thinking at each level could be specified. This typical student understanding is represented in the ESS construct map shown in Figure 2 (a) by general descriptions of what the student understands, and (b) by limitations to that thinking in the form of misconceptions, labeled as “common errors.” Common errors used to define level 1 include explanations for day/night and the phases of the Moon involving something covering the Sun or Moon, respectively.

=====
Insert Figure 2 about here
=====

In addition to defining student understanding at each level of the continuum, the notion of common errors helps to clarify the difference between levels. Misconceptions, represented as common errors in one level, are resolved in the next level of the construct map. For example, students at level 3 think that it gets dark at night because the Earth goes around the Sun once a

day—a common error for level 3—while students at level 4 no longer believe that the Earth orbits the Sun daily but rather understand that this occurs on an annual basis.

The top level of the ESS construct map represents the understanding expected of 8th graders in national standards documents. Because students' understanding of ESS develops throughout their schooling, it was important that the same continuum be used to describe the understandings of both 5th and 8th grade students. However, the top level is not expected of 5th graders; equally, we do not expect many 8th grade students to fall into the lowest levels of the continuum.

The main thrust of this paper is to postulate a family of statistical models that can capture the relationship (or, rather, potential relationships) between construct maps and the learning progression. One concern I have is that the reader may think that this is just a matter of abstract niceties, or quibbling about obscure jargon. In order to help the reader see that the concept of a construct map (or its equivalent) is central to the successful development of assessments, I include below a description of the other parts of the BAS. If the reader is already familiar with BAS, then they should go straight on to the section headed "Mapping out a learning progression using construct maps," unless they are in need of a refresher on the BAS.

Principle 2: Match between Instruction and Assessment

The main motivation for the progress variables so far developed is that they serve as a framework for the assessments and a method of making measurement possible. However, this second principle makes clear that the framework for the assessments and the framework for the curriculum and instruction must be one and the same.

Building Block 2: The items design. The items design governs the match between classroom instruction and the various types of assessment. The critical element to ensure this in the BEAR assessment system is that each assessment task and typical student responses are matched to certain levels within at least one construct map.

Returning to the ESS example, the OMC items were written as a function of the underlying construct map, which is central to both the design and interpretation of the OMC items. Item prompts were determined by both the domain as defined in the construct map and canonical questions (i.e., those which are cited in standards documents and commonly used in research and assessment contexts). The ESS construct map focuses on students' understanding of the motion of objects in the Solar System and explanations for observable phenomena (e.g., the day/night cycle, the phases of the Moon, and the seasons) in terms of this motion. Therefore, the ESS OMC item prompts focused on students' understanding of the motion of objects in the Solar System and the associated observable phenomena. Distractors were written to represent (a) different levels of the construct map, based upon the description of both understandings and common errors expected of a student at a given level and (b) student responses that were observed from an open-ended version of the item. Two sample OMC items, showing the correspondence between response options and levels of the construct map are shown in Figure 3. Each item response option is linked to a specific level of the construct map. Thus, instead of gathering information solely related to student understanding of the specific context described in the question, OMC items allow us to link student answers to the larger ESS domain represented in the construct map. Taken together, a student's responses to a set of OMC items permit an estimate of the student's level of understanding, as well as providing diagnostic information about that specific misconception.

=====
Insert Figure 3 about here
=====

Principle 3: Management by Teachers

For information from the assessment tasks and the BEAR analysis to be useful to instructors and students, it must be couched in terms that are directly related to the instructional goals behind the progress variables. Open-ended tasks, if used, must be quickly, readily, and reliably scorable.

Building Block 3: The outcome space. The outcome space is the set of categorical outcomes into which student performances are categorized for all the items associated with a particular progress variable. In practice, these are presented as scoring guides for student responses to assessment tasks. This is the primary means by which the essential element of teacher professional judgment is implemented in the BEAR Assessment System. These are supplemented by “exemplars:” examples of student work at every scoring level for every task and variable combination, and “blueprints,” which provide the teachers with a layout showing opportune times in the curriculum to assess the students on the different variables.

Principle 4: Evidence of High Quality Assessment

Technical issues of reliability and validity, fairness, consistency, and bias can quickly sink any attempt to measure along a progress variable as described above, or even to develop a reasonable framework that can be supported by evidence. To ensure comparability of results across time and context, procedures are needed to (a) examine the coherence of information gathered using different formats, (b) map student performances onto the progress variables, (c) describe the structural elements of the accountability system—tasks and raters—in terms of the achievement variables, and (d) establish uniform levels of system functioning, in terms of quality control indices such as reliability.

Building Block 4: Wright maps. Wright maps represent this principle of evidence of high quality. Wright maps are graphical and empirical representations of a construct map, showing how it unfolds or evolves in terms of increasingly sophisticated student performances.

Broader Perspectives on the BAS

We typically use a multi-dimensional Rasch modeling approach to calibrate the maps for use in the BEAR Assessment System (see Adams, Wilson, & Wang (1997) and Briggs & Wilson (2001), for the specifics of this model). These maps have at least two advantages over the traditional method of reporting student performance as total scores or percentages: First, it allows teachers to interpret a student’s proficiency in terms of average or typical performance on representative assessment activities; and second, it takes into consideration the relative difficulties of the tasks involved in assessing student proficiency. Later in this paper, we will use a somewhat different approach to modeling in order to integrate hypotheses about links among constructs within a learning progression. This will alter some of the statistical aspects of the modeling, but many of the types of analyses and results will remain similar

In this brief summary, we have demonstrated a way in which large-scale assessments can be more carefully linked to what students are learning. The key here is the use of construct maps to provide a common conceptual framework across curricula. Construct maps developed and used in the ways we have described here can mediate between the level of detail that is present in the content of specific curricula and the necessarily more vague contents of standards documents. These construct maps also create a “conceptual basis” for relating a curriculum to standards documents, to other curricula, and to assessments that are not specifically related to that curriculum.

