DESIGNING LEARNING PROGRESSIONS WITH THE BEAR ASSESSMENT SYSTEM

Karen Draney, University of California, Berkeley

INTRODUCTION

The Berkeley Evaluation and Assessment Research (BEAR) Assessment System has been used as a tool for developing Learning Progressions for the past decade. Projects with which we have been involved in this process include Living by Chemistry (Claesgens, Scalise, Wilson, and Stacy, 2009), Assessing Data Modeling and Statistical Reasoning (ADMSR) (Lehrer, et. al., 2007), Diagnostic E-Learning Trajectories Approach (DELTA) applied to rational number reasoning for grades 3-8 (Diakow, Feldman, and Wilson, 2009; Wilson and Feldman, 2008), and Michigan State University’s Carbon Cycle project (Mohan, Chen, and Anderson, in press). This paper will describe the BEAR Assessment system in terms of its principles and its components, as well as give examples of these components and how they contribute to the Learning Progression, from two of our science projects, LBC and the Carbon Cycle.

The Bear Assessment System is an integrated approach to developing assessments that provide meaningful interpretations of student work relative to the cognitive and developmental goals of a curriculum. It is grounded by four key principles guiding assessment development and includes four associated building blocks that are tools for constructing robust assessments aligned with curricular goals and instructional activities. These principles are:

- Assessment should be based on a developmental perspective of student learning.
- What is taught and what is assessed must be clearly aligned.
- Teachers are the managers and users of assessment data.
- Classroom assessment must uphold traditional standards of validity and reliability.

These four principles are described in considerably more detail in Wilson’s paper (this volume). They also relate to the Assessment Triangle developed by the National Research Council Committee on the Foundations of Assessment and published in their report, *Knowing What Students Know* (Pellegrino, Chudowsky & Glaser, 2001). The Assessment Triangle, shown in Figure 1, is a model of the essential connections and dependencies present in a coherent and useful assessment system. Meaningful connections among the three vertices, cognition, observation, and interpretation, are deemed essential for assessment to have a positive impact on learning. Thus, assessment activities (the observation vertex) must be aligned with the knowledge and cognitive processes one wishes to affect through the instructional process (the cognition vertex), and the scoring and interpretation of student work must reflect measures of the same knowledge and cognitive processes (the interpretation vertex).

Insert Figure 1 here
The BEAR Assessment System includes four building blocks for constructing quality assessments:

- Construct Maps
- Items Design
- Outcome Space
- Measurement Model

These building blocks follow from the four principles given above, and map to the NRC Assessment Triangle, as shown in Figure 1. They operationalize the BEAR Assessment System principles.

We will now introduce the two projects whose work we will use as examples in the remainder of the paper.

**Living by Chemistry:**

Living by Chemistry (LBC) is a high school chemistry curriculum project that proposes framing the “big ideas” of chemistry to provide developmental cohesion across the curriculum and promote conceptual understanding. The resulting learning progression framework, called Perspectives of Chemists, is intended to allow measurement of individual conceptual change in chemistry over time, in order to inform on the patterns and characteristics of the conceptual “change space” in the domain. The framework is based on integrating conceptual change theory with National and California State Science Standards, expert opinion, interviews with teachers, surveys of topics in high school chemistry textbooks, and classroom observations of students. The resulting “frame” was tested and further developed in two small pilot studies, in which student task responses were analyzed for progressions of student understanding, calibrated with item response theory measurement models, and mapped to proposed conceptual models of chemistry. A larger initial field study at the secondary level was conducted, which included further exploration of the framework at the university level as part of the UC Berkeley ChemQuery project.

