Berkeley Evaluation & Assessment Research Center
Graduate School of Education
University of California, Berkeley
Berkeley, CA 94720-1670
http://bearcenter.berkeley.edu
Technical Report Series
No. 2005-03-01

The BEAR Assessment System: A Brief Summary for the Classroom Context

Cathleen A. Kennedy

March 30, 2005

The BEAR Assessment System: A Brief Summary for the Classroom Context

The Bear Assessment System (BAS; Wilson & Sloan, 2000) 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 building blocks (each associated with one of the principles; Wilson, 2005) that are tools for constructing meaningful 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 sound standards of validity and reliability.

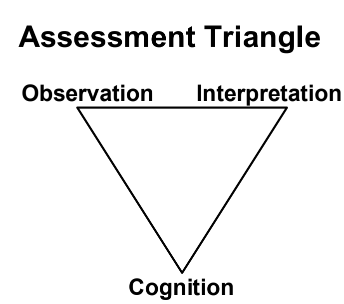


Figure 1. The National Research Council's assessment triangle.

These four principles 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* (NRC, 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. In this triangle, assessment activities (the *observation* vertex) must be aligned with the knowledge and cognitive processes (the *cognition* vertex) one wishes to affect through the instructional process, and the scoring and interpretation of student work (the *interpretation* vertex) must reflect measures of the same knowledge and cognitive processes. Meaningful connections among the three vertices, cognition, observation, and interpretation, are deemed essential for assessment to have a positive impact on learning. We refer to this whole process as *construct modeling*.

Four Building Blocks

The BEAR Assessment System includes four building blocks for constructing quality assessments: Construct Maps, Items Design, the Outcome Space, and the Measurement Model. These building blocks map to the NRC Assessment Triangle, as shown in Figure 2 below, and operationalize the BEAR Assessment System principles.

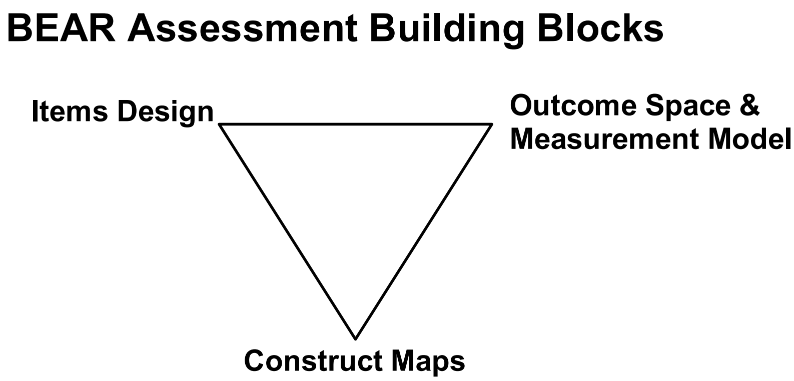


Figure 2. BEAR Assessment System building blocks.

A *construct map*, which defines a latent variable 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 are to 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[1] 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

A **construct** is what you want to measure.

[1] The BAS can be applied to assessment situations in which single or multiple continua are the underlying form of the construct. It is not designed for other underlying forms, such as latent classes.

of the curriculum. An example of a generic construct map is shown in Figure 3. In this example, three qualitatively different levels of knowledge are defined.

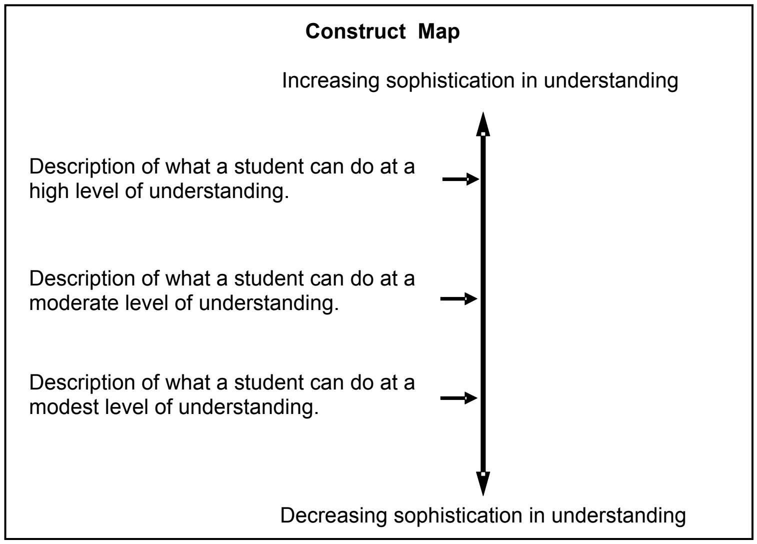


Figure 3. A generic Construct Map describing how students are expected to perform at three levels of understanding on the construct.

Items provide the mechanism to gather data about what students know and can do.

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 (and then multiple pieces of information may be mapped onto multiple construct maps).

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 are to 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 assessment 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. The outcome space and the associated construct map become the evidentiary foundation for teachers to use on a daily basis in their classrooms, on both formal assessments and in informal instructional contexts.

