PEL
Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana
2014, 51(1), 156-174

Pontificia Universidad
Católica de Chile
Facultad de Educación

Considerations for Measuring Learning Progressions where Target Learning is Represented as a Cycle

Consideraciones para la medición de las progresiones de aprendizaje en donde el aprendizaje objetivo se representa como un ciclo

Mark Wilson

University of California, Berkeley, USA

Abstract

This paper explores some of the conceptual complexities associated with developing a learning progression where its final object is represented as a cycle. By way of introduction to this conceptual and theoretical review, the paper first describes some of the conceptualizations underlying the work of the BEAR Center in developing learning progressions. The core of all of these developments is the construct map, the first building block of the BEAR Assessment System (BAS). After introducing the concept of a learning progression, this paper summarizes the elements of the BAS, emphasizing the central role of the construct map. This paper then describes a series of several different ways to interpret the relationship between the idea of a construct map and the idea of a cycle-based progression. Along the way, the paper discusses some strengths and limitations of these conceptualizations, focusing on issues related to both education and measurement. The paper concludes with some general reflections.

Keywords: measurement, assessment, learning progressions, assessment structure

Post to:
Mark Wilson
University of California, Berkeley, USA
Graduate School of Education, University of California, Berkeley
Berkeley, CA 94720, USA
Email: markw@berkeley.edu
This paper is an extension and adaptation of the original article Measuring progressions: Assessment structures underlying a learning progression, published in the Journal of Research in Science Teaching, Special Issue: Learning Progressions, 46(6), 716-730, August 2009.

© 2014 PEL, http://www.pensamientoeducativo.org - http://www.pel.cl

ISSN: 0719-0409 DDI: 203.262, Santiago, Chile
doi:10.7764/PEL.51.1.2014.12

Resumen

En este informe se exploran algunas de las complejidades conceptuales que guardan relación con el desarrollo de una progresión de aprendizaje en donde su objetivo final se representa como un ciclo. A modo de introducción a esta revisión conceptual y teórica, primero se describen algunos de los conceptos que subyacen al trabajo del Centro BEAR en pos de desarrollar progresiones de aprendizaje. Al centro de todos estos desarrollos está el mapa conceptual, el pilar principal del sistema de evaluación BEAR (SEB). Después de presentar el concepto de una progresión de aprendizaje, se resumen los elementos del SEB, al tiempo que se recalca la función central del mapa conceptual. Luego, se describen diferentes maneras de interpretar la relación entre la idea de un mapa conceptual y la idea de una progresión orientada a un ciclo. Asimismo, a lo largo del informe se discuten algunas fortalezas y debilidades de estos conceptos, con un enfoque en los temas relacionados con la educación y la medición. Finalmente, se concluye con algunas reflexiones generales.

Palabras claves: medición, evaluación, progresiones de aprendizaje, estructura de evaluación

The idea of a learning progression is currently still up for debate (Alonzo & Gotwals, 2012). However, it is really just the most recent chapter in the evolution of an older tradition regarding how students develop consistently across different areas. This goes back at least as far as Piaget's work (Piaget, 1950, 1971), but also has had a strong impact on Wright's work (e.g., Wright & Masters, 1981; Wright & Stone, 1979), as well as a role in establishing the conceptual foundation of tests such as TIMSS, PISA and PIRLS. For this idea to be successful, however, it will be essential to develop assessments that can measure a person's growth in a learning progression. This paper, a conceptual and theoretical review, addresses one of the most important challenges that learning progressions pose for measurement practices, which is how to measure a progression where target learning is represented as a *cycle*. There are many examples of cycles that are extant in educational settings: some are focused on the knowledge of cycles occurring in nature, such as the carbon cycle, whereas others are cycles of activities in which students engage as they master a skill or technique; this paper focuses on the latter type of cycle. The BEAR assessment system (BAS) is an example of one such cycle, described in some detail in one of the sections below. For now, it suffices to say that it is a cycle of activities carried out by a person developing a measurement (such as a test or attitude scale), and consists of four sequential parts that are each necessary to the success of the effort. Of course, in carrying out that cycle of activities, it is implicit that the measurer who is creating the instrument must also understand the concept of the cycle itself, and also its application to the particular circumstances of that instrument. Concluding the fourth part of the sequence leads back to the first, allowing the cycle to repeat in iterations, perhaps several times in a single development effort.

It must be made clear from the start that the concept of a cycle does not readily map into the more standard conventions of psychometrics, such as classical test theory or item response theory. In the standard situation upon which both of these approaches are based, the construct under measurement is assumed to be a unidimensional structure, and within the achievement domain, it is most often assumed to grow in one particular direction. A cycle does not inherently embody such a structure, and may, in fact, be thought of as antithetical to such an approach. The following paper discusses how the standard item response modelling approach can be adapted to the multidimensional aspect of a cycle, and then, how that multidimensional approach can be modified to incorporate a cycle.

The remainder of this paper introduces the concept of a learning progression, and then describes how the BEAR Assessment System (BAS) (Wilson, 2005; Wilson & Sloane, 2000) could be viewed as appropriate for a simple unidimensional learning progression, as well as its expansion to a multidimensional profile (Adams, Wilson, & Wang, 1997). The next part of the paper describes how the BEAR approach can then be expanded to incorporate the more complex structures that will likely be present in cycle-based learning progressions. The paper concludes with a brief discussion of the ideas presented here, including possible steps for the future and an exploration of major issues that may arise.

In this paper, the manner in which the measurement approach supports the learning progression will be referred to as the *assessment structure* for the learning progression. Of course, there are other measurement approaches that one could take besides the BAS; however, because the specifics of these other approaches are outside the scope of this current effort, the paper focuses only on using the BAS.

Learning progressions

A group consensus of curriculum developers and assessment researchers who work on learning progressions suggested the following broad description of their area of research at a foundational educators meeting:

> Learning progressions are hypothesized descriptions of the successively more sophisticated ways student thinking about an important domain of knowledge or practice develops as children learn about and investigate that domain over an appropriate span of time (Center for Continuous Instructional Improvement, 2009, p. 37).

The description is deliberately encompassing, allowing it to be used in a wide variety of ways. However, the term is meant to mean something more than just an ordered set of ideas, pieces of curriculum or instructional events. In addition, the group felt it was necessary for the definition of learning progressions to include the idea of a student's thinking as it progresses through a series of levels of sophistication, while remaining broad enough to allow for complications, such as non-linear ordering and different orderings for different parts of the learning progression.

Although, as mentioned above, the idea of a learning progression is associated with many other older and venerable ideas about education, the history of the specific term in the context of education is relatively brief (CCII, 2009). It dates back to the publication of an NRC report (National Research Council, 2006) focused on assessment in K-12 education, meaning that the term has been linked to assessment from the very beginning. Nevertheless, in the short time elapsed since the report, there has been a scarcity of extant literature regarding the relationship between these two ideas, although this may well change in the near future. A second NRC report (National Research Council, 2007) also featured the concept and expanded on classroom applications. The term *learning trajectory* is an older term used in mathematics education, with a rather parallel meaning (e.g., Clements, Wilson, & Sarama, 2004). Several assessment initiatives and perspectives are discussed in these reports, including references to the seminal 2001 NRC report *Knowing What Students Know*.