**The Carbon Cycle**

Understanding the ecological carbon cycle is critical to our view of environmental science literacy—the capacity to understand and participate in evidence-based discussions of the effects of human actions on environmental systems. Environmentally science literate citizens need to understand how sea ice available to polar bears in the Arctic is connected to processes inside leaf cells in the Amazon rain forest and to American consumers’ choices about what car to buy. The global climate is changing and with this change comes increasing awareness that actions of human populations are altering processes that occur in natural ecosystems. The “carbon cycle” is no longer a cycle, on either local or global scales; most environmental systems—especially those altered by humans—are net producers or net consumers of organic carbon. Similarly, humans have altered the global system so that there is now a net flow of carbon from forests and fossil fuels to atmospheric carbon dioxide.
Recently in the Carbon Cycle project, we have been involved in the development of the set of construct maps which comprise the learning progression for student understanding of organic carbon. We have been engaged in qualitative analysis of student responses to a number of open-ended items, administered to upper elementary, middle school, and high school students, as well as some clinical interview questions, all associated with the carbon cycle (see also Mohan, et. al., 2006). We have used this analysis, as well as our goals for student learning by the end of high school, and the scientific understanding necessary to make informed environmental decisions, to construct the set of progress variables. We originally selected five progress variables; these include Structure of Systems, Tracing Matter, Tracing Energy, Citizenship, and Change Over Time. The focus of the research for the first portion of the project has been on the Tracing Matter and Tracing Energy variables; we will use these as examples throughout the rest of the paper.

Illustrating the building of a learning progression: The BEAR Assessment System exemplified

Construct Maps

A construct map, which defines a latent trait or construct, is used to represent a cognitive theory of learning consistent with a developmental perspective. This building block is grounded in the principle that assessments be designed with a developmental view of student learning. This means that the underlying purpose of assessment is to determine how students are progressing from less expert to more expert in the domain of interest, rather than limiting the use of assessment to measure competence after learning activities have been completed. Generally, we want to describe a continuum of qualitatively different levels of knowledge from a relatively naïve level to a more expert level. The knowledge, skill, or ability one wishes to measure is referred to as the construct, while the definitions of what it means to have more or less of the construct, in a cognitive sense, is referred to as the construct map. A construct map serves as a mechanism for defining and representing what students know and can do at several levels; it facilitates the alignment of responses on student work with the cognitive goals of a curriculum. A collection of construct maps taken together can comprise a learning progression.

How the construct maps are arranged when they are collected together is an open question (see Wilson, in press; Wilson, this volume). Both of the learning progressions represented as examples in this paper are hypothesized to be arranged such that the levels of the learning progression are horizontally consistent across several construct maps – i.e. a student operating at a given level in one of the constructs will be operating at more or less the same level across the other constructs in the learning progression.

Figure 2 shows the three construct maps (each represented within a column) for the LBC curriculum: Matter, Change, and Energy (headers ranging across the top of the table). The leftmost column provides general names for each level of all three construct maps, which follow the same structure. These levels range from the lowest (Notions) at the bottom of the page, to highest (Generation). Within each cell of the table are questions that students at that level of performance on that construct are able to answer for the first time when they reach that level. For example, students at the Formulation level on the Matter construct are able to answer questions about the interaction between atoms.
Figure 3 shows two different views of the Carbon Cycle learning progression. The top view is the general structure of all of the components of the learning progression. The leftmost column shows the levels of the learning progression, lowest at the bottom of the page and highest at the top. Across the top of this page are the various ways that student learning performances are classified, by Principles (these include Tracing Matter and Tracing Energy), Practices (what scientists do), and the processes through which matter and energy are traced (such as combustion, cellular respiration, photosynthesis, and decay).

The bottom view of the learning progression shows two different ways in which students could progress through this learning progression, starting in upper elementary school, and continuing through high school. The view on the left side is what, during the scope of this project, we have seen given the current curricula used in the schools we have investigated. The view on the right is what we would consider a more successful view. In this view, students would progress through the same general levels of reasoning, but without forming many of the misconceptions they currently develop.

Figure 4 shows a more detailed view of one of the constructs in the Carbon Cycle learning progression (Tracing Matter, one of the two that have been fully fleshed out so far in the project). This describes the specific accomplishments and limitations we see at each level of the learning progression as they appear in items that tap the Tracing Matter construct. For example, when students at Level 3 answer such items, they can talk about what happens at both atomic-molecular and macroscopic levels during chemical equations, but they do not consistently recognize that gases have mass, and they do not always understand the chemical identities of a substance.