Scores align student responses with levels of knowledge.

We often organize *moderation sessions* in which teachers discuss with one another how they might score individual student work products. One objective is to improve consistency between scorers and to help teachers become more consistent in their own scoring. Perhaps more importantly, his 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 a) what it means to be located at one level or another, and b) what soft of evidence is required to conclude that one knows where a student is located on the construct map.

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 sound quality standards. Of course, the classroom context is different from that of traditional standardized assessment, so the manifestations of quality may well differ. This principle helps ensure that the inferences drawn from classroom assessment activities are 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. 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 (MRCMLM) (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. An example of this alignment is shown in Figure 4. For example, we can describe what students at a certain level on a construct can be expected to do based upon items aligned at that level and below it on the same scale. When a person and a dichotomous item are parallel to one

Measurement models connect the evidence to inferences about what students know.

another on the map, the person has a 50-50 chance of responding correctly to that item. The person has a higher (than 50-50) chance of responding correctly to items below his or her location, and a poorer chance of responding correctly to items above his or her location. Similar statements can be made for specific scores on polytomous items. In the example shown in Figure 4, Student A has a 50-50 chance of responding correctly to item 3, and a better chance of performing correctly on items 1 and 2.

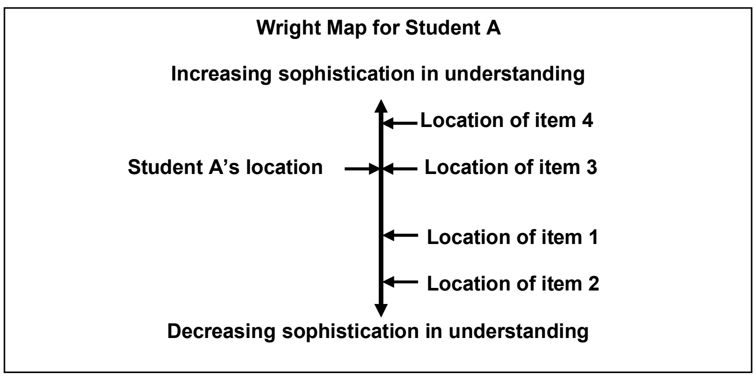


Figure 4. Wright Map showing alignment of person measures on the left with item measures on the right along a construct (cognitive or developmental domain).

The BEAR Assessment System includes software to facilitate the design of assessments and model respondent proficiencies. The software, called GradeMap (Kennedy, Wilson & Draney, 2005), produces maps to help teachers interpret student positions on the constructs of interest. One example, shown in Figure 5, is a Performance Map, which shows how a student has performed on a series of assessment activities over time. Note that the titles on the right edge of the map are brief descriptions of the levels from the construct map for the associated curriculum. In this example, the construct map for the *Designing and Constructing Investigations* construct differentiates five levels of performance: off target, incorrect, incomplete, correct, and beyond correct. Each point on the map shows the location of the student's most active learning at that point in time. This differs from the concept of *competence* in that each point represents the location on the construct where the student had a 50-50 chance of performing at that level or above.[2] In this example, the student's responses tended to be at the *Incomplete* level for activities 1 through 28, and then her responses become mostly *Correct* after that.

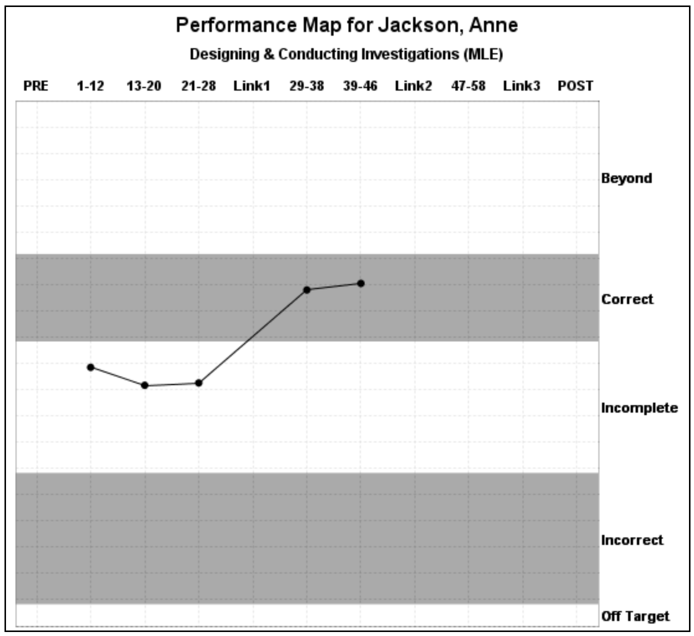
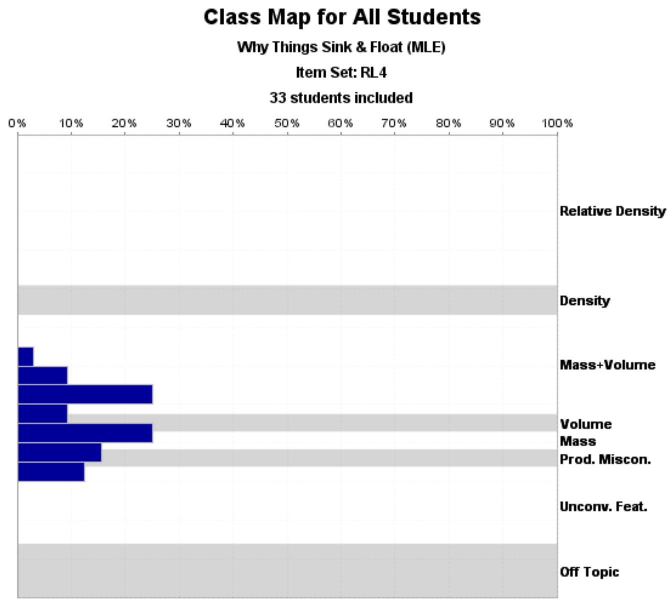