With the assessments to be used across curricula structured by construct maps, the problem of item development is lessened – ideas and contexts for assessment tasks may be adapted to serve multiple curricula that share construct maps. The cumulative nature of the curricula is expressed through (a) the increasing difficulty of assessments and (b) the increasing sophistication needed to gain higher scores using the assessment scoring guides. Having the same underlying structure makes clear to teachers, policy-makers, and parents what is the ultimate purpose of each instructional activity and each assessment, and also makes easier the diagnostic interpretation of student responses to the assessments.

Mapping out a learning progression using construct maps

This section of the paper concentrates on just the first of the building blocks described above—the construct map—and one potential relationship with the idea of a learning progression, also described above. I have labeled this as the assessment structure. It might seem to have been a waste of time to describe all four of the building blocks when only the first is being used in the rest of the paper, but the concern is that, unless the place of the construct map in the entire BAS approach is understood, its relevance and importance in the following discussion would be misunderstood. At relevant points in the discussion, issues concerning the items, the outcome space and the measurement model will also be mentioned. But the main focus of this article is on the conceptual relationship between the construct map and a learning progression, hence these other matters, although they are of great importance for any actual realization of a construct map, will not be fully explored.

The relationship between the construct maps that make up the learning progression may be quite complex (See Wilson (2008) and Wilson (in press) for examples of other relationships between the construct maps and the learning progression.) For instance, there could be an assumption that certain of the constructs led to one construct rather than another. This could be illustrated as in Figure 5. Here, the attainment of levels of a construct would be seen as being dependent on the attainment of high levels of specific “precursor” constructs. An example of such thinking, this time in the case of the Molecular Theory of Matter for the middle school level under development with Paul Black of King's College, London, is shown in Figure 6 (Wilson & Black, 2007). In this example, each of the boxes can be thought of as a construct map, but the relationship between them is left unspecified in this diagram. In particular, the Density and Measurement and Data Handling constructs are seen as providing important resources to the main series of constructs, which is composed of the other four constructs, Properties of Objects, Properties of Atoms and Molecules, Conservation and Change, and Molecular Theory of Macro Properties.

=====
Insert Figures 5 & 6 about here
=====

A more complicated way of seeing such a possibility is shown in Figure 7, where there are links hypothesized that are between specific levels of one construct, and specific levels of other constructs (rather than the “top to bottom” relationships shown in Figure 8).

=====
Insert Figures 7 & 8 about here
=====

Structured Constructs Models (SCMs)

With respect to the measurement models that one would use to model the data arising from assessments based on the construct map structures described above, a great deal will depend on the nature of those structures, and the hypothesized links among them. Statistically speaking, the most common frameworks are essentially comprised of correlated dimensions, so that a multidimensional item response model (Adams, Wilson & Wang, 1997) would be suitable. However, the approach in Figure 7 would constitute a variant of structural equation models (SEM). In this view, each of the construct maps would be a single SEM variable, and the arrows between would be the usual SEM paths, effectively representing regression relationships between the lower (predictor) variable and the higher (criterion) variable. In contrast the approach where the arrows arise from and point to levels inside the boxes would constitute a more complicated model that goes beyond the usual SEM formulation. In this model, the “SEM paths” run not between the construct maps (i.e., between the SEM variables), but from and to specific levels within those variables.

There are various ways that one might formally model what these diagrams could represent. First, let us pause and make sure that we are clear on what we do indeed intend those arrows to represent. In many circumstances, it can be reasonable to hypothesize that reaching a certain level on one construct map is a necessary (or perhaps a favored) prelude to reaching a certain level on a second construct map, but that is not a continuous relationship, as for SEM—it is only for that pair of levels. For example, in the Atomic-Molecular Model example cited above, there is a connection from the highest level of the Density construct map to the mixtures level of the Molecular Theory of Macro Properties (MTMP) construct map. This connection does not pertain to any other level of the Density construct, nor does it lead to any other level of the MTMP construct.

One way to mathematically represent this would be to see these two levels of the two different construct maps as each being one of a series of ordered latent classes (i.e., the levels) for each construct, where the regression relationship relates to the probability of a person being in the two specific latent classes—i.e., being in one level of the first construct makes it more likely that the person would be in a particular level of the second construct. This is illustrated in Figure 9, where the small ovals represent the successive levels of each construct, and the specific link is

from level j of the first construct to level k of the second (of course, there could be more such links, but I will illustrate with just one).

=====

Insert Figure 9 about here

=====

To write this as an equation, consider each of the constructs as represented by an ordered set of latent classes (corresponding, say, to the levels of the construct map for each): $\{\phi_1, \phi_2, \dots, \phi_R\}$ for the first (predictor) construct, and $\{\eta_1, \eta_2, \dots, \eta_D\}$ for the second (criterion) construct. The individual student n is characterized by an indicator vector for group membership in each construct ϕ_n , and η_n . For the R potential person groups, $\phi_n = (\phi_{n1}, \dots, \phi_{nR})$, where ϕ_{nr} takes the value of 1 if the person n is in group k and 0 if not, and similarly for the second construct. The model assumes that each person belongs to one and only one ϕ group and one η group; thus, only one of the ϕ_{nr} is theoretically nonzero, and similarly for η_{nd} . Consistent with the term “latent”, the values of neither ϕ_n nor η_n are observable.

Then the link between two specific levels, say j and k , between the first and second constructs, respectively, is given by an equation of the form given in Equation 1:

$$\Pr(\eta_{nj}=1) = f_{jk}(\Pr(\phi_{nk}=1)). \tag{1}$$

Note that this probability is the posterior probability. The function f_{jk} may take a variety of forms, the simplest being a linear relationship. I will use F to designate the entire set of possible functions f_{jk} —of course, in this simple case, only one is not null. I call this a "Structured Constructs Model" (SCM). There could be many ways to hypothesize more complex versions of this type of model, but I will stick to this simple example for now.

What is not yet clear is how the observations of items relate to these constructs. Suppose that item responses X_i and Y_h relate to the first and second constructs respectively, where $i=1,2,\dots,I$, and $h=1,2,\dots,H$. Again, there could be more complex situations, where, there are, say, item bundles that relate to both constructs, etc., but, again, I will stick to this simple formulation.