Items Design and Outcome space

The items design and outcome space are in theory quite different, but in practice highly intertwined, as it is quite difficult to develop items without considering what kinds of responses might be given; and, in the case of closed-ended items, one must develop the possible outcomes along with the items. Therefore, these two elements of the system will be considered together.
The items design building block is a framework for designing tasks to elicit specific kinds of evidence about student knowledge, as described in one or more construct maps. The guiding principle is that assessment should be seamlessly integrated into the instructional activities of a course. That is, assessment is not merely tacked on at the end of instructional units, but is embedded in normal classroom activity and may even be, from the student’s point of view, indistinguishable from instruction. This necessitates unambiguous linkages between assessment activities and curricular content so that assessment results can be used to improve teaching and learning activities in an ongoing manner. Items are written with the intention of producing evidence of specified levels of understanding along a construct. In some cases, for example in embedded assessments, it may be desirable for a collection of items in a particular instrument to focus on only one or two levels at a time; in other cases, such as in end of unit summative assessments, one may want a collection of items that cover the complete range of the construct. In addition, items may be designed to measure multiple aspects of knowledge in a single comprehensive response, thereby producing more information from an amount of student effort that might otherwise provide evidence of only one aspect of knowledge.

The outcome space describes in detail the qualitatively different levels of responses associated with the construct map for a particular prompt or stimulus. This building block operationalizes the principle that teachers be the primary managers of assessment in the classroom. To accomplish this, they must have not only the data that they need to assess student learning, but also the skill set required to use that data effectively. This implies a data-driven approach to teaching, in which teachers use evidence to draw inferences about student knowledge and understanding. This evidence can then suggest actions that teachers and students can take to improve learning outcomes. In addition, teacher judgment becomes a valued and integral part of the assessment feedback loop. The purpose of the outcome space, then, is to facilitate identification of student responses corresponding to a particular level on a construct; teachers use the outcome space to assign scores to student work. We often organize moderation sessions in which teachers discuss with one another how they might score individual student work products to improve consistency between scorers and to help teachers become more consistent in their own scoring. This activity also helps teachers internalize the developmental goals of the curriculum as they focus attention on student locations along the construct continua and reflect on what it means to be located at one level or another.

The BEAR system requires that every item used in an assessment provide evidence of performance on one or more of the construct maps, and that every response to an item be associated with a level of that construct map. For example, Figure 5 shows an item from one of the LBC assessments that provides evidence about a student’s level on the Matter construct. This is an open-ended item which allows the student to answer at a wide variety of levels. It reflects the philosophy of LBC, whose developers had noticed that students who were capable of solving traditional chemistry problems at what seemed like a high level, would revert back to answering at the Notions level whenever the item was not like one they had seen in their chemistry class, although it involved the same principles.

-------------------------------
Insert Figure 5 here
-------------------------------
Figure 6 shows two different sets of responses to the LBC smell item. The first set of responses give evidence that the students are at the Notions level of the Matter construct map. They use only common sense ideas about ideas they are familiar with, and general reasoning skills, but do not attempt to employ any knowledge about chemistry. The second set of responses show evidence of understanding at the Recognition level. They are able to state the general chemical principle that is involved, but cannot give specifics about the particular problem at hand.

Not all items associated with the BEAR Assessment System must be open-ended. Figure 7 shows an item from a middle and high school assessment used in the Carbon Cycle project. Although this item includes an open-ended explanation component, which we are using in part for validation purposes, a great deal about a student’s level on the construct map can be told simply by looking at the pattern of “yes” and “no” answers that the student circles to the various potential energy sources for plants. Students who circle “Water” as an energy source, for example, are probably students with a Level 1 understanding of energy. They are defining energy in terms of general needs, and anything the plant “needs” to grow will probably be labeled as an energy source. Students who circle “light,” “air,” and “plants make their own energy,” and no other options, probably have a Level 3 understanding of energy. They have learned about photosynthesis, and that it involves both air and light, but they still have significant misunderstandings about this phenomenon, and about energy in general. They still are not successfully conserving energy, nor are they differentiating between matter and energy.