Figure 5. Performance Map for Anne Jackson produced by GradeMap showing maximum likelihood proficiency estimates on the "Designing and Conducting Investigations" construct at five time points.

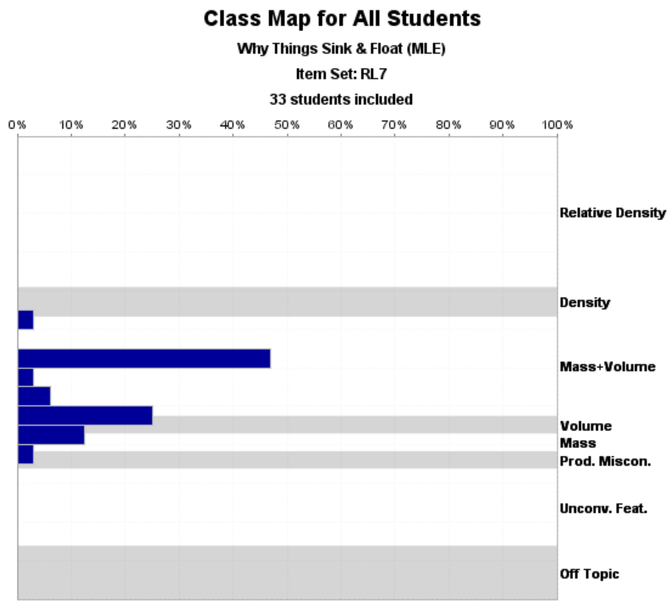
Teachers can also analyze the performance of a class as a whole with an output map called the Frequency Map (which has been named a Class Map in the example). In the example shown in Figure 6, taken from a curricular unit about buoyancy, a teacher can see that after completing a formative assessment at the fourth week of instruction, about 25% of the students' responses were below the targeted *Mass* level; that is, these

[2] Note that these figures can be recalculated to correspond to, say, an 80% probability of correct responses.

students were unable to explain how changing the mass of an object affects how it might sink or float.



Map of student proficiencies after 4th week.



Map of student proficiencies after 7th week.

Figure 6. Frequency maps of student proficiencies at the 4th and 7th weeks of the curriculum.

The teacher can then compare that map with a map of student proficiencies after the seventh week to see that the class has improved overall. Now, only a small percentage of students cannot explain how mass affects buoyancy, and well over 50% are able to explain that mass and volume work together to affect floating and sinking (i.e., at least 50% of the students had a proficiency level at or above the *Mass + Volume* level), which is the targeted learning objective during the seventh week of instruction. This type of information can help teachers focus classroom activities in ways that will be most beneficial to their particular students. For more complete examples of GradeMap reports and their interpretations, refer to the *GradeMap Users Guide* (Kennedy, 2005).

As these two examples demonstrate, the GradeMap software[3] aligns student proficiencies with the constructs targeted by a curriculum. This can improve the interpretability of student work and help teachers focus on the specific needs of their students *in the context of the developmental perspective of the curriculum*.

Assessments designed and developed within the BEAR Assessment System framework employ these principles and building blocks. This approach helps ensure that appropriate evidence is produced to draw reliable inferences about the student proficiencies of interest, and that those inferences can be interpreted in a straightforward way that is meaningful for teachers, students, and other educational stakeholders.

[3] See http://bearcenter.berkeley.edu/GradeMap/

References

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

Kennedy, C.A. (2005). *GradeMap v4.0 users guide*. University of California, BEAR Center: Berkeley. Available online at http://bearcenter.berkeley.edu/GradeMap/docs/UserGuide4_0.pdf

Kennedy, C.A., Wilson, R. and Draney, K. (2005). *GradeMap (version 4.1)* [computer program]. University of California, BEAR Center: Berkeley.

National Research Council. (2001). *Knowing what students know*. Committee on the foundations of assessment. J.W. Pellegrino, N. Chudowsky, R. Glaser (Eds.) Washington, D.C.: National Academy Press.

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

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