The probability of person n responding at level r of item i related to the first construct is given as:

$$P\left(X_{nir} = x_{nir} \mid \phi_{nk} = 1, \beta_i\right) = \frac{\exp \sum_{s=1}^r (\beta_{is})}{\sum_{t=1}^H \exp \sum_{s=1}^t (\beta_{is})}, \tag{2}$$

where β_i is a vector of latent class parameters, specific to item i , governing the allocation of persons to classes within the first construct ϕ . Similarly, the probability of person n responding at level d of item h for the second construct is given as:

$$P\left(Y_{nhd} = y_{nhd} \mid \eta_{nh} = 1, \delta_h\right) = \frac{\exp \sum_{s=1}^d (\delta_{hs})}{\sum_{t=1}^D \exp \sum_{s=1}^t (\delta_{hs})}, \quad (3)$$

where δ_h is the corresponding set of parameters of item h for the second construct.

Note that the parameters β_{is} have specific interpretations. Assuming still that $\phi_{nk}=1$ (i.e., person n is indeed in level k), then, for $s=k$, β_{ik} governs the probability that a person at level k will indeed give a response to item i that is also in level k . However, for $s=k+t$ ($t>0$), β_{is} governs the probability that a person at level k will give a response to item i that is above level k . Under some circumstances, this might be interpreted as “guessing,” and this is how it is usually labeled in other latent class models (e.g., NIDA (Maris, 1999), DINA (Junker & Sijtsma, 2001), Fusion (Roussos, DiBello, Stout, Hartz, Henson & Templin, 2007), etc.). However, this label is limiting in its meaning, as the reason for the higher response could be that (a) the person did indeed know the higher level response for this particular item (even though they don’t in a general sense), (b) the rating of that response may have been erroneous, (c) the person may indeed have guessed, or even, faked, the response to item i , or (d) some other cause not covered by (a) through (c). Thus, I will call this set of phenomena “over-responding,” rather than guessing. Similarly, for $s=k-t$ ($k>t>0$), β_{is} governs the probability that a person at level k will give a response to item i that is below level k . This is usually called “slipping” in other latent class models. Again, one can consider different possibilities for why it occurs: The reason for the lower response could be that (a) the person did indeed not know the higher level response for this particular item (even though they do in a general sense), (b) the rating of that response may have been erroneous, (c) the person may have made a responding error (e.g., filled in a different bubble than they intended, etc.) for the response to item i , or (d) some other cause not covered by (a) through (c). I will call this set of phenomena “under-responding,” rather than “slipping,” as that corresponds well to “over-responding,” even though “slipping” also is quite appropriate. Of course, the same remarks can be made for the δ_h parameters.

Under some circumstances, it may be that the β_{is} parameters could be constrained in ways that correspond to the circumstances. For example, some researchers might postulate that there is no possibility for persons to respond at a higher level than k (over-responding is absent): Then for $s>k$, one might set $\beta_{is} = 0$, or, $\beta_{is} = c$, where c is a value corresponding to a very small probability. This might then be modified to make this restriction for just the levels that are at least two above k , etc. Another circumstance might occur where the items are of sufficiently similar characteristics that it would make sense to constrain the β_{is} parameters (other than β_{ik}) to be constant across items: i.e., $\beta_{is} = \beta_i$, for $s \neq k$. Under certain circumstances, it might even make sense to constrain the parameters to be symmetric about $s=k$, although that seems unlikely in a cognitive context, where under-responding seems always to be more likely than over-responding. Of course, the same remarks can be made for the δ_h parameters.

Specification of an SCM such as that described above can be expressed under a Generalized Linear and Mixed Model (GLMM) framework. The probability that a person with group membership parameter ϕ_n will respond in category r to item i is given by:

$$P\left(X_{nir} = x_{nir} \mid \underset{\sim}{\phi}_n, \underset{\sim}{\beta}\right) = \prod_g P(X_{nir} = x_{nir} \mid \underset{\sim}{\phi}_{ng} = 1, \underset{\sim}{\beta}_i)^{\phi_{ng}}. \quad (4)$$

Similarly, for the second construct, the probability that a person with group membership parameter $\underset{\sim}{\eta}_n$ will respond in category d to item h is given by:

$$P\left(Y_{nhd} = y_{nhd} \mid \underset{\sim}{\eta}_n, \underset{\sim}{\delta}_h\right) = \prod_{g'} P(Y_{nhd} = y_{nhd} \mid \underset{\sim}{\eta}_{ng'} = 1, \underset{\sim}{\delta}_h)^{\eta_{ng'}}. \quad (5)$$

As item responses are assumed to be independent given $\underset{\sim}{\phi}_n$, $\underset{\sim}{\eta}_n$, the item parameters and the link function parameters, F , the modeled probability of a response vector is:

$$P\left(X_n = x_n, Y_n = y_n \mid \underset{\sim}{\phi}_n, \underset{\sim}{\eta}_n, \underset{\sim}{\beta}, \underset{\sim}{\delta}, F\right) = \prod_g \prod_i \prod_r \left\{ P(X_{nir} = x_{nir} \mid \underset{\sim}{\phi}_{ng} = 1, \underset{\sim}{\beta}_i)^{\phi_{ng}} \right\}^{x_{nij}} \quad (6)$$

$$\times \prod_{g'} \prod_h \prod_d \left\{ P(Y_{nhd} = y_{nhd} \mid \underset{\sim}{\eta}_{ng'} = 1, \underset{\sim}{\delta}_h)^{\eta_{ng'}} \right\}^{y_{nhd}}$$

subject to the constraints of Equation 1. Different estimation approaches are embodied in different programs—for example, programs that will implement this are LatentGOLD (Vermunt & Magidson, 2008), *gllamm* (Rabe-Hesketh, Pickles & Skrondal, 2001), HUGIN (Andersen, Olesen, Jensen, & Jensen, 1989) and WinBugs, with the last providing the greatest flexibility to estimate most complex versions of this family of models.

Note that the results from such an estimation will be an allocation of probabilities for each person to each of the ordered latent classes—i.e., the levels. Even though the fundamental assumption is that each person is in just one level, we still get probabilities—this is not a contradiction, as we are still assuming that the person is in just one level, but the probabilities relate to how much we believe they belong to each. This will allow a visual representation that is analogous to the Wright Map mentioned above, but that does not have the same metric properties as that representation—there is no analogy here of distance along map to probability, as the discrete nature of the latent classes do not allow such an interpretation. Instead, the classes are simply ordered as in Figure 9. Thus, the SCM map will look far more like the construct map itself, so long as there was no evidence in the results that the construct map should be changed (see next paragraph).