Looking at the above definition of a learning progression, the challenges to measurement methods and practices based on the idea of a set of "successively more sophisticated ways of student thinking" become increasingly clear. First, for simplicity, assume that there is just one such fully-ordered set of these ways of thinking. In this case, the challenge is to relate the underlying construct to the description of these ways of thinking. What is needed is a way to "bootstrap" the resulting measurements back to the qualitative features of the cognitive structure through the observed characteristics of student responses to the tasks, items or other methods used to generate data. The BAS (described in some detail below) provides one approach to this simpler, unidimensional version of a learning progression. However, it may be too simplistic to assume a complete ordering of the ways of thinking. The topic of learning progressions may involve sub-concepts, or sub-dimensions, each of which has its own ordering, where the sub-dimensions may also have complex relationships with one another, such as being in an ordered cycle. The second half of this paper therefore extends the ideas embodied in the BAS to respond to this challenge.

The BEAR Assessment System

The BEAR Assessment System is based on the idea that good assessment addresses the need for sound measurement through four principles: (a) a developmental perspective, (b) a match between instruction and assessment, (c) management by teachers and (d) the gathering of high-quality evidence. These four principles, in addition to four building blocks that embody them, are shown in Figure 1. The next section of this paper addresses each of these principles and building blocks, emphasizing the first. This approach was initially developed by Wilson and Sloane (2000), as an extension of the Rasch modeling approach propounded by Wright (Wright & Masters, 1981; Wright & Stone, 1979). The BAS approach uses the four principles mentioned above and focuses on using Rasch models to implement the last. See Wilson (2005) for a detailed account of an instrument development process that works through these steps, and

Black, Wilson and Yao (2011) for a discussion of how the resulting assessment system can be related to a learning progression. Information about the BAS is also available online at the BEAR Center website: https://bearcenter.berkeley.edu/page/bearassessment-system. The account given below is broadly based on the descriptions in these references.

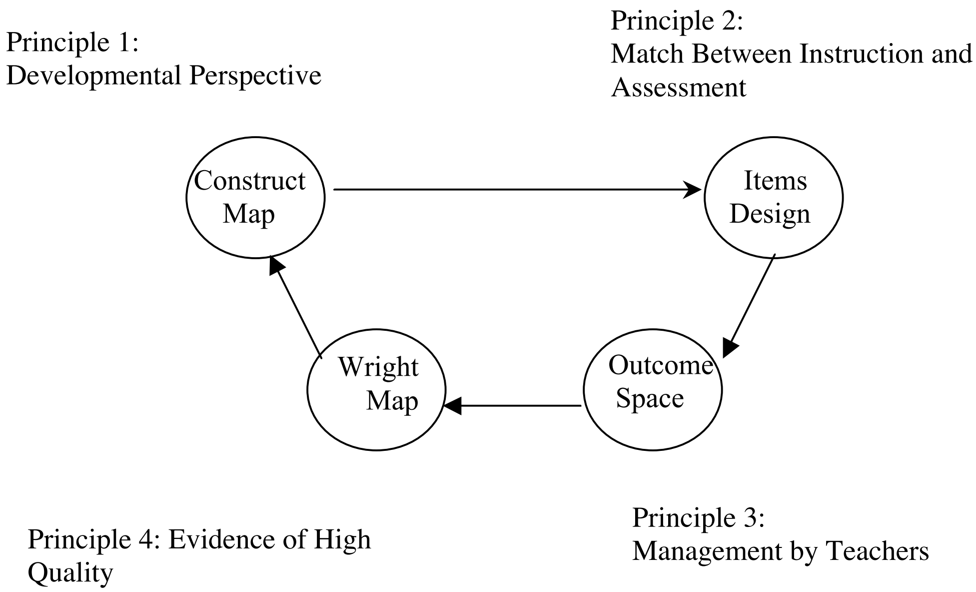


Figure 1. The principles and building blocks of the BEAR Assessment System (BAS).

Principle 1: A developmental perspective

A developmental perspective of student learning involves assessing how student understanding of particular concepts and skills develops over time, as opposed to, for instance, making a single measurement at some final or supposedly significant point in time. Establishing appropriate criteria for implementing a developmental perspective has been a challenge to educators for many years. What to assess and how to assess it, whether to focus on generalized learning goals or domain-specific knowledge, as well as the implications of a variety of teaching and learning theories, all impact what approaches might best inform developmental assessment.

Building block 1: Construct maps. Construct maps (Wilson, 2005) embody the developmental perspective on assessing student achievement and growth, the first of the four principles. 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 that they can be interpreted within a curriculum and potentially across curricula, but specific enough to guide the development of other components. When instructional practices are linked to the construct map, it is also indicative of teaching goals. 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 the very least maintaining the potential to remain independent of the content of a specific curriculum. A clear definition of what students are expected to learn, as well as a theoretical framework of how that learning is expected to unfold as the student progresses through the instructional material (i.e., in terms of learning performance), is necessary to establish the construct validity of an assessment system.

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 (National Research Council, 2001). To more clearly understand what a progress variable is, consider the following example.

The first example explored in this brief introduction is a measure for a single domain from a test of statistics and modeling for middle school students. The Assessing Data Modeling and Statistical Reasoning (ADMSR) project is a collaborative effort between measurement and learning specialists to develop a curricular and embedded assessment system in the areas of statistical reasoning and data modeling (Lehrer, Kim, Ayers, & Wilson, in press). The instruments to measure student abilities in data modeling domains were designed and implemented under the guidance of the Berkeley Evaluation and Assessment Research (BEAR) Center following the framework of the BEAR Assessment System.

The first building block, the construct map, is a description of a latent variable or construct and is an ordering of qualitatively different levels of performance focusing on one characteristic. A construct map is used to represent a cognitive theory of learning consistent with a developmental perspective. Figure 2 shows an example of one of the construct maps from the ADMSR project, the Conceptions of Statistics (CoS) construct map. The CoS construct proposes a series of landmarks as students come to first recognize that statistics measure qualities related to distribution, such as center and spread, and then go on to develop an understanding of statistics as generalizable and subject to sample-to-sample variation.

Conceptions of statistics	
C0S4 – Investigate and anticipates qualities of a sampling distribution.	
CoS**4D**	Predict and justify changes in a sampling distribution based on changes in properties of a sample.
CoS**4C**	Predict that, while the value of a statistic varies from sample-to-sample, its behavior in repeated sampling will be regular and predictable.
CoS**4B**	Recognize that the sample-to-sample variation in a statistic is due to chance.
CoS**4A**	Predict that a statistic's value will change from sample-to-sample.
CoS3 – Consider statistics as measures of qualities of a sample distribution.	
CoS**3F**	Choose/Evaluate statistic by considering qualities of one or more samples.
CoS**3E**	Predict the effect on a statistic of a change in the process generating the sample.
CoS**3D**	Predict how a statistic is affected by changes in its components or otherwise demonstrate knowledge of relations among components.
CoS**3C**	Generalize the use of a statistic beyond its original context of application or invention.
CoS**3B**	Invent a sharable (replicable) measurement process to quantity a quality of the sample.
CoS**3A**	Invent an idiosyncratic measurement process to quantify a quality of the sample based on tacit knowledge that others may not share.
CoS2 – Calculate statistics	
CoS**2B**	Calculate statistics indicating variability.
CoS**2A**	Calculate statistics indicanting central tendency.
CoS1 – Describe qualities of distribution informally.	
CoS**1A**	Use virtual qualities of the data to summarize the distribution.

