

Running Head: Formative Assessments

Let's Talk Formative Assessment...*and* Evaluation?

May 26, 2009

Karee E. Dunn, Ph.D.

University of Arkansas

and

Sean W. Mulvenon, Ph.D.

University of Arkansas

Address Correspondence to:

Karee E. Dunn  
248 Grad Ed  
College of Education  
University of Arkansas  
Fayetteville, AR 72701  
Phone: (479) 575-5593  
Email: [kedunn@uark.edu](mailto:kedunn@uark.edu)

## Abstract

The implementation of No Child Left Behind (NCLB) in 2002 and subsequent sanctions for lower performing school systems has resulted in a number of both positive and negative consequences. One negative effect of NCLB is that teachers report a great deal of anxiety with regard to state-mandated assessments. The authors' propose that one cause for the fear and anxiety teachers experience with regard to high-stakes state-mandated summative exams is the inability of educational researchers and policy makers to supply teachers with assessment-related language to support their instructional efforts. The design of research that supports an effective model for the use of state-mandated assessment outcomes should extend beyond NCLB related sanctions and provide useful direction to educators. The purpose of this article is to highlight the absence of a sound and usable vocabulary related to summative and formative assessment models for classroom educators. Moreover, the authors describe a new way of defining formative and summative assessment in which assessment is separated from the evaluation of assessment outcomes (what we call formative and summative evaluation) as well as a practical example of how this terminology can be used to unlock the power of high-stakes tests.

### Let's Talk Formative Assessment...*and* Evaluation?

In the wake of NCLB, most teachers are experiencing high anxiety in the high-stakes testing era. However, the purpose of NCLB was not to inject fear into teachers; the purpose was instead to inject data-driven decision making into schools. A possible cause for the fear and anxiety teachers experience with regard to high-stakes mandated summative exams is rooted in the failure of assessment-related language and research to provide an effective model for improving teaching and learning through the use of state-mandated assessment data. As Sir Francis Bacon stated long ago, “knowledge is power.” Many teachers do not feel empowered when dealing with assessment issues as there is a glaring absence of understanding in both the classroom and the literature with regard to how to fully use the power of both summative and formative assessments in education. The purpose of this article is to highlight the absence of a sound and usable vocabulary related to summative and formative assessment, and to describe a new way of defining formative and summative assessment in which assessment is separated from the evaluation of assessment outcomes — formative and summative evaluation.

As part of a series of studies being designed to evaluate the assessment and methodological practices used in formative assessment, a review of the literature revealed limited empirical evidence demonstrating that the use of formative assessments in the classroom directly resulted in marked changes in educational outcomes (Dunn & Mulvenon, 2009). Basically, what began as a perfunctory review of literature on formative assessments for a manuscript on statistical methods, evolved into a mission to clean up the “messy” area of formative and summative assessment. It is important for teachers, administrators, researchers, and policy makers to share a common language related to assessment so as to unlock the power of assessment and create positive changes in student achievement.

## Review of the Literature

Over the past several years, a growing emphasis on the use of formative assessment has emerged, yet formative assessment has remained an enigma in the literature (Dunn & Mulvenon, 2009; Leung & Mohan, 2004). When reading formative assessment literature and focusing on the issue of solidifying a definition of the term, an interesting and problematic theme arose. Formative assessment and its various manifestations were defined not only by inherent characteristics, but also by the use of the assessment outcomes. Formative assessment's status as an ethereal construct has been perpetuated in the literature due to the lack of an agreed upon definition. The vagueness of the definition directly contributes to the weaknesses found in the related research and dearth of empirical evidence identifying best practices related to formative assessment (Dunn & Mulvenon, 2009). Without a clear understanding of *what* is being studied, sound empirical evidence supporting best formative practices will more than likely remain in short supply.

For example, Black and Wiliam (1998) defined formative assessment as “all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (p. 10). The Council of Chief State School Officers (CCSSO) defined formative assessment differently according to the specifications provided by Formative Assessment for Students and Teachers (FAST), a department of CCSSO. FAST defined formative assessment as a process used during instruction to provide feedback for the adjustment of ongoing teaching and learning for the purposes of improving student achievement related to instructional objectives (Melmer, Burmaster, & James, 2008). In 2006, Popham stated that an assessment is formative to the degree that the information collected from the assessment is used during the assessed

instructional period to improve instruction so as to meet the needs of the students assessed. In 2008, Popham defined formative assessment as a planned process during which the teacher or students use assessment-based evidence to adjust ongoing learning and instruction. Without any inter- or intra-individual consensus as to what the term formative assessment means, it is difficult to have a well-formed body of research.