These results can be useful not only in identifying people who are clearly at a specific level, but also people whose responses show them to be “misfitting” the fitted model (in a way that is quite parallel to the use of “fit” and “misfit” in item response modeling). Thus, there may be people who are roughly split between too levels, and there may be people who seem to have no strong plurality for any one level. Of course, when there are large numbers of such people who misfit, then we would consider that the misfit information was not just informative about those individuals, but might also be informative about misfit of the specified model. This could lead to alterations in the construct map and/or the SCM, depending upon the specific patterns of misfit.

Discussion & Conclusion

In this paper I have tried to outline some possible underlying statistical models that one could build to undergird the relationship between a set of constructs and a learning progression. This has been done from very specific measurement perspective, that of the construct map that forms the heart of the BEAR Assessment System. I make no excuses about this focus, as is displayed in Wilson (in press), even taking such a particularistic view, there are a great many ways that the construct map concept could be deployed to give structure and form to the assessments to support a learning progression. Other measurement approaches could equally be used, but these would require separate development in separate papers. Laying out these possibilities is helpful to thinking about what the issues and limitations of such an approach might be.

One interesting question is whether the proposed SCM formulation is indeed a step forward from the venerable SEM approach. Indeed, it would be an interesting comparison to compare the results from the two modeling approaches. This could not be done in a technically precise manner with the classical SEM framework, where the modeling was based on the variance-covariance matrix, but it could be carried out in an approximately similar way using a full item response modeling approach as is available in the *gllamm* (Rabe-Hesketh *et al*, 2001), or M-PLUS (Muthén, & Muthén, 1998-2007) software. Under these circumstances, the SEM solution would provide a robust “baseline” model for the SCM model, allowing one to compare the value of the preciseness of the level-to-level hypothesized links with the more relaxed variable-to-variable links in the SEM approach. Equally, an SEM solution could also be seen as a pre-cursor to a more fine-grained SCM solution, allowing for rapid estimations, and (perhaps) pilot work done on smaller samples. Thus the two approaches could have complementary purposes.

The ideas about SCMs that have been presented here can be extended to deal with other complexities of the assessment context. Clearly the simple uni-connection SCM model described can be extended to dealing with multiple connections and more constructs. Integration of the SCM and SEM types of models would allow one to entertain the possibility that some constructs were composed of (ordered) latent classes while, at the same time, some others were composed of continua. A more subtle extension would be to entertain the idea that some constructs could be better-represented as being simultaneously both class-like and continuum-like. One way to interpret that would be to see that the latent classes represented in Figure 9 could, in addition to the current conception, be seen as being arrayed along a continuum, so that the “distance” between them became meaningful in a probabilistic sense—thus going beyond the current conception, where the latent-class representation of the levels means that there is no distinction in the links between each of the levels. Similar possibilities have been considered in related domains, such as in the modeling of stage-like development, where specific models such as the Saltus model (Wilson 1989, Draney & Wilson, 2007) have been developed and applied. The development of the SCM, as shown in Equations 1 through 6, has been made explicitly with this extension in mind.

A further complexity arises if one asks a different question about the relationships among the ideas in a learning progression: What if the links that are being hypothesized in a diagram like Figures 5 and 7 are not relationships among persons, but among items? This possibility lays open a completely different formulation of the modeling situation, where the links shown are hypothesized are seen as being between item characteristics (i.e., between item parameters), rather than between person characteristic (i.e., between person parameters). This can be seen as

delving into the “item-side” of the situation as well as the “person-side” (De Boeck & Wilson, 2004), and opens up possibilities to start distinguishing between those two sides. A class of models has been developed that is designed to encompass this possibility, the random item profile (RIP) models (De Boeck, 2008), and simple examples have already been published, the so-called “MIRID” models (Butter, De Boeck & Verhelst, 1998; Lee & Wilson, 2009)

Although the example SCM that is discussed is, intentionally, a very simple one, one can easily see that there are a great many possible ways that the construct maps could be deployed to support a learning progression. This flexibility is important, as one would not want to have the potential usefulness of a learning progression to be constricted by the underlying conceptualization of construct maps.

It is also clear that there are some important decisions that will need to be made when one is thinking about designing an SCM appropriate to a given assessment structure (i.e., to the constructs and their associated learning progression). Being aware of the range of possibilities described here, and possibilities beyond these (briefly mentioned above), will help the developers of a learning progression in thinking about the form they want their learning progression to take, and, especially, how they will relate it to the assessments they will use. Considering issues such as whether one would prefer an assessment structure that was best thought of as a multidimensional item response model, or as an SCM, something in-between, will be an important step in developing a new learning progression, or in modifying an existing one.

Equally clear, these choices will also have important ramifications for the other building blocks, the items design and the outcome space. Looking to the topic of assessment structures per se, this paper has really just scratched the surface of an important aspect of the application of measurement ideas in science education in particular, and, potentially, across the whole range of areas of educational achievement. Unidimensional and multidimensional item response models have been a mainstay of the measurement in educational achievement domains for the last few decades. Seeing how these can be extended into the complex areas allowed by SEM-like approaches, and the more subtle SCM approaches described above will be an interesting and challenging task in the future.