Figure 2. The Conceptions of Statistics (CoS) Construct Map from the ADMSR Learning Progression.

A construct map assumes that the construct being measured represents a continuum of ability with lower levels of the construct at the bottom and more expert levels at the top end of the map. This continuum of ability is broken up into qualitatively distinguishable reference points along the continuum (i.e., the levels). Within each of the levels, there are sub-levels (which may or may not be ordered depending on the content).

Principle 2: Match between instruction and assessment

The main purpose of the constructs is to 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 must be the same as the framework for the curriculum and instruction.

Building block 2: The items design. The items design building block 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.

In this second building block of the BAS, items are designed to elicit specific kinds of evidence about a respondent's ability with regards to the construct map levels. The goal of a set of items in the BAS is to generate student responses at every level of the construct map. These items may vary by type. For the ADMSR project, the items consisted mostly of short constructed response items, but included some multiple choice items as well. An example of the ADMSR item "Kayla's Project" is shown in Figure 3.

Kayla's project

Kayla completes four projects for her social studies class. Each is worth 20 points.

Kayla's Projects —	Points Earned
Project 1	16 points
Project 2	18 points
Project 3	15 points
Project 4	???

The mean score Kayla received for <u>all four</u> projects was 17.

1. Use this information to find the number of points Kayla received on **Project 4**. Show your work.

Figure 3. The "Kayla's Project" Item from ADMSR.

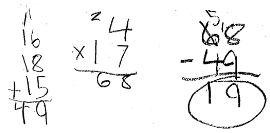
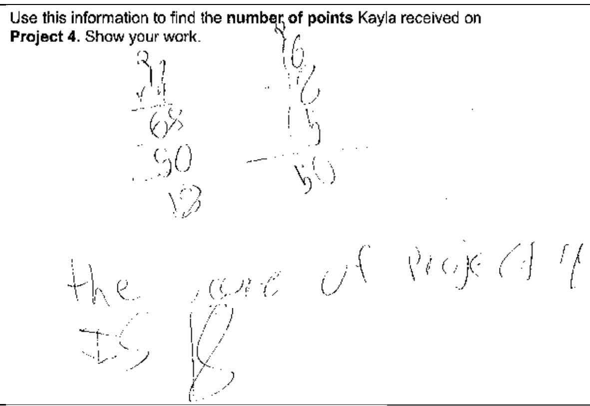
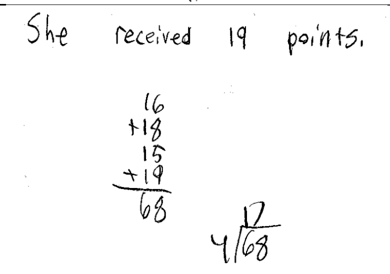
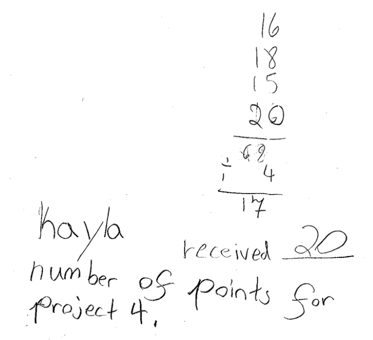
The Kayla's Project item assesses a small part of student understanding of the conceptions of statistics construct. By asking students to calculate a missing value given all other values and a known mean, we are able to assess their understanding of the mean and how it is composed of component values.

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. Teachers already have a full-time job teaching their classes, and no amount of new information will be useful to them if they cannot access it quickly and efficiently. Hence, these open-ended tasks 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 of 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 at which to assess the students on the different variables.

After the items have been administered to the respondents, the results are interpreted using the third building block, the outcome space, which describes in detail how a respondent's answers to items are linked back to the different levels of the construct map. Every item in the ADMSR instruments provides evidence of a respondent's level on one or more of the seven constructs. For the ADMSR project, scoring exemplars were created to explicitly score student responses as a level on a construct map. A set of scoring exemplars for the Kayla's project item is shown in Figure 4.

Kayla's Project Exemplars: Conceptions of Statistics (CoS)		
Levels	**Response Exemplars**	**Example of Student Response**
CoS3D	**Predict how a statistic is affected by changes in its components or otherwise demonstrate knowledge of relations among its components.** Students use strategies correctly that indicate they understand the relations between the individual scores and the mean.	• "The differences between the mean and each score are -1, 1, -2, so the last difference must be 2 and the score must be 19." • N*Mean strategy 
CoS3D-	**Predict how a statistic is affected by changes in its components or otherwise demonstrate knowledge of relations among its components.** Students use a strategy described in CoS3 D but make mistakes in calculations.	
CoS2A	**Calculate statistics indicating central tendency.** Students show knowledge of mean by using a guess and check strategy to find the missing value.	
CoS2A-	**Calculate statistics indicating central tendency.** Students write correct equation to calculate the missing measure. Students provide incorrect answer, but their work shows that they know the procedure.	

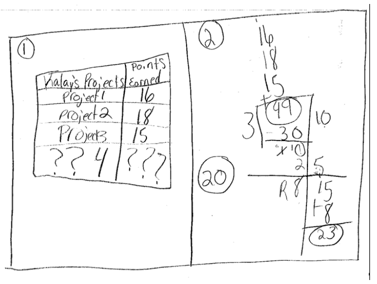
NL(ii)	Students do some idiosyncratic calculation with the measures. Answer is incorrect with no work shown but within a relevant range of answers. Students begin to carry out a strategy correctly but not to completion.	 1.  2.
NL(i)	Attempts item but answers are irrelevant, unclear, implausible, unreasonable, or demonstrate that student did not understand the item.	"I don't know." *
M	Missing response	

Figure 4. The scoring exemplar for Kayla's Project.

The highest performing respondents to the Kayla's project item are scored at level 3 of the CoS construct. At level 3, students are able to employ more flexible strategies to solve this problem. For example, they understand that if the mean of the four scores is 17, the scores must add up to 68. Once past this first step, they are able to find the missing score by subtracting the given values from 68. Students at level 2 on the CoS construct understand how to calculate the mean and use the formula as they normally would when provided a set of values. At this level, students need to use a guess and check strategy in order to solve the problem, but are able to calculate an answer. Students who gave relevant responses, but did not provide evidence of performing at a level on the CoS construct were scored *NL(ii)*, while those who gave irrelevant responses were scored *NL(i)*. These *NL* responses are coded this way to represent responses to the item with no relation to the levels on the CoS construct map. Finally, respondents who saw the item but did not provide a response were scored as missing.