To further complicate the issue of defining formative assessment based upon the assessment itself as well as the use of evidence from the assessment, formative assessments serve a myriad of feedback-related purposes such as diagnosis, prediction, and evaluation of teacher and student performance (Black & Wiliam, 1998). For example, Perie, Marion, and Gong (2007) argue that assessment issues can be clarified if assessment is defined by its purpose. From this perspective, *formative assessment* is defined as assessment used by teachers and students to adjust teaching and learning, as compared to *interim assessment* that informs policymakers or educators at the classroom, school, or district level about student achievement levels and curriculum effectiveness. Defining assessments in this fashion may create confusion for consumers of assessment products and literature. For example, one assessment could be used by students and teachers to inform the learning process as well as by administrators to create policy changes, essentially contradicting this definition.

Moreover, a great deal of assessment literature is aimed at delineating between formative and summative assessment, yet summative assessment can be used for formative purposes and vice versa (Bell & Cowie, 2000). It is important to note that we acknowledge that the purpose for which any assessment is developed and validated is an important aspect of assessment. However, a test that was designed to give formative feedback is only formative if the teacher uses it to

provide feedback for the student. If the teacher only uses the formative assessment to provide a grade, is that assessment still formative?

According to the definitions presented above, the mere assessment of performance into a grade category (i.e., “A” or “B”) should be labeled formative because it provides feedback related to the learning progress of the student and may be used for future instructional interventions. For example, if a student takes a summative test on fractions and does well overall, but misses every item related to the division of fractions, is the teacher only supposed to use the assessment data to assign a grade? Clearly this would not be appropriate. But, if the teacher uses that data to provide the learner feedback and create instructional interventions, is the test now formative and not summative even though it was used to assign a grade? Most literature categorizes any assessment used to assign grades as summative. This example illustrates the issues that arise when the use of the data that results from an assessment is incorporated into the definition of the design of the assessment.

Although an assessment may be designed and packaged as a formative or summative assessment, it is the actual methodology, data analysis, and use of the results that determine whether an assessment is formatively or summatively evaluated. For example, Wininger (2005) used a summative assessment as a formative assessment by providing both quantitative and qualitative feedback about the results of the examination. He called this formative summative assessment. Winniger’s discussion exemplifies the complications that arise when one defines an assessment by its usage. Thus, we propose that it is necessary to clarify the lexicon and literature related to formative assessment so as to delineate between the design of an assessment and the evaluation of the resultant data (use of the outcome). An assessment is an assessment, and the manner in which an assessment is evaluated or used is a related but separate issue.

The concept of setting assessment apart, as something unique from evaluation, researchers and educational stakeholders alike may begin to speak the same language related to the usage of these assessments and produce research that results in more powerful academic outcomes. For example, a hammer is a hammer regardless of how it is used. If a hammer were defined by its use, it would make the discussion of the tool much more difficult in the remodeling of a home. It is easier to simply ask for and receive a hammer than to provide a description of the intended use (i.e., If you ask for an item that can make a hole in the wall, you might receive a sledge hammer in lieu of a hammer).

By separating the design of an assessment from the evaluation (use) of its results, the authors' perspective harkens back to Scriven's (1967) original presentation of formative evaluation. Scriven described formative evaluation as the evaluation of an ongoing and malleable educational program. It was Bloom (1969) who attempted to transfer the term formative from *evaluation* to *assessment*. Perhaps this is where an understanding of the process of defining formative assessments first became problematic.

The authors argue that defining formative assessment as a test and formative evaluation as the specific use of assessment data (be it formative or summative data) is more amenable to both classroom application and academic discourse. Thus, the authors proposed that formative or summative assessment data may undergo formative or summative evaluation. The purpose of this manuscript was to provide a clear and more user-friendly lexicon related to formative and summative assessment by carefully examining and separating the issue of assessment from the issue of assessment outcome evaluation.

### Separating Evaluation from Assessment

In this section, the authors reintroduce and redefine formative evaluation as well as separate the issue of assessment from the issue of evaluation of assessment-based data. The authors define summative and formative assessment as well as summative and formative evaluation of assessment-based data. Summative assessments are those assessments designed to determine a students' academic development *after* a set unit of material (i.e., assessment *of* learning) (Stiggins, 2002). Formative assessments are assessments designed to monitor student progress *during* the learning process (i.e., assessment *for* learning) (Chappuis & Stiggins, 2002).