References

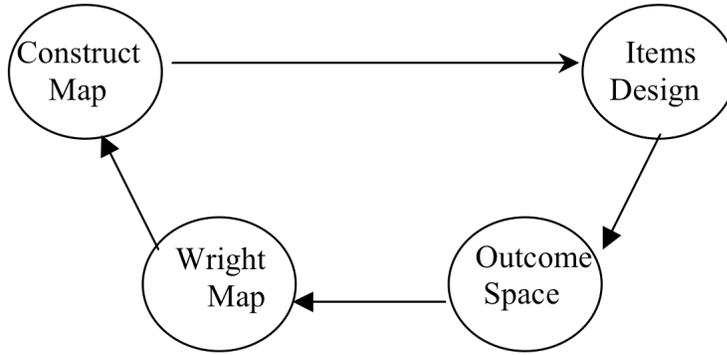
- Adams, R. J., Wilson, M. & Wang, W-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1–23.
- Andersen, S.K., Olesen, K.G., Jensen, F.V., & Jensen, F. (1989). HUGIN - A shell for building bayesian belief universes for expert systems. *International Joint Conference on Artificial Intelligence 1989*, 1080-1085
- Baxter, J. (1995). Children's understanding of astronomy and the earth sciences. In S.M. Glynn & R. Duit (Eds.), *Learning science in the schools: Research reforming practice* (pp. 155-177). Mahwah, NJ: Lawrence Erlbaum Associates.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain*. New York; Toronto: Longmans, Green.
- Briggs, D., Alonzo, A., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, 11(1), 33-63.
- Briggs, D. & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, 4(1), 87-100.
- Butter, R., De Boeck, P., & Verhelst, N. D. (1998). An item response model with internal restrictions on item difficulty. *Psychometrika*, 63, 1-17.
- Center for Continuous Instructional Improvement (CCII). (2009). *Report of the CCII Panel on Learning Progressions in Science*. CPRE Research Report, Columbia University, New York.
- De Boeck, P. (2008). *Random item IRT models*. Presidential address presented at the annual meeting of the Psychometric Society, Durham, NH.
- De Boeck, P. & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer-Verlag.
- Draney, K, & Wilson, M. (2007). Application of the Saltus model to stage-like data: Some applications and current developments. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models* (pp. 119-130). New York: Springer.
- Duncan, R.G., & Hmelo-Silver, C. (in press). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal for Research in Science Teaching*.
- Glaser, R. (1990). *Testing and assessment: O tempora! O mores!* Pittsburgh: Learning Research and Development Center, University of Pittsburgh.
- Haladyna, T. M. (1994). Cognitive taxonomies. In T. M. Haladyna (Ed.), *Developing and validating multiple-choice test items* (pp. 104–110). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hewson, P. W. (1992). *Conceptual change in science teaching and teacher education*. Paper presented at a meeting by the National Center for Educational Research, Documentation and Assessment. Ministry of Education and Science, Madrid, Spain, 1992.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Lee, Y-S., & Wilson, M. (2009). An extension of the MIRID model for polytomous responses and Random Effects. Paper presented at the annual meeting of the American Educational Research Association, San Diego.

- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187-212.
- Masters, G. N., Adams, R. J., & Wilson, M. (1990). Charting of student progress. In T. Husen & T. N. Postlethwaite (Eds.), *International Encyclopedia of Education: Research and Studies. Supplementary Volume 2*. Oxford: Pergamon Press. (pp. 628-634)
- Masters, G., and Forster, M. (1996). *Progress maps. Assessment resource kit*. Victoria, Australia: Commonwealth of Australia.
- Mohan, L., Chen, J., & Anderson, C. W. (2008). *Developing a K-12 learning progression for carbon cycling in socio-ecological systems*. Center for Curriculum Materials in Science Research Report, Michigan State University (Downloaded from <http://edr1.educ.msu.edu/EnvironmentalLit/publicsite/html/carbon.html>)
- Muthén, L.K. and Muthén, B.O. (1998-2007). *Mplus User's Guide*. Fifth Edition. Los Angeles, CA: Muthén & Muthén.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. J. Pellegrino, N. Chudowsky, & R. Glaser (Eds). Washington, DC: National Academy Press.
- National Research Council. (2006). *Systems for state science assessment*. Committee on Test Design for K-12 Science Achievement. M. Wilson, & M. Bertenthal (Eds.). Washington, D.C.: National Academy Press.
- National Research Council. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8*. Committee on Science Learning, Kindergarten through Eighth Grade. R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.). Washington, D.C.: National Academy Press.
- 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(2), 211-227.
- Programme for International Student Assessment. (2005). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Organization for Economic Co-operation and Development.
- Rabe-Hesketh, S., Pickles, A. and Skrondal, A. (2001). *GLLAMM Manual. Technical Report 2001*. Department of Biostatistics and Computing, Institute of Psychiatry, King's College, London.
- Roussos, L.A., DiBello, L.V., Stout, W., Hartz, S.M., Henson, R.A., & Templin, J.L. (2007). The Fusion model skills diagnostic system. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education*. Cambridge University Press, Cambridge, UK.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS User Manual, Version 1.4*. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.
- Vermunt J. K., & Magidson J. (2008). *Latent Gold 4.5 user's guide*. Belmont, MA: Statistical Innovations Inc.
- Vosniadou, S., & Brewer, W.F. (1994). Mental models of the day/night cycle. *Cognitive Science*, 18, 123-183.
- Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, 105(2), 276-289.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Wilson, M. (in press). Measuring progressions: Assessment structures underlying a learning progression. *Journal for Research in Science Teaching*.
- Wilson, M. (2008, March). *Measuring progressions*. In A. C. Alonzo & A. W. Gotwals (Chairs), *Diverse perspectives on the development, assessment, and validation of learning progressions in science*. Symposium conducted at the annual meeting of the American Educational Research Association, New York. Retrieved March 30, 2008, from <http://myweb.uiowa.edu/alonzo/aera2008.html>
- Wilson, M., & Black, P. (2007, March). *The idea of a learning progression as a core for both instruction and assessment*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13(2), 181–208.

Principle 1:
Developmental Perspective

Principle 2:
Match Between Instruction and
Assessment



Principle 4: Evidence of High
Quality

Principle 3:
Management by Teachers

Figure 1. The principles and building blocks of the BEAR Assessment System.