The closed response part of this item is similar to the Ordered Multiple Choice (OMC) item type described in Briggs, Alonzo, Schwab, and Wilson (2006). An OMC item looks exactly like a multiple choice item, however, rather than scoring the item “right/wrong”, each possible response choice is associated with a level of the construct map. While the evidence gathered from a single such item is not enough to conclude much about a particular student’s level of functioning, a pattern of such evidence across a number of items would be very informative, and would work well in a large-scale situation.

The final building block of the BEAR Assessment System is the measurement model, which defines how inferences about student understandings are to be drawn from the scores. The principle underlying this building block is that classroom assessment should adhere to the same quality standards as high stakes assessment. Of course, the classroom context is different from that of high stakes assessment, so the manifestations of quality may well differ. This principle helps ensure that the inferences drawn from classroom assessment activities are both meaningful in their own right, and also consistent across multiple instruments. The latter is particularly
useful when making longitudinal comparisons to determine change or progress. In this stage of assessment development, associations between items and constructs are specified; we refer to this process as construct modeling. The psychometric approach we often use to model the data is a multidimensional Rasch-based item response model known as the multidimensional random coefficients multinomial logit model (MRCML) (Adams, Wilson & Wang, 1997). This model provides a convenient way to develop person proficiency measures and item difficulty measures using the same scale; a technique that facilitates the interpretation of student measures on the construct. A description of how a variety of relationships between constructs might yield a learning progression, and be modeled psychometrically, is given in Wilson (this volume).

To communicate with a wide variety of audiences, we use a visual representation of these models called a Wright map (Wilson, 2005), in honor of Dr. Benjamin Wright, a longtime proponent of this type of measurement model and this type of graphical representation. In such a map, item and level difficulties and person proficiencies can be graphed on the same scale. Lower difficulty items (and levels of items scored with more than two outcomes), and lower proficiency persons, appear at the bottom of the scale, and higher difficulty items, and higher proficiency persons, at the top. If a person appears at the same level as an item level difficulty, then that person has a .5 chance of scoring at that level (or higher) on that item, and a .5 chance of scoring below. Such a graphical representation allows us to check our theoretical hypotheses about how item levels relate to one another, and to the population of persons we draw from.

Figure 8 shows a Wright map for a small set of items related to the Matter construct from the LBC curriculum. On the far left is the logit scale (a scale centered at zero, and generally running from about -6.0 to about +6.0, although in theory they can range to infinity). Just to the right of that is the person distribution; each individual represented by an X. On the right side of the page is a representation of the difficulty of scoring at each level of interest on four different items (you can see that the items have been scored not just 0/1/2, but for levels 1 and 2, a 1-/1/1+ and 2-/2/2+. These additional scores were necessary to differentiate the various levels of student performance seen by the scorers involved in the project. The horizontal lines across the page represent the average value at which the step between each level on the construct map and the below it occurred (e.g. the average value at which the step between 1- and 1 occurred, and so on). These are based on all of the items, not only the four illustrated in this map. As you can see, most of the steps between a given two levels are fairly consistent; however, there are some which occur at a higher or lower level than might be expected. For example, in this Figure on question 3, the step between 1- and 1 happened at an easier value than might be expected. Such information can be used to investigate specific items.

In Figure 9, a similar Wright map is shown for data from the Carbon Cycle project’s Tracing Matter variable, although this graph includes data for all items associated with this variable, and also includes the average difficulty of each item across all of the possible responses (this allows us to determine if any items, or item types, are especially easy or difficult for the students). There is in this figure considerable random scatter within any one score level;
however, there does appear to be at least some consistency such that students at a particular ability level tend to answer most items at the same score level (if you recall, that was the theoretical structure of not only a given construct map, but across all construct maps related to the carbon cycle). There were in this dataset so few scores at level 4 of the construct map that this level is not represented here.