Principle 4: Evidence of high quality

Technical issues related to reliability, validity, fairness, consistency and bias can quickly sink any attempts at measuring using a construct as described above, or even efforts to develop a reasonable framework that can be supported by evidence. To ensure that results can be compared across time and varied contexts, there must be procedures 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 —such as the items and the raters— in terms of the underlying constructs and (d) establish uniform levels of system functioning, in terms of quality control indices such as reliability.

Building block 4: Wright maps. The final building block of the BAS is the measurement model. The measurement model provides a principled way to use the information about respondents and their responses to items coded in the outcome space to locate the respondents and items on the construct map (Wilson, 2003). Different measurement models can be applied to a given instrument. Typically, the results of the measurement model are represented using a *Wright map*, a graphical and empirical representation of a construct map that shows how it unfolds or evolves empirically in terms of increasingly sophisticated student performances.

Mapping out a learning progression using construct maps

The remainder of this paper concentrates on just the first of the building blocks described above —the construct map— and its potential relation to the idea of a learning progression, also described above. I have labeled this as the assessment structure. Because it is necessary to understand how the construct map fits into the overall BAS approach to appreciate its relevance and importance in the subsequent discussion, I have described all four building blocks. When relevant, the discussion will also mention issues related to the items, the outcome space and the measurement model. However, the main focus of this article is the conceptual relationship between the construct map and a learning progression; consequently, although these elements are important to actually produce a construct map, this article will not fully explore them.

One approach to understanding how construct maps are related to learning progressions is to think of the latter as a *set* of the former, each comprising a dimension of the learning progression, where the levels of the construct maps relate (in some way) to the levels of the learning progression (Draney, 2009; Wilson, 2009). A very simple (unidimensional) example of this is shown in Figure 5 —here the learning progression is represented as a sequence of larger and larger clouds of concepts— where their size and relative height are meant to represent the increase in sophistication (and applicability) of the levels of understanding. The person in the bottom left-hand corner is a curriculum developer who has proposed this idea for a learning progression. The single construct map is shown as superimposed on this set of learning progression levels (symbolized by a construct map much like that in Figure 2, reduced to the status of an icon) and allows one to map out the learning progression, similar to the way in which miles mark out the distances along a strip map. In this case, the construct map levels are a fairly good match for the different learning progression structures, but it is not always true that the levels and the Wright map match well. When they do not, the developers must go back to the drawing board, as the problem may derive from any of several issues, including the construct levels, the items, the scoring guide or even a technical issue with the statistical model. Because all options are possible, when something goes wrong, they all must be investigated.

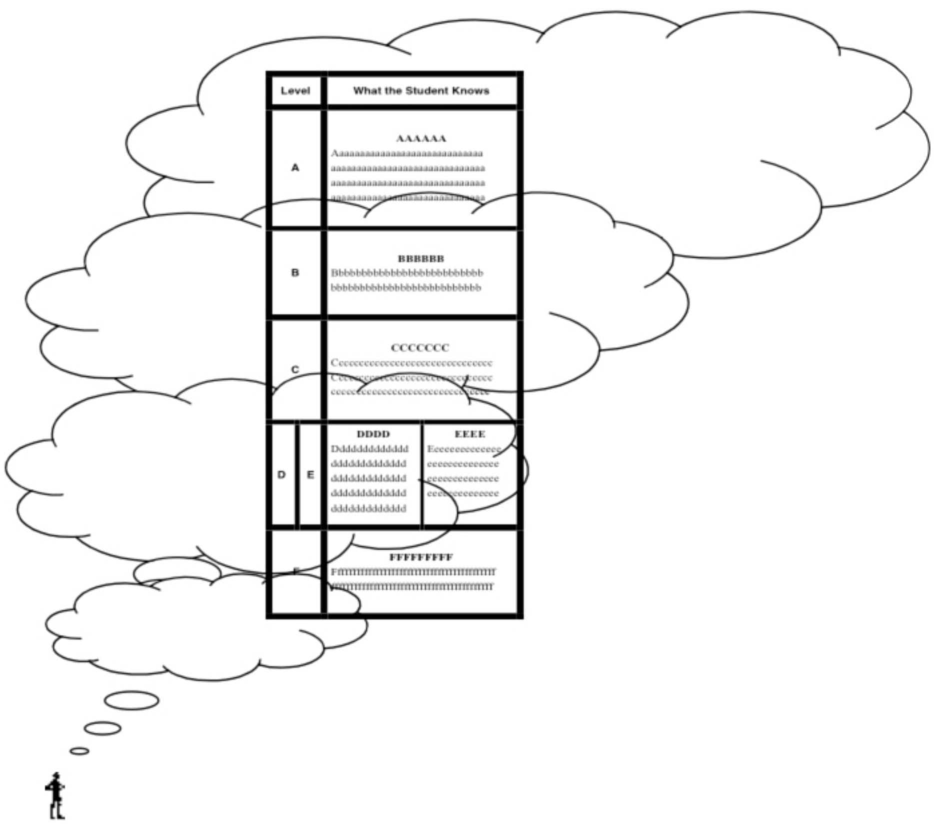


Figure 5. A learning progression with an associated construct map.

However, because more complex learning progressions quickly arise in any pragmatic learning situation, the simple relationship given in Figure 5 must be adapted to incorporate complexities. Beyond a single dimension, a collection of construct maps can be used together to map out a more complex learning progression (Draney, 2009; Wilson, 2009). Perhaps the simplest way to envisage this would be to consider each construct map as a single dimension of a multidimensional structure, shown in a generic illustration for a three-dimensional learning progression in Figure 6. This figure, along with the understanding that the relationship between the three dimensions is estimated using a correlation coefficient, is for a multidimensional item response model (or sometimes, if the observed variables are continuous, a factor analysis model). In addition, learning progressions often involve more complex relationships among dimensions, where a certain level of a dimension must be attained before success can be reached at a particular level of another dimension. These are intriguing formulations and interested readers are advised to refer to Wilson (2009, 2012) for a discussion of their implications for measurement. However, these are beyond the scope of this paper.

