

# What does “balance” mean in a balanced assessment system?

Stuart Kahl  
RMC Research Corporation

[Note: This document appeared as a blog post on 1/3/2019 on the RMC website.]

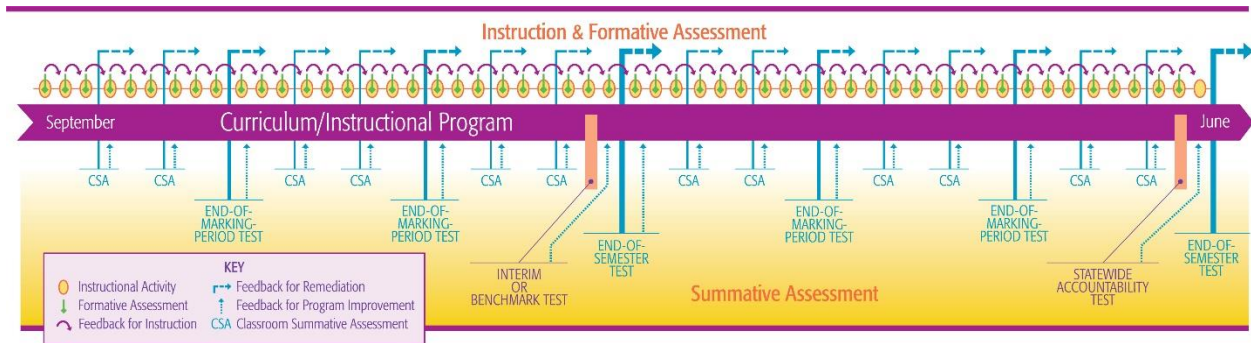
It’s fitting that an early post in this blog is a piece on “balanced assessment systems” because it’s an all encompassing concept relative to the assessments experienced by students. We can zero in on specifics throughout the year. The term has become a widely used one, yet when folks are asked to define it, the answers are quite variable. I won’t hazard a guess as to the origin of the term, but I can say with certainty that a few years after researchers Black and William (yes, spelled correctly with one “l”) published a 1998 landmark piece on formative assessment in *Education Leadership*, the concept of a balanced assessment system was given significantly increased attention.

Oftentimes, people will describe a balanced assessment system by listing components – e.g., formative, interim, and summative assessment. This doesn’t really address the idea of “balance,” which gets at the notion of appropriate use/emphasis. It’s difficult to discuss relative emphasis on *formative assessment* because the formative assessment that research says is so effective is really a multi-step instructional process that is ongoing and in which just one step is gathering evidence of learning. Unfortunately, some have taken the term to mean “frequent testing.” It is not. Most student work gathered or observed for formative purposes should not count toward students’ course grades since it is to be gathered “during the learning” -- before the students have reached the level of attainment (relative to specific learning targets) they will reach upon completion of an instructional unit. I’ll leave the topic of grading practices relative to formative assessment to later blog posts.

I like to distinguish among four components of a balanced assessment system:

- classroom formative assessment (described above),
- classroom summative assessments,
- external interim and benchmark assessments,
- external summative assessments (e.g., state assessments).

Other critical components of the system are the curriculum and instruction, which are derived from content standards, which also define the content of the measurement instruments.



Copyright © 2006 Measured Progress, Inc. Reproduced with permission .

I used the graphic above in several presentations in 2006 and 2007. I actually violated one of my own principles in it – the down arrow in the key should be labeled “Formative Assessment Evidence Gathering” since it is referring to just that one step within the instructional process of formative assessment.

**Classroom summative assessments**, such as end-of-unit or end-of-marking-period tests, count toward grades. **External assessments** are those involving instruments developed or purchased by the district or state. Different kinds of external tests can be administered during the course of the school year, and people often use the names of these “through-course” tests interchangeably. I use **interim** to refer to assessments that use general achievement measures covering a full year’s course content in order to monitor growth or provide early warning regarding students or curricular areas needing additional attention prior to end-of-year **summative** testing that covers the same material. **Benchmark assessments** cover recently taught material. Since interim and benchmark tests are administered after the learning, they are really **summative** measures, which document or certify learning. In fact, teachers’ end-of-unit or end-of-marking period tests are benchmark tests and are used for summative purposes...hence the name “classroom summative.”

#### Formative vs Summative -- Illustrative Examples

Formative Assessment	Summative Assessment
A chef tasting soup and deciding it needs more salt	A restaurant guest tasting the soup and saying, “Wow, this soup is good!”
A student driver being quizzed by a friend on the contents of the driver’s manual and a driving instructor observing the student’s performance behind the wheel and providing feedback	A driver’s license candidate taking the written and performance components of the state’s driver’s test to demonstrate the knowledge and skills needed to earn a license

At a large conference in San Francisco in 2006, I accused publishers of becoming an obstacle to improved instructional practice by pirating the term “formative” and putting it in front of the name of every item in their catalogs. The commercial off-the-shelf products at that time were really interim or benchmark tests used for summative purposes – after the learning. I was

concerned that school administrators not familiar with the process of formative assessment but hearing of its effectiveness would purchase them thinking their mere use would improve student performance. Of course, as higher stakes were attached to the results of state tests, external interim and benchmark tests were used more and more in preparation for year-end summative testing and, according to research, became the major cause of over testing (an imbalance), which so rattled teachers, students, and parents.

As long as I'm dealing with vocabulary, I'll try to clarify two more terms, even though I find it difficult to avoid using them interchangeably. **Assessment** is a process, and a **test** is a tool used in an assessment. Oftentimes, either term works. Enough.....

Some final points I want to make regarding balanced assessment systems pertain to other ways I believe assessments (or tests) need to be balanced. I believe there should be a balance struck between attention to foundational or basic knowledge/skills and attention to deeper learning (higher order thinking or the application of the basics). The latter is so often shortchanged in both local and state testing that has been dominated for some time by the multiple-choice format. (The popularity of online testing and the immediacy of its results has exacerbated the problem.) This balance would be related to a balance of item or task types in tests and to depth of knowledge, one of multiple aspects of test alignment to curriculum standards advanced by Norman Webb of the University of Wisconsin. State assessment instruments are subjected to alignment studies that address depth of knowledge as well as other types of alignment to standards, including balance of representation and range of knowledge. All these alignments deal with the adequacy of coverage of the content domain/standards being assessed. Finally, regarding federally mandated state assessments, I believe there should be a balance between state external and local testing components, both of which should contribute significantly to school accountability data, the latter including curriculum-embedded performance measures.

For a more in-depth description of components and characteristics of a balanced assessment system, one might read the 2010 commissioned paper by Brian Gong of the National Center for the Improvement of Educational Assessment, "Using Balanced Assessment Systems to Improve Student Learning and School Capacity: An Introduction."

This early blog post covers a lot of ground superficially, but I'm hoping it stimulates lots of comments and questions. It just scratches the surface of topics such as grading practices, test types and purposes, test quality, item formats, over testing, online testing, and many more. I look forward to hearing from you and discussing many assessment-related topics in greater depth.