---------------------------------

Insert Figure 9 here

---------------------------------

Implications and Future Directions

With more widespread use of such learning progressions, built around constructs based on such a developmental perspective, student progress in the understanding of subject matter could be traced across a longer timeframe than is currently done. In the LBC example, the learning progression is designed to show progress through high school and early college chemistry; the Carbon Cycle project has worked on tracing student development from upper elementary through high school. The BEAR system is, however, deliberately neutral to the issue of grain size. Longer timeframes and coarser grainsize are illustrated in these two projects; we have also worked with shorter timeframes and finer grainsizes in other projects (e.g. the Assessing Science Knowledge project, an NSF-funded grant by the Full Option Science System to develop embedded assessments as well as pre and posttests for 16 elementary school science curriculum modules (Woodward, Draney, and Tan, 2005). Each of these is designed for a 9-12 week timeframe; thus, a much finer grainsize must be exemplified in the learning progression in order to see progress from the beginning to the end of the unit.

In this paper we have described two examples of learning progressions, showing each piece of the system as it is exemplified in a particular context. However, there are many other possible contexts in which the system could function. We have, for example, not developed learning progression in which the relation of the various constructs is very different from the two showed here (see Wilson, this volume). It is entirely possible, however, given the difference in curriculum and teaching methods in elementary and secondary school, that one should not use the same set of constructs at the two levels, even if the ultimate learning goals are the same. It may be that a lower elementary school student’s understanding of such issues as a food web, the growing of plants, and/or phase changes in matter, are the constructs that at the secondary level become the tracing of matter and energy. Such issues must be investigated further.
References


Figure 1. The BEAR building blocks and the assessment triangle.
Figure 2: A learning progression in Chemistry, as devised by the Living by Chemistry and ChemQuery projects

<table>
<thead>
<tr>
<th>Levels</th>
<th>Matter</th>
<th>Change</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>What new experiments can we design to gain a deeper understanding of matter?</td>
<td>What new reactions can be designed to generate desired products?</td>
<td>How can we use these models to optimize this type of change?</td>
</tr>
<tr>
<td>Construction</td>
<td>How can we understand composition, structure, properties and amounts?</td>
<td>How can we understand type, progression, and extent of change?</td>
<td>How do different models explain the chemical and physical changes that may occur?</td>
</tr>
<tr>
<td>Formulation</td>
<td>How can we think about interactions between atoms?</td>
<td>How can we think about rearrangement of atoms?</td>
<td>How can chemists predict the extent of energy transfer and the resulting products of chemical and physical change?</td>
</tr>
<tr>
<td>Recognition</td>
<td>How do chemists describe matter?</td>
<td>How do chemists describe change?</td>
<td>How do chemists explain energy transfer for chemical and physical changes?</td>
</tr>
<tr>
<td>Notions</td>
<td>What do you know about matter? (initial ideas, logic, real world knowledge)</td>
<td>What do you know about change? (initial ideas, logic, real world knowledge)</td>
<td>What do you know about energy in relationship to chemistry?</td>
</tr>
</tbody>
</table>
Figure 3: Two different views of the Carbon Cycle learning progression

<table>
<thead>
<tr>
<th>Levels of Achievement</th>
<th>Practices, principles, and processes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practices: Inquiry, accounts, decisions</td>
</tr>
<tr>
<td>4: Qualitative model-based accounts</td>
<td></td>
</tr>
<tr>
<td>3: “School science” narratives</td>
<td></td>
</tr>
<tr>
<td>2: Force-dynamic with hidden mechanisms</td>
<td></td>
</tr>
<tr>
<td>1: Force-dynamic narratives</td>
<td></td>
</tr>
</tbody>
</table>

**Learning performances** for specific processes and Levels of Achievement:
Inquiry, accounts, citizenship decisions about processes in socio-ecological systems

**Level 4:** Principled Accounts of Processes in a Hierarchy of Systems (Successful Constraints)

**Level 3:** Detailed Narratives with Unsuccessful Constraints

**Level 3:** Principled Accounts at Molecular Scale without Chemical Details

**Level 2:** Hidden Mechanisms About Events

**Level 2:** Successful Conservation at Macroscopic Scale

**Level 1:** Macroscopic Force-Dynamic Narratives about Actors and Events

Learning Trajectory: Structure-First focusing on Naming

Learning Trajectory: Principles-First focusing on Explaining
Figure 4: Detailed view of the Tracing Matter construct