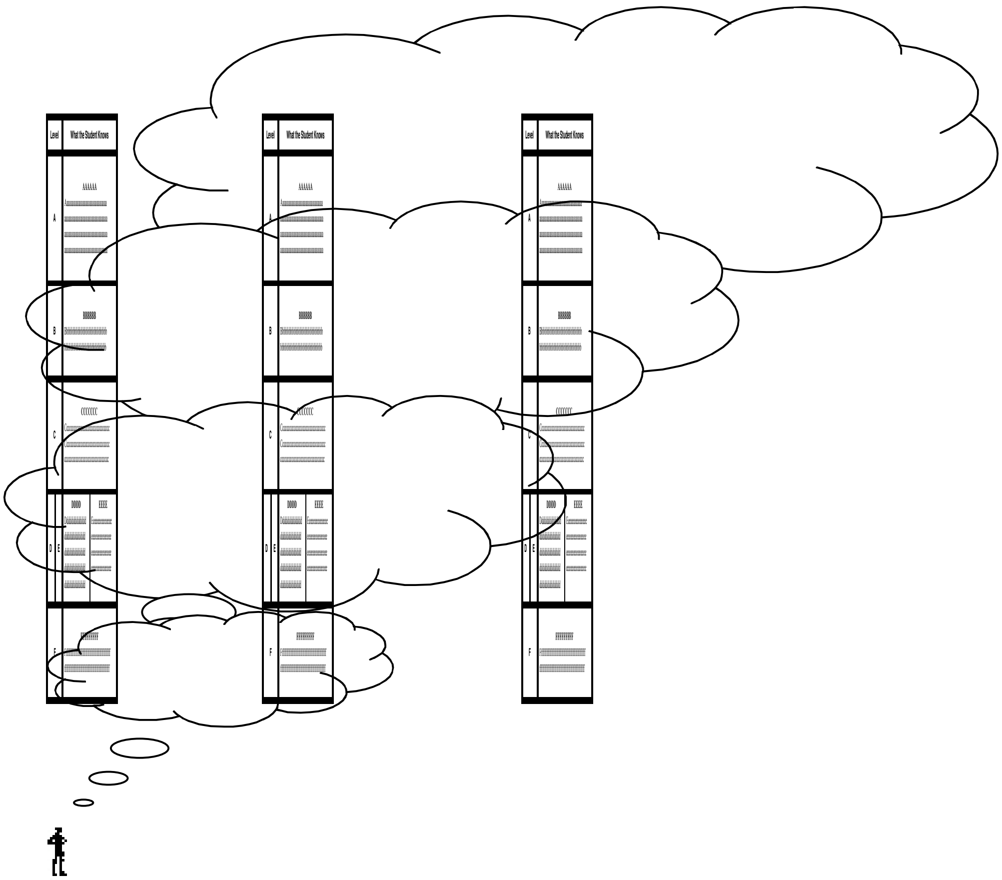


Figure 6. One possible relationship—the levels of the learning progression are levels of several construct maps.

Measuring a learning progression where target learning is represented as a cycle

The description above provides a now well-established narrative of how a learning progression, with its basic idea of directionality towards greater sophistication, can be supported by unidimensional constructs, which then may be combined into multidimensional structures. The Conceptions of Statistics construct is an example of a unidimensional learning outcome, which could also be augmented with other data modeling constructs to form a multidimensional learning outcome. Now I will shift my focus to a different scenario, which involves adapting this measurement approach to assess a somewhat more complex cognitive structure, where target learning is represented as a cycle. By this I mean that the central target of the learning progression is for students to master a multi-step cycle of cognitive activities. There are many examples of this form in the higher levels of school curriculum and it is a fundamental form for many university-level subjects. A typical example (in its simpler form) in the middle school science curriculum, typically extended in high school curricula and at the university level, is the carbon cycle (see Jin, Zhan, & Anderson, 2013) for a discussion of the scientific cycle itself and the complexities of measuring it). Although the carbon cycle itself is a cyclical phenomenon, this paper is rather focused on the cyclical learning process involved. There are many such cycles extant in a wide range of literatures and the four building blocks of the BEAR Assessment System constitute a clear example (see Figure 1).

These cycles can be represented generically, as in Figure 7. Here, the curriculum developer is contemplating, say, a four part cycle. One way to model this structure would be to insert a construct map into each part of the cycle, as in Figure 8. The hypothesis here is that as a student's sophistication in understanding and applying the cycle develops, the student's command of each of the parts will also grow. This seems an elementary and obvious first step towards measuring a learning progression that embodies the cycle as its target: the progression is within the parts of the cycle.

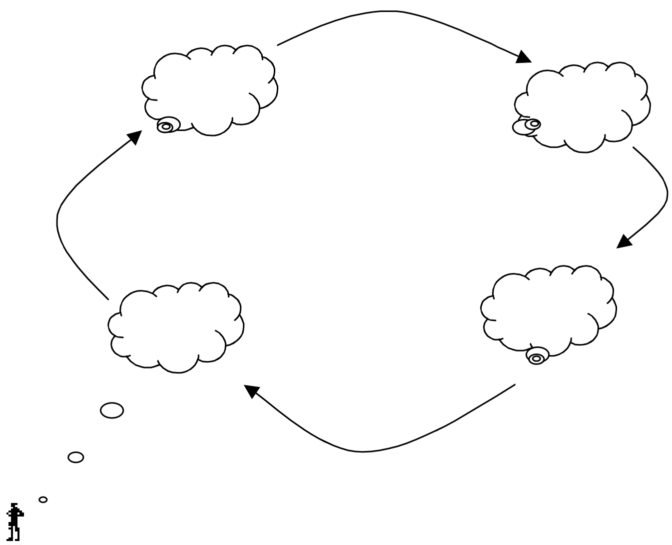


Figure 7. A representation of a learning cycle.

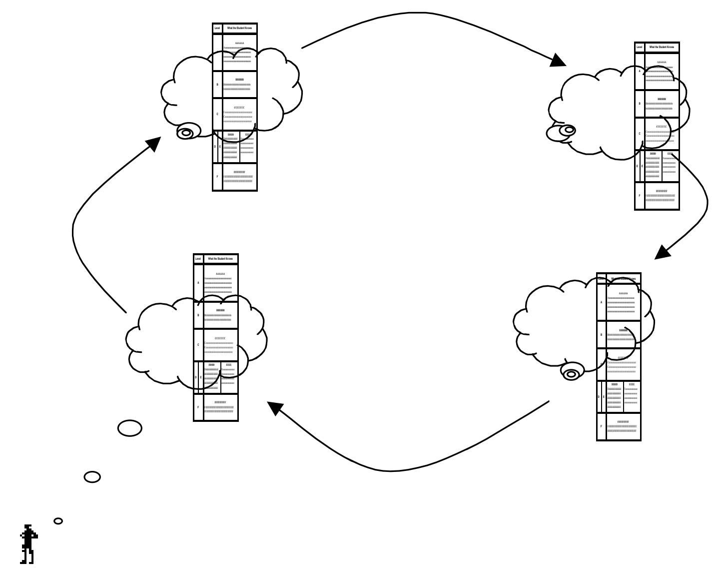


Figure 8. Each part of the learning cycle could be represented as a construct map.

Duckor, Draney and Wilson (2009) provide one example of developing an approach to measuring cycles. Their paper describes the *Measuring Measuring* project, which applies the BAS to student understanding and accomplishments in the BEAR Assessment System (i.e., they were studying how to measure graduate student performance in introductory measurement classes where the curriculum was structured in terms of the BAS). For example, Figure 9 shows their representation of the Understanding Construct Maps construct. They chose as their assessment framework a structure that parallels the one in Figure 8. Figure 10 shows their basic model. As demonstrated, there is one measured construct for each part of the BAS cycle. The aforementioned paper then details how they designed items and constructed outcome spaces for the four constructs (i.e., the four building blocks), and then estimated item and person locations to show in Wright maps.

