
Building Balanced Assessment Systems to Guide Educational Improvement

Doris Redfield, Ed Roeber & Rick Stiggins with contributions from Frank Philip

The 2008 National Conference on Student Assessment is about the appropriate and effective use of a variety of assessment processes, all of which should serve student learning. An important key to effective assessment practice is clearly communicating the purpose and intended use of the results to all those participating in or affected by an assessment program.

What makes an assessment appropriate for a particular use is determined by the alignment among the assessment purpose, assessment target, and the assessment process. Similarly, what makes an assessment system balanced is determined by the alignment of each component to the user's information needs such that the needs of all users are met. This concept of "balance" does not imply an equal number of tests of each kind or that each assessment or test should carry the same weight within the broader system.

The purpose of the opening panel of the 2008 National Conference on Student Assessment is to provide perspectives on the needs for, and functions of, assessment systems balanced to provide information needed at the student, classroom, school, district, state, and national levels to facilitate student learning. Other purposes are to define terms clearly and provoke lively conversation throughout the conference and beyond about the current status of assessment, where it needs to go, and how to achieve balanced, coherent systems that truly support student achievement.

Sometimes it is useful to consider the meaning of the words used to describe an entity or phenomenon. Here we consider the words behind the phenomenon of *building balanced assessment systems to guide educational improvement*.

- **Building** suggests a process or work in progress. It also suggests that the process may be iterative or unending, that the end result will consist of more than one component, and effort that is often team effort.
- **Balanced** does not necessarily mean equal or of equal weight. A balanced assessment system is a configuration of different assessment types and processes to fulfill purposes that align with expectations for learning and performance (Bazmore, Cippoletti, Howard, et al., 2008). It meets the needs of all users.
- **Assessment** can be used as a verb or a noun. When used as a verb, it connotes a process that draws upon a variety of tools to generate data to inform the formulation of conclusions, recommendations, instructional actions, policies, etc. When used as a noun, assessment may refer to a particular tool, such as a test.
- A dictionary definition of **systems** is "a regularly interacting or interdependent group of items . . . forming a unified whole" or . . . "network for serving a common purpose." Assessment systems are networks of the structures, supports, and tools that, together, meet the specified purpose or purposes for assessment. Hence, systems may differ from state-to-state and district-to-district, insofar as their purpose or purposes for assessment vary.
- **Guide** means to lead, direct, show the way, or influence action or thinking toward a particular end. In the assessment world, the "end" would be student learning – learning that may be demonstrated in a number of ways, at a number of levels, and for one or more purposes. An important requirement in an effective system is that the "way" be guided by purpose and use.

- **Educational** is a broad term that encompasses numerous aspects of the educational enterprise such as learning standards or expectations, curriculum frameworks and curriculum, instructional practices, assessment, pre-service training, and in-service professional development.
- **Improvement** refers to meaningful change in an important outcome and in a positive direction.

Guiding Principles

While building balanced systems to guide educational improvement may take on a number of different looks, the process should be guided by certain "building codes" or principles:

1. **Purpose(s)**. The purpose(s) of assessment need to be clear and clearly articulated for, and at, each level of the system. For example, there should be shared understandings of state and local assessment policies as well as the purposes of state, district, school, classroom, and individual student assessment practices. To be effective, assessment systems must be built around the reality that different users need different information in different forms and at different times in order to fulfill their decision making responsibilities. Balanced systems honor all information needs at classroom, building, district, state, and national levels.
2. **Assessment Adequacy**. The types of assessments included in the system should be appropriate and valid for meeting the specified purposes of each system component and the system as a whole. Assessment tools and processes used at each level and for each purpose within the system must be of high quality and adequate for the kinds of decisions to be made. For example, if results will be used to make decisions about individual or institutional rewards or sanctions, the results must be obtained using highly reliable tools and processes that yield highly reliable results. If the assessment will be used to guide day-to-day instructional decisions, then formal documentation of technical adequacy becomes less important; rather, fair and

accurate immediate feedback followed by instructional intervention is key. In all cases, the quality of an assessment tool or process is dependent upon the clarity and appropriateness of the achievement expectations to be assessed. These achievement targets must be clearly and completely defined. Otherwise, the results of the assessment will be undependable and may lead to counterproductive instructional or policy decisions.