<table>
<thead>
<tr>
<th>Level</th>
<th>Accomplishments</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 4: Qualitative model-based accounts</td>
<td>Model-based accounts of all carbon transforming processes.</td>
<td>Difficulty with quantitative reasoning.</td>
</tr>
<tr>
<td>Level 3: “School science” narratives</td>
<td>Stories of events at atomic-molecular, macroscopic, and large scales.</td>
<td>Mass of gases not consistently recognized. Incomplete understanding of chemical identities of substances.</td>
</tr>
<tr>
<td>Level 2: Force-dynamic with hidden mechanisms</td>
<td>Stories involving hidden mechanisms (e.g., body organs). Recognition of events at microscopic scale. Tracing matter through most physical changes Coherent stories of food chains. Macroscopic force-dynamic narratives about actors and events (e.g., needs of plants and animals).</td>
<td>Matter (especially gases) not clearly distinguished from conditions or forms of energy. Macroscopic events are associated with specific organs rather than cellular processes. Focus on reasons or causes for events rather than mechanisms. Vitalistic explanations for events involving plants and animals.</td>
</tr>
<tr>
<td>Level 1: Event-based narratives</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5: An open-ended item related to the Matter construct from the LBC curriculum

Both of the solutions have the same molecular formulas, but butyric acid smells bad and putrid while ethyl acetate smells good and sweet. Explain why these two solutions smell differently.
Figure 6: Sample responses to the LBC item, at two different levels of the construct map

<table>
<thead>
<tr>
<th>Level</th>
<th>Example Responses</th>
<th>Analysis</th>
</tr>
</thead>
</table>
| 1     | **Response A:** I think there could be a lot of different reasons as to why the two solutions smell differently. One could be that they're different ages, and one has gone bad or is older which changed the smell. Another reason could be that one is cold and one is hot.  
**Response B:** Using chemistry theories, I don't have the faintest idea, but using common knowledge I will say that the producers of the ethyl products add smell to them so that you can tell them apart.  
**Response C:** Just because they have the same molecular formula doesn't mean they are the same substance. Like different races of people: black people, white people. Maybe made of the same stuff but look different. | These students use ideas about phenomena they are familiar with from their experience combined with logic/comparative skills to generate a reasonable answer, but do not employ molecular chemistry concepts. |
| 2     | **Response D:** They smell differently b/c even though they have the same molecular formula, they have different structural formulas with different arrangements and patterns.  
**Response E:** Butyric acid smell bad. It's an acid and even though they have the same molecular formula but they structure differently. | Both responses appropriately cite the principle that molecules with the same formula can have different structures, or arrangements of atoms within the structure described by the formula. However the first answer shows no attempt and the second answer shows an incomplete attempt to use such principles to describe the simple molecules given in the problem setup. |
1. Which of the following is/are energy source(s) for plants? Circle yes or no for each of the following.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Water</td>
<td>Yes / No</td>
<td></td>
</tr>
<tr>
<td>b. Light</td>
<td>Yes / No</td>
<td></td>
</tr>
<tr>
<td>c. Air</td>
<td>Yes / No</td>
<td></td>
</tr>
<tr>
<td>d. Nutrients in soil</td>
<td>Yes / No</td>
<td></td>
</tr>
<tr>
<td>e. Plants make their own energy.</td>
<td>Yes / No</td>
<td></td>
</tr>
</tbody>
</table>

Please explain ALL your answers, including why the things you circled “No” for are NOT sources of energy for plants.
Figure 8: A Wright map of data gathered during a study of the LBC curriculum

<table>
<thead>
<tr>
<th>Student performance estimates</th>
<th>Item difficulty estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1</td>
</tr>
<tr>
<td>2.0</td>
<td>2</td>
</tr>
<tr>
<td>1.0</td>
<td>2-</td>
</tr>
<tr>
<td>0</td>
<td>1+</td>
</tr>
<tr>
<td>-1.0</td>
<td>1</td>
</tr>
<tr>
<td>-2.0</td>
<td>1-</td>
</tr>
</tbody>
</table>

$x = $ one student

$x$
Figure 9: A Wright map from the Carbon Cycle project