Understanding Construct Maps

	Respondents		*Responses to items*
High	*Respondents* who can integrate normative and criterion referenced aspects of the construct map. They understand the construct map as a hypothesis about the empirical distribution of e.g. item difficulties and person proficiencies and try to align items design, outcome space, and measurement model with map.	Integrative 5	*Response to items* indicates understanding of where and when the particular construct map representation can be employed to strengthen/weaken inferential links between specific aspects of measurement system. Also demonstrates capacity to compare theoretical expectations against empirical findings.
	Respondents who can explain why some persons and items have more or less of the construct being measured. They may also be able to articulate the relationship between both.	Multiestructural 4	*Response to items* indicates understanding of how developing the orderliness of the Construct map aids in the development of items to populate scale, sketch out initial scoring strategy, provide validity check on content.
	Respondents who can describe the construct map in terms of a single concept or definition. They recognize the need for descriptions of ordered levels. They may also begin to develop sub- constructs to deal with complexity.	Definitional 3	*Response to items* indicates basic understanding of criteria for developing a Construct map. Shows that respondent can detect issues with construct definition, orderliness, dimensionality, etc.
	Respondents who can begin to describe all the goals, standards, factors, scales, etc. of interest but have not yet proposed to measure any single phenomena. They may be rigid and inflexible about the need to narrow and focus on a single construct map.	Discordant 2	*Response to items* indicates emerging notion of construct, but defined in multiple or vague ways. Shows that respondent may not be aware of inferential nature of measurement and the role of hypothesizing in advance.
Low	*Respondents* who ignore or are not attentive to any notion of cognitive or construct-based theory.	Pre-measurement 1	*Response to items* indicates a lack of concept or understanding of notion of construct or is off-topic.

Figure 9. The Understanding Construct Maps construct map.

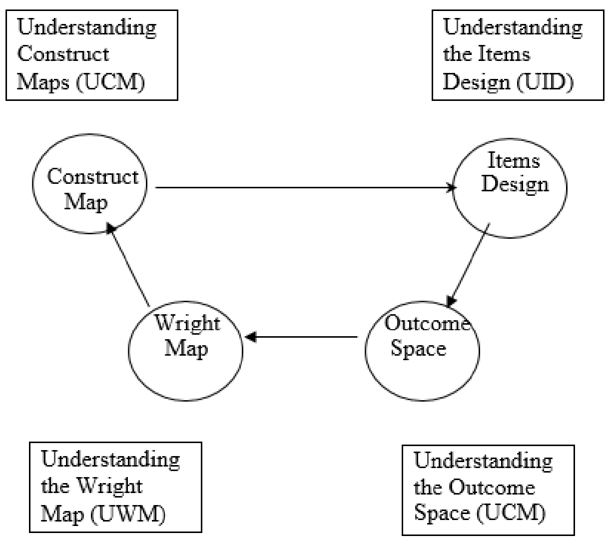


Figure 10. A representation of the four building blocks, each as a construct map, from the Measuring Measuring project.

However, this particular assessment framework, while nicely consistent with the single and multidimensional one framework shown above (i.e., for ADMSR) is somewhat unsatisfactory in this case. It does not explicitly address the existence of logical connections between the constructs. For example, each building block is logically dependent on gaining certain knowledge and accomplishments in the preceding blocks, precisely because they are all part of the BAS cycle. Hence, it would make sense to incorporate this idea into a measurement model in order to conceptualize how to measure these connections. This idea might consist of several different aspects. Consider one aspect, illustrated in Figure 11. This figure demonstrates how an educator might think of the increasing sophistication a student would experience in initially understanding BAS. The student might start with (a) an understanding of just one building block, say, the construct map (although it could also be the items design), as shown in the bottom left-hand corner of Figure 11. Then, add in (b) an understanding of how the construct map relates to the items design, followed by (c) an understanding of the three-way relationship between the construct map, the items design and the outcome space. Finally, (d) achieve an understanding of the whole BAS cycle, the relationships amongst all for building blocks. Note that, in the diagram, each of the building blocks is still covered by its own construct map —hence, there are five constructs represented in Figure 11: the Building Blocks Relationships construct, which has just been described, and one for each building block, as before.

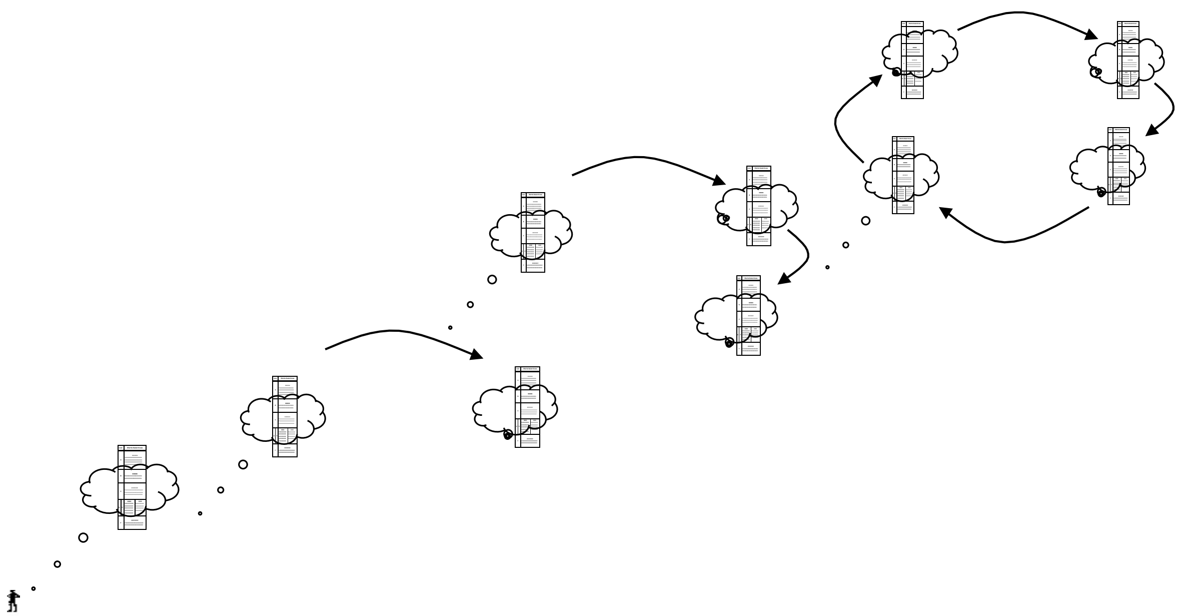


Figure 11. One way to represent development of understanding of a cycle.

Another form of development in understanding a cycle could occur after the sequence of development described in Figure 11. Assuming that students do indeed now have an understanding of the full four blocks, then the way in which students understand the building blocks might evolve, becoming more sophisticated as they develop through the construct levels. This is illustrated in a very generic way in Figure 12, where the different pattern inside each cycle of building blocks indicates the qualitative changes taking place. There are numerous ways that the understanding of the entire cycle could deepen, depending on the area of cognition represented. For example, the cycle itself could be applied to analyze a specific question or issue, or, it could be augmented by another component of a different nature than the original parts.