Although assessments may be designed for formative or summative purposes, the authors argue that resultant data may be interpreted either formatively or summatively. The authors further argue that the inclusion of how the data is used is problematic and results in issues in the literature due to the possibility of evaluating either type of assessment data formatively or summatively (i.e., summative data may be formatively evaluated). For example, weekly quizzes may be designed to provide a grade, but they may also be used to assess students' learning progress within a given instructional unit. The authors define the terms formative evaluation and summative evaluation in terms of the use of assessment data; thereby, separating the issue of assessment design from assessment use.

For our purposes, summative evaluation is defined as the evaluation of assessment based data for the purposes of assessing academic progress at the end of a specified time period (i.e., a unit of material or an entire school year) for the purposes of establishing a students academic standing relative to some established criterion (i.e., state mandated assessments for NCLB).

Formative evaluation is defined as the evaluation of assessment-based evidence for the purposes



of providing feedback to and informing teachers, students, and educational stakeholders about the teaching and learning process.

Formative evaluation also informs policy, which then affects future evaluation practices, teachers, and students. The reciprocal relationship between policy and formative assessment is graphically represented by the Key Model for Academic Success (See Figure 1). This model supports Shepard's (2000) assertion that it is not necessary to separate assessment from teaching; instead, teaching practices can and should be informed by and coincide with assessment practices and outcomes.

---

Insert Figure 1 About Here

---

Having defined what is meant by “formative evaluation,” it is important to further separate the issue of assessment from the issue of formative or summative evaluation. In doing so, we hope to provide a more clearly defined nomenclature to frame the investigation of both effective implementation of formative evaluation as well as the effect of formative evaluation on student performance. Assessments are instruments for collecting information, in this case information about students' academic performance, including progress with regard to specific materials. Evaluation is a separate, but related issue that has to do with the use of assessment-based data. Although an assessment may be designed to be formative or summative, the data acquired by the administration of either type of assessment may be used for formative and/or summative purposes. In other words, formative evaluation or summative evaluation may be applied to either formative or summative assessment data. What arises from the application of

this assessment lexicon is a practical model for a system of assessment and evaluation (See Figure 2).

---

Insert Figure 2 About Here

---

### How This Lexicon Can Support Classroom Teachers

One potential use of this lexicon for the development of formative evaluation models that utilize summative assessment data is currently under construction at the National Office for Research on Measurement and Evaluation Systems (NORMES)<sup>1</sup> at the University of Arkansas. By utilizing existing and accessible data from the Arkansas mandated Benchmark exam, NORMES faculty and staff employ measurement theory and statistical modeling to develop detailed reports of gaps in student knowledge at various levels within school systems. State mandated tests are not designed to be diagnostic at the classroom or student level; however, they may provide global diagnostic information that can be used to help improve instruction. As Gierl (2007) more succinctly stated, “cognitively-based microlevel analysis of examinees’ skills and misconceptions...has implications for formative assessment specifically, and educational improvement more generally” (p.325).

Through the use of such information, school administrators and teachers can evaluate and create curriculum plans. The formative evaluation of summative Benchmark exam data identifies conceptual gaps exhibited by groups of students. Thus, revisions and additions to existing instructional curriculum can be made to ameliorate learning gaps for future students. Moreover,

teachers, parents, and students will be able to better understand areas of weakness for individual students and plan accordingly for the academic success of each and every child. Essentially, through the application of formative evaluation techniques to summative assessment data, NORMES is developing tools that will better equip educational practitioners to design curriculum programs that support the success of entire school systems as well as individual students.