Level	Description
5 8 th grade	<p>Student is able to put the motions of the Earth and Moon into a complete description of motion in the Solar System which explains:</p> <ul style="list-style-type: none"> • the day/night cycle • the phases of the Moon (including the illumination of the Moon by the Sun) • the seasons
4 5 th grade	<p>Student is able to coordinate apparent and actual motion of objects in the sky. Student knows that</p> <ul style="list-style-type: none"> • the Earth is both orbiting the Sun and rotating on its axis • the Earth orbits the Sun once per year • the Earth rotates on its axis once per day, causing the day/night cycle and the appearance that the Sun moves across the sky • the Moon orbits the Earth once every 28 days, producing the phases of the Moon <p>COMMON ERROR: Seasons are caused by the changing distance between the Earth and Sun.</p> <p>COMMON ERROR: The phases of the Moon are caused by a shadow of the planets, the Sun, or the Earth falling on the Moon.</p>
3	<p>Student knows that:</p> <ul style="list-style-type: none"> • the Earth orbits the Sun • the Moon orbits the Earth • the Earth rotates on its axis <p>However, student has not put this knowledge together with an understanding of apparent motion to form explanations and may not recognize that the Earth is both rotating and orbiting simultaneously.</p> <p>COMMON ERROR: It gets dark at night because the Earth goes around the Sun once a day.</p>
2	<p>Student recognizes that:</p> <ul style="list-style-type: none"> • the Sun appears to move across the sky every day • the observable shape of the Moon changes every 28 days <p>Student may believe that the Sun moves around the Earth.</p> <p>COMMON ERROR: All motion in the sky is due to the Earth spinning on its axis.</p> <p>COMMON ERROR: The Sun travels around the Earth.</p> <p>COMMON ERROR: It gets dark at night because the Sun goes around the Earth once a day.</p> <p>COMMON ERROR: The Earth is the center of the universe.</p>
1	<p>Student does not recognize the systematic nature of the appearance of objects in the sky. Students may not recognize that the Earth is spherical.</p> <p>COMMON ERROR: It gets dark at night because something (e.g., clouds, the atmosphere, “darkness”) covers the Sun.</p> <p>COMMON ERROR: The phases of the Moon are caused by clouds covering the Moon.</p> <p>COMMON ERROR: The Sun goes below the Earth at night.</p>
0	No evidence or off-track

Figure 2. Construct Map for Student Understanding of Earth in the Solar System

Item appropriate for fifth graders:

It is most likely colder at night because	
A. the Earth is at the furthest point in its orbit around the Sun.	Level 3
B. the Sun has traveled to the other side of the Earth.	Level 2
C. the Sun is below the Earth and the Moon does not emit as much heat as the Sun.	Level 1
D. the place where it is night on Earth is rotated away from the Sun.	Level 4
© WestEd, 2002	

Item appropriate for eight graders:

Which is the best explanation for why we experience different seasons (winter, summer, etc.) on Earth?	
A. The Earth's orbit around the Sun makes us closer to the Sun in summer and farther away in winter.	Level 4
B. The Earth's orbit around the Sun makes us face the Sun in the summer and away from the Sun in the winter.	Level 3
C. The Earth's tilt causes the Sun to shine more directly in summer than in winter.	Level 5
D. The Earth's tilt makes us closer to the Sun in summer than in winter.	Level 4
© WestEd, 2002	