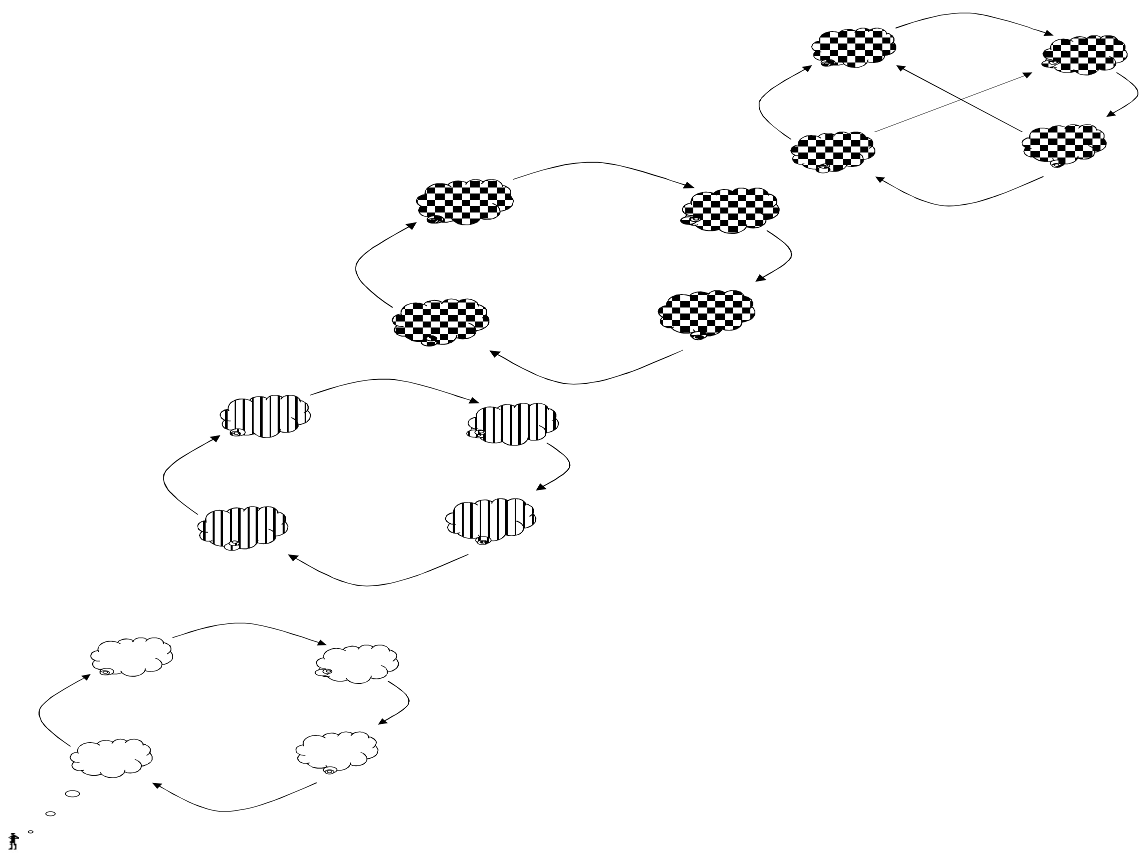


Figure 12. A second way to represent development of understanding of a cycle.

A paper by Brown and Wilson (2011) discusses one example of this type of learning progression. The study takes the basic BAS cycle (as in Figure 1) and adds complexity in two steps (see Figure 13). In this figure, the bottom left-hand corner shows the original representation of the BAS cycle; the next representation in the middle of the figure shows the same four building blocks, but here the four links have been augmented by two more diagonal lines showing the remaining links among the four blocks —these each symbolize important secondary elements of the BAS. The diagonal running from outcome space to construct adds a very important sub-cycle to the BAS, sometimes called the inner *qualitative triangle*, consisting of those two plus the items design. A measurement developer might cycle around the inner triangle several times before collecting enough data in a single instance to use the measurement model —hence appreciating that this is an important accomplishment. The other diagonal, running from the measurement model back to the items design signifies the important procedural step of checking the technical "fit" (Wilson, 2005) of the items (and other fixed measurement effects, such as raters, etc.) to (a) ensure that the formal assumptions of the statistical model have been satisfied and (b) to substantively interpret any lack of such fit. The third representation, in the top right-hand corner, represents another important step in appreciating and applying the BAS —this time involving an extra block. This block is not usually included in the list of the four BAS measurement building blocks because it is actually outside of the BAS and is not formally part of a measurement system. It represents the research questions under investigation, which constitute the motivation for employing the BAS to develop the assessments. Of course, a given assessment might be used with different sets of research questions (and, given the effort needed to perform the job well, one hopes that this would indeed be the case). Nevertheless, the research questions provide an important guiding framework for applying all four building blocks, as well as for understanding how they operate together.

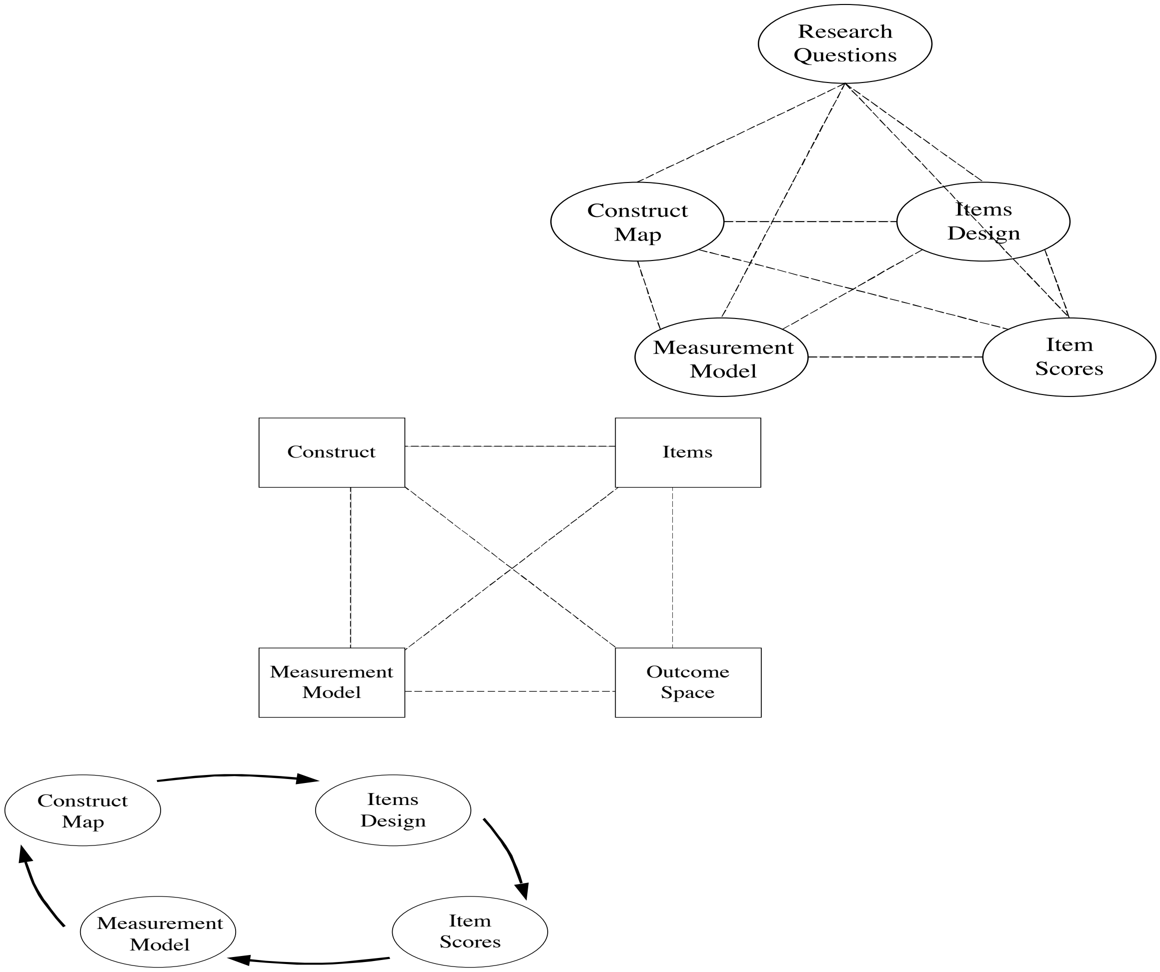


Figure 13. The three levels of understanding of the BAS cycle, as envisaged in Brown & Wilson (2011).