3. **Communication of Results.** For balanced systems to serve productively, results must be communicated in a timely and understandable manner to the intended user(s).
4. **Supports.** Adequate supports need to be provided so that the purposes of the system can be met. For example, if a purpose is to provide teachers and students with real time information to guide instruction and learning, adequate professional development opportunities will need to be provided.

Shared Vocabulary

Much of the current buzz surrounding assessment concerns the idea and practice of summative, formative, interim, and benchmark assessments. However, not everyone means the same thing by these terms. Excerpts from *Speaking the Same Language* (Redfield, 2008) are provided as a basis for communication (and, no doubt, debate) throughout the course of the opening panel presentation at the 2008 National Conference on Student Assessment, the conference, and, hopefully, beyond. Ultimately, purpose and use determine assessment type.

Summative Assessment

The purpose of *summative* assessment is to “sum up” or describe what has been learned over time. Black and Wiliam (1998) and Stiggins, Arter, Chappuis et al. (2006) refer to summative assessments as assessments of learning. Summative assessment results are most often expressed as a number or score. The instructional time period covered by summative assessments may range from a single lesson to an entire course of study or even several courses of study. The results of summative assessments are usually used for accountability purposes such as assigning student grades, evaluating a teacher’s instruction, or evaluating a school’s success. *Accountability* assessments are summative assessments used to hold individuals or groups accountable for specified outcomes.

Summative assessments can be either

- **norm-referenced** (i.e., allowing for the comparison of a student’s performance or score to the performance or scores of a norm group of similar students) or
- **criterion-referenced or standards-based**, (i.e., allowing for the comparison of the student’s performance to a standard or criterion).

Formative Assessment

The purpose of formative assessment is to guide teaching and learning processes. Hence, formative assessments tend to include both formal and informal processes. Black and Wiliam (1998) and Stiggins, Arter, Chappuis et al. (2006) refer to a subset of formative assessments as assessments for learning. Assessments for learning occur during instruction, as part of the instructional process, and are not intended to be used for accountability purposes. Rather, they are intended to be used to provide teachers and students with corrective feedback and guidance relative to student learning needs.

Many formative assessments are selected or developed by teachers and used as part of on-going instructional practices. A definition of formative assessment as implemented by teachers in their classrooms (“classroom formative assessment”) that a number of assessment practitioners are finding useful is one offered by the Council of Chief State School Officers (2006): “Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students’ achievement of intended instructional outcomes.”

Benchmark or Interim Assessments

Interim or benchmark assessments are sometimes used summatively and sometimes formatively. While test publishers typically describe interim assessments as formative, such tests tend to be short-cycle summative tests. The purpose for administering the assessment and how the results are used determines whether it is formative or summative. A more detailed description of the purposes for interim benchmark assessments, as well as their uses, can be found in the Perie, Marion, & Gong (2007) paper on this topic.

References

- Bazemore, M., Cippoletti, B., Howard, T., Jones, L., Redfield, D.L., & Roberts, M. (2008, June). Connecting the dots: Balanced comprehensive assessment systems. CCSSO’s National Conference on Student Assessment, Orlando, FL.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Educational Assessment: Principles, Policy and Practice*, 5(1), 7-74. Also summarized in an article entitled, Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Perie, M., Marion, S, & Gong, B. (2007). A framework for considering interim assessments. National Center for the Improvement of Educational Assessment, Dover, NH.
- Redfield, D.L. (2008). Speaking the same language. Assessment Academy Website (contact author – doris.redfield@edvantia.org for access).
- Stiggins, R.J., Arter, J.A., Chappuis, J., & Chappuis, S. (2006). Classroom assessment for student learning: Doing it right – using it well. Educational Testing Service, Portland, OR.