Figure 3. Sample OMC items based upon Earth in the Solar System Construct map

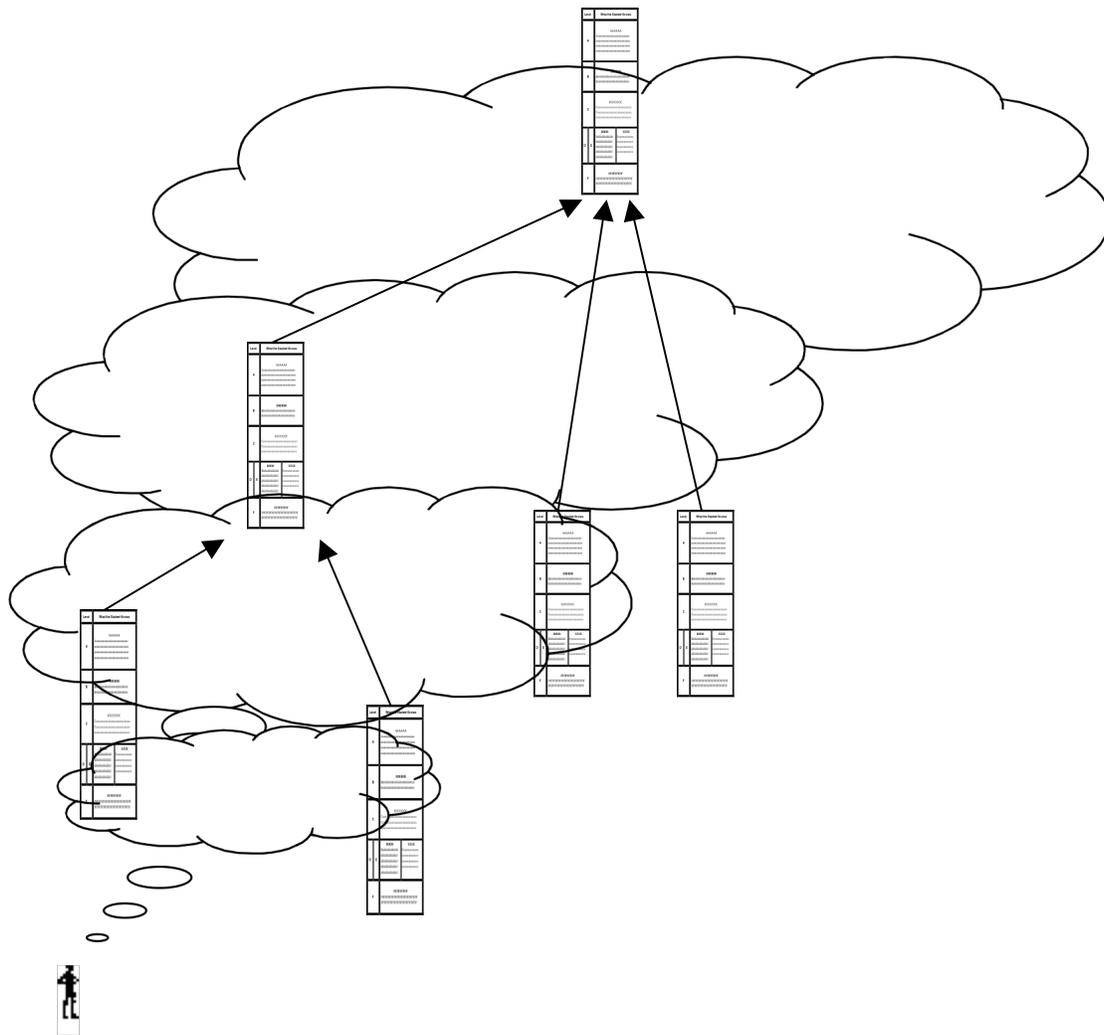
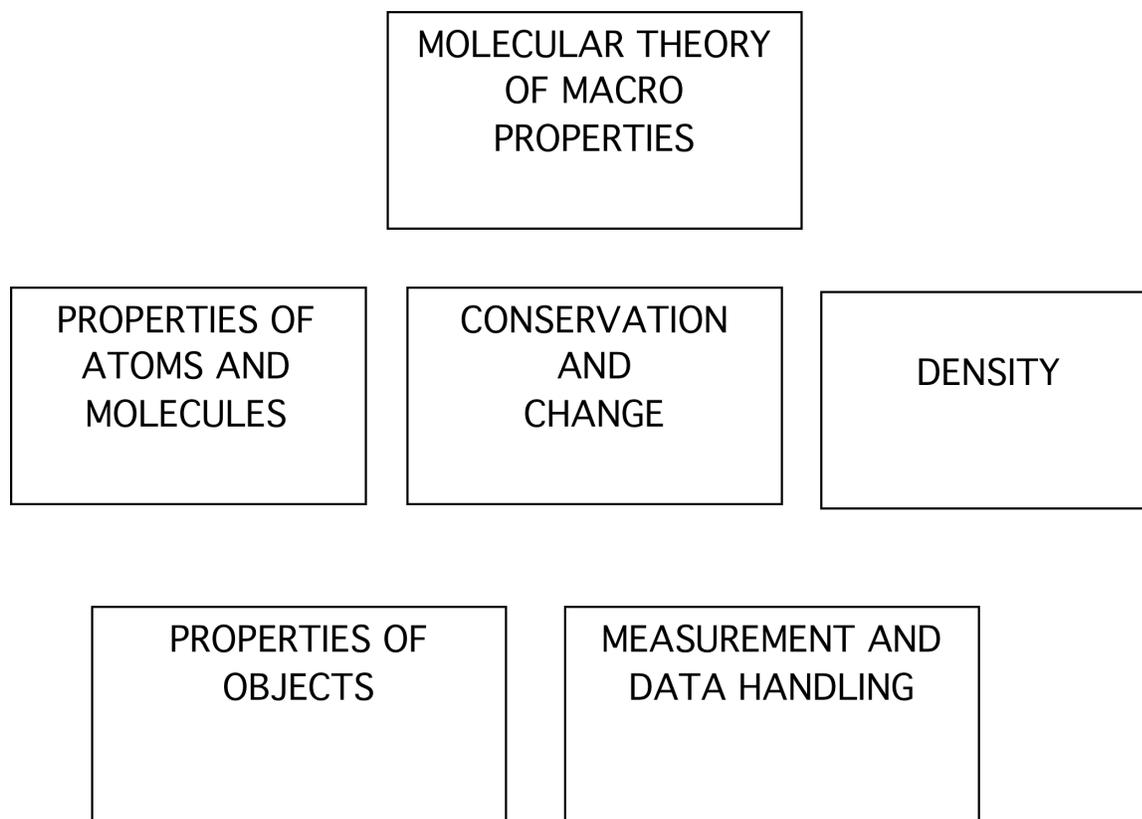


Figure 5. In this situation, there is a complicated dependency relationship between the construct maps in the learning progression.



Molecular Theory of Matter

Figure 6. A set of constructs hypothesized to constitute a Molecular Theory of Matter.