Discussion

The abstract structures described and discussed above are quite complex, certainly more complicated than those typically used as the underlying model in item response theory. It is my view that a clear exposition of the modeling approach described here will become crucial for progress in the assessment of learning progressions as they are applied in increasingly broader and deeper ways. Indeed, learning progressions pose multiple challenges to the typical assessment development and psychometric modeling techniques that have dominated assessment for the last 100 years. In particular, the limitation to unidimensional conceptualization of classical test theory and standard item response modeling will prove inadequate (and already has). At minimum, more attention must be focused on multidimensional models and their more complex extensions, as explored above.

The different approaches to measuring a learning progression described in this paper, where target learning is represented as a cycle, constitute only a limited sample of the many other possibilities that quickly become extant when delving into the ideas that curriculum developers have as they create curricula and associated learning progressions. For example, to cite just one complexity common among such efforts, links between the different levels of the construct maps may be necessary. These would usually be hypothesized on the basis of cognitive theories underlying the learning progressions and/or task analyses of the items used to measure the constructs. In this case, there could be specific level-to-level linkages in the statistical model (e.g., Wilson, 2012), which would involve considerable new work on both the statistical modeling needed and how to interpret such links. A second complexity that should be fairly clear from this text is that developers may want to (and probably would) layer a model similar to what is illustrated in Figure 12 over one like the model shown in Figure 11. This would provide a more complete picture of development from the very spotty beginning of a learning progression to a much more complex final stage. Many more complexities can, and will, need to be developed.

The different modeling strategies will need to be coordinated with the instructional strategies employed in the curricula intended for student learning. The assumption shown in Figure 12 has rather different educational implications than what is shown in Figure 11. If the Figure 11 model was deemed appropriate, the four parts of the cycle could simply be taught one at a time, ending with the integration of the fourth. If the Figure 12 model was determined to be suitable, then the whole cycle could be taught basically at the same time, gradually improving each part as the teaching iterated around the cycle several times. Scientific evidence will not necessarily reveal that one such approach is better than the other; in fact, it may well be the case that instructional strategies could be developed to succeed in either case. The assessment structure must therefore be designed taking into account the perspective of the curriculum involved.

In presenting the constructs in this paper, I have not made great distinctions among the psychometric models that might actually be used to embody the constructs in an empirical sense. However, it is important to note that many of the representations used in this paper make the implicit assumption that the results of psychometric analyses can be represented in the Wright map format (which makes the connection back to the construct map very explicit). It turns out that this is a significant requirement on both the items that are used for the assessments and on the psychometric models that can be used for the data analysis: the psychometric models need to be amenable to the Wright map style of presentation, and this means that they will need to come from within the Rasch family of models. This means that the items would all need to fit the specific Rasch model used reasonably well (Wilson, 2005).

Finally, the presentation in this paper is quite exploratory. The formulation of a set of sound development steps (e.g., steps such as instrument development, analysis, learning from reflection on the results of the analysis and returning to development) will require actually developing multiple assessments for numerous learning progressions. It will also take time to properly digest the lessons of these exercises. This paper therefore merely presents the first few steps of a path that portends to be long and interesting.

The original article was received on October 9th, 2013
The revised article was received on January 13th, 2014
The article was accepted on January 22nd, 2014

References

Adams, R. J., Wilson, M., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1-23. doi: 10.1177/0146621697211001

Alonzo, A. C., & Gotwals, A. W. (Eds.). (2012). *Learning progressions in science*. Rotterdam, The Netherlands: Sense Publishers.

Black, P., Wilson, M., & Yao, S. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives, 9*, 71-123. doi: 10.1080/15366367.2011.591654

Brown, N. J. S., & Wilson, M. (2011). A model of cognition: The missing cornerstone of assessment. *Educational Psychology Review, 23*(2), 221-234. doi: 10.1007/s10648-011-9161-z

Center for Continuous Instructional Improvement (CCII). (2009). *Report of the CCII panel on learning progressions in science*. New York: CPRE Research Report, Columbia University.

Clements, D. H., Wilson, D. C., & Sarama, J. (2004). Young children's composition of geometric figures: A learning trajectory. *Mathematical Thinking and Learning, 6*, 163-184.

Draney, K. (2009). *Designing learning progressions with the bear assessment system*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA.

Duckor, B., Draney, K., & Wilson, M. (2009). Measuring measuring: Toward a theory of proficiency with the constructing measures framework. *Journal of Applied Measurement, 10*(3), 296-319.

Jin, H., Zhan, L., & Anderson, C. W. (2013). Developing a fine-grained learning progression framework for carbon-transforming processes. *International Journal of Science Education, 35*(10), 1663-1697. doi: 10.1080/09500693.2013.782453

Lehrer, R., Kim, M.-J., Ayers, E., & Wilson, M. (in press). Toward establishing a learning progression to support the development of statistical reasoning. In J. Confrey, & A. Maloney (Eds.), *Learning over time: Learning trajectories in mathematics education*. Charlotte, NC: Information Age Publishers.

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

National Research Council (NRC). (2006). *Systems for state science assessment. Committee on Test Design for K-12 Science Achievement.* Washington, D.C.: National Academy Press.

National Research Council (NRC). (2007). *Taking science to school: Learning and teaching science in grades K-8. Committee on Science Learning, Kindergarten through Eighth Grade.* Washington, D.C.: National Academy Press.

Piaget, J. (1950). *The psychology of intelligence*. London: Lowe & Brydone.

Piaget, J. (1971). *Biology and knowledge*. Chicago: University of Chicago Press.

Wilson, M. (2003). On choosing a model for measuring. *Methods of Psychological Research, 8*(3), 1-22.

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

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

Wilson, M. (2012). Responding to a challenge that learning progressions pose to measurement practice: hypothesized links between dimensions of the outcome progression. In A. C. Alonzo, & A. W. Gotwals (Eds.), *Learning progressions in science* (pp. 317-344). Rotterdam, The Netherlands: Sense Publishers.

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

Wright, B. D., & Masters, G. N. (1981). *Rating scale analysis*. Chicago: MESA Press.

Wright, B. D., & Stone, M. (1979). *Best test design*. Chicago: MESA Press.