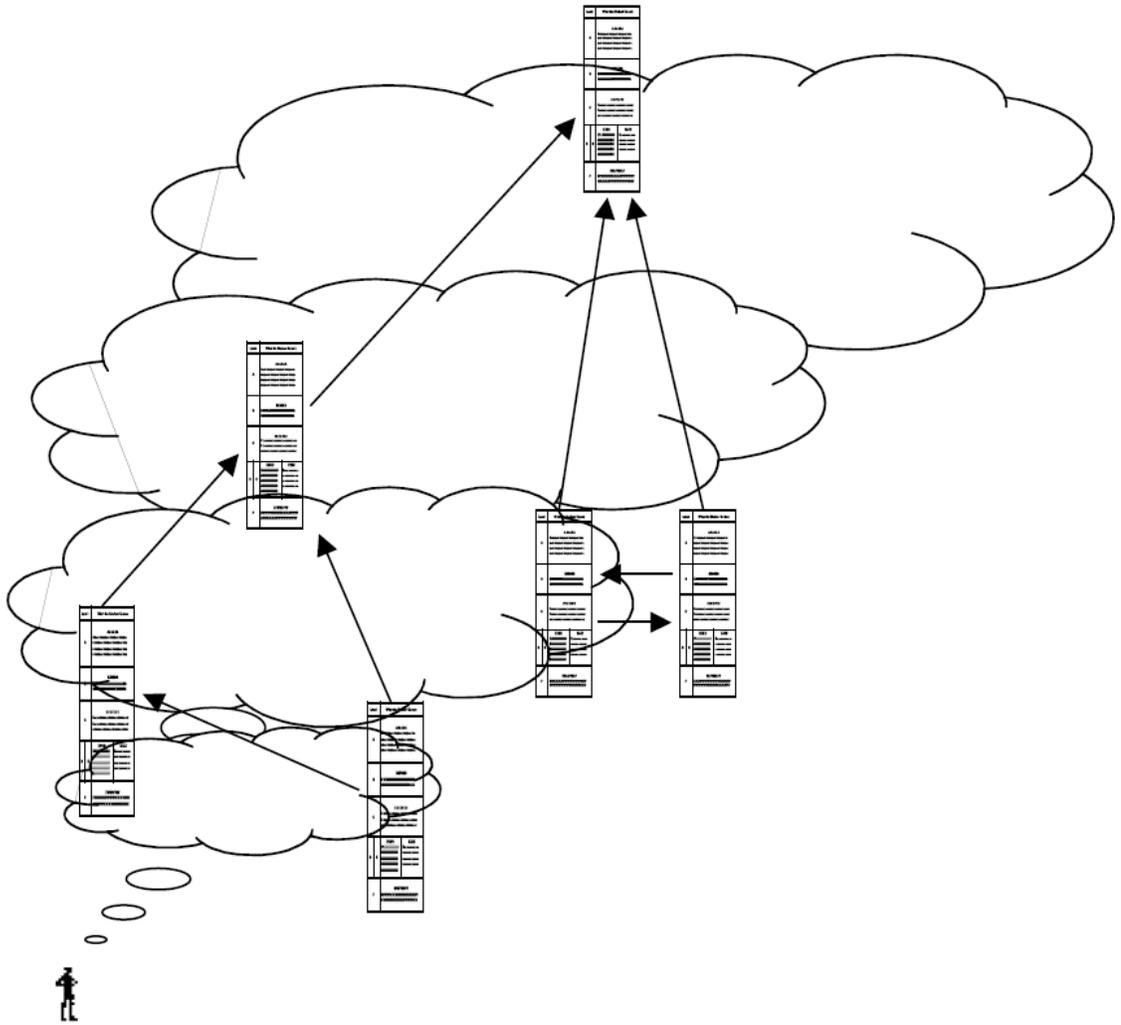


Figure 7. In this situation, the relationship between the construct maps is from level to level.

IN BOXES - PROGRESS VARIABLES
 BOX AS A WHOLE - CONSTRUCT VARIABLES
 BETWEEN BOXES - LINKS
 AS TOOLS (-.-.-) OR BASIS (-----) OR PART OF EVIDENCE (?)

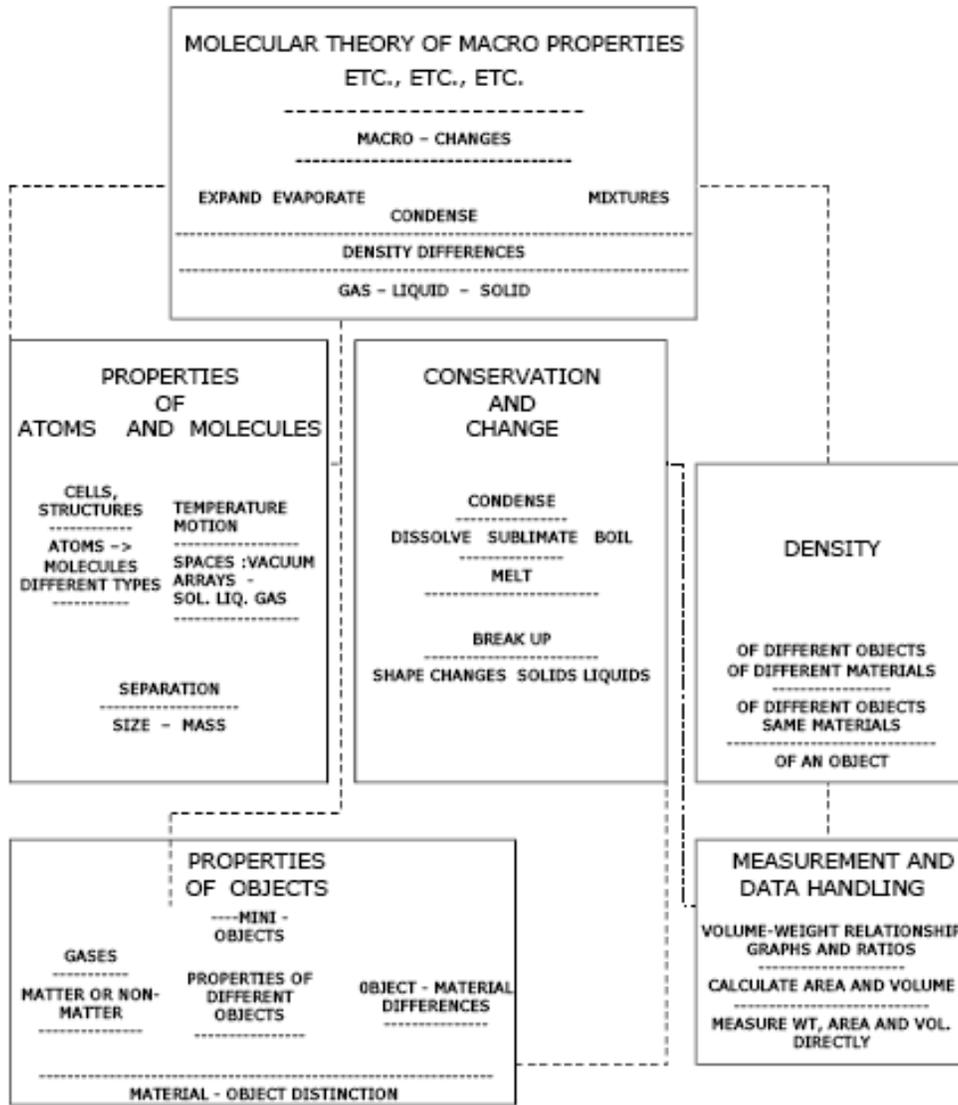


Figure 8. A more detailed version of the relationships shown in Figure 6.

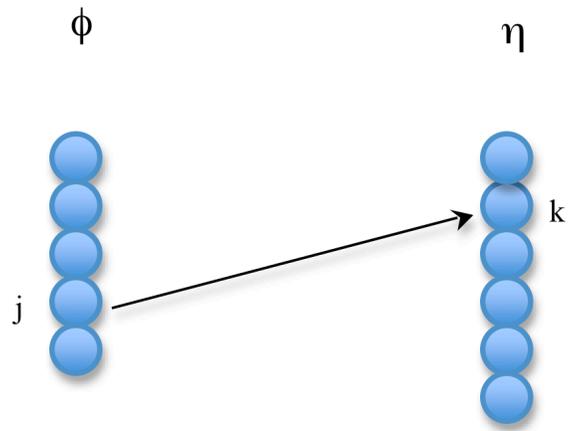


Figure 9. Diagram of a simple SCM.