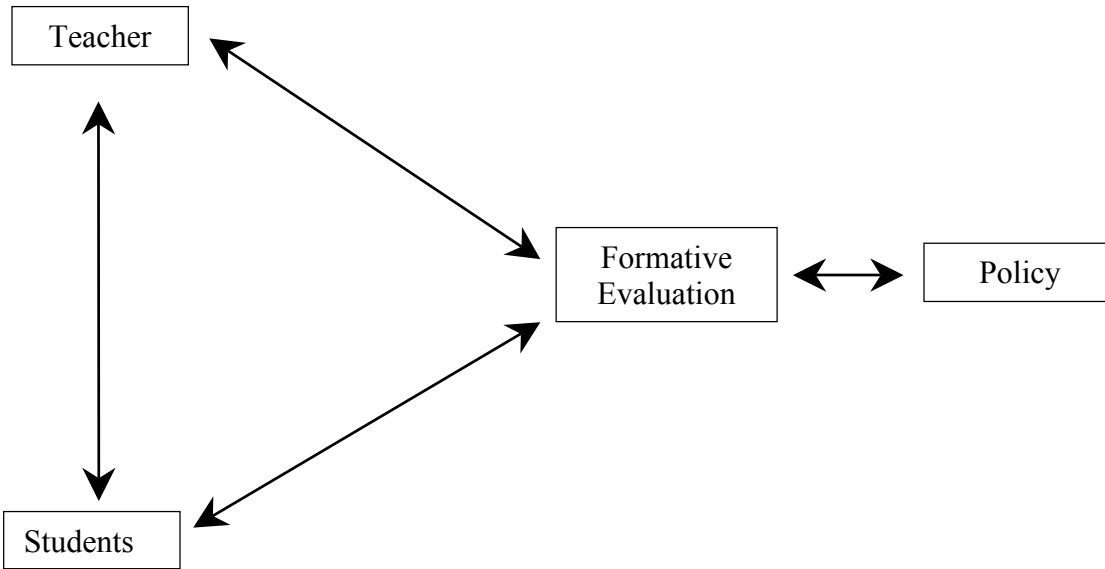
### Conclusion

Formative evaluation of assessment outcomes has numerous possibilities for positively impacting teachers and learners. By communicating in the shared language of “formative evaluation,” educational stakeholders and practitioners may begin to view high-stakes testing in a more positive light and the dream of NCLB may be one step closer to success. Within the context of formative evaluation, end-of-year benchmark exams can become something other than a dreaded exercise in futility for teachers and administrators. These exams can become the curriculum maps for future school improvement. Furthermore, when researchers, teachers, and administrators alike begin to speak the same language it will result in improved research models that can be used to help design more effective summative and formative assessments for summative and formative evaluation. More importantly, formative evaluation of all assessments can truly lead to effective data-driven decision-making that leads to real changes in achievement for today’s children.

## References

- Bell, B., & Cowie, B. (2000). The characteristics of formative assessment in science education. *Science Education, 85*, 536–553.
- Black, P., & William, D. (2003). ‘In praise of educational research’: Formative assessment. *British Educational Research Journal, 29*(5), 623-637.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Taylor (Ed.), *Educational evaluation: New roles, new means: The 68<sup>th</sup> yearbook of the National Society for the Study of Evaluation, Part II* (pp. 26-50). Chicago: University of Chicago Press.
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership, 60*(1), 40-43.
- Dunn, K. E., & Mulvenon, S. W. (2009). A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence of the Impact of Formative Assessments in Education. *Practical Assessment, Research, and Evaluation, 14*. Retrieved May 26, 2009, from <http://pareonline.net/genpare.asp?wh=0&abt=14>
- Gierl, M. J. (2007). Making diagnostic inferences about cognitive attributes using the rule-space model and attribute hierarchy method. *Journal of Educational Measurement, 44*(4), 325-340.
- Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing, 21*(3), 335-359.
- Melmer, R., Burmaster, E., & James, T. K. (2008). *Attributes of effective formative assessment*. Washington, DC: Council of Chief State School Officers. Retrieved October 7, 2008, from <http://www.ccsso.org/publications/details.cfm?PublicationID=362>

- Perie, M., Marion, S., & Gong, B. (2007). The role of interim assessments in a comprehensive assessment system: A policy brief. Retrieved October 1, 2008 from <http://www.nciea.org/publications/PolicyBriefFINAL.pdf>
- Popham, W. J. (October 2006). *Defining and enhancing formative assessment*. Paper presented at the Annual Large-Scale Assessment Conference, Council of Chief State School Officers, San Francisco, CA.
- Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, and M. Scriven (Eds.), *Perspectives of curriculum evaluation, Volume I* (pp. 39-83). Chicago, IL: Rand McNally.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Stiggins, R. J. (2002). Assessment Crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, 83(10), 758-765.
- Wininger, R. S. (2005). Using your tests to teach: Formative summative assessment. *Teaching Psychology*, 32(2), 164-166.



*Figure 1.* Key Model for Academic Success (Dunn & Mulvenon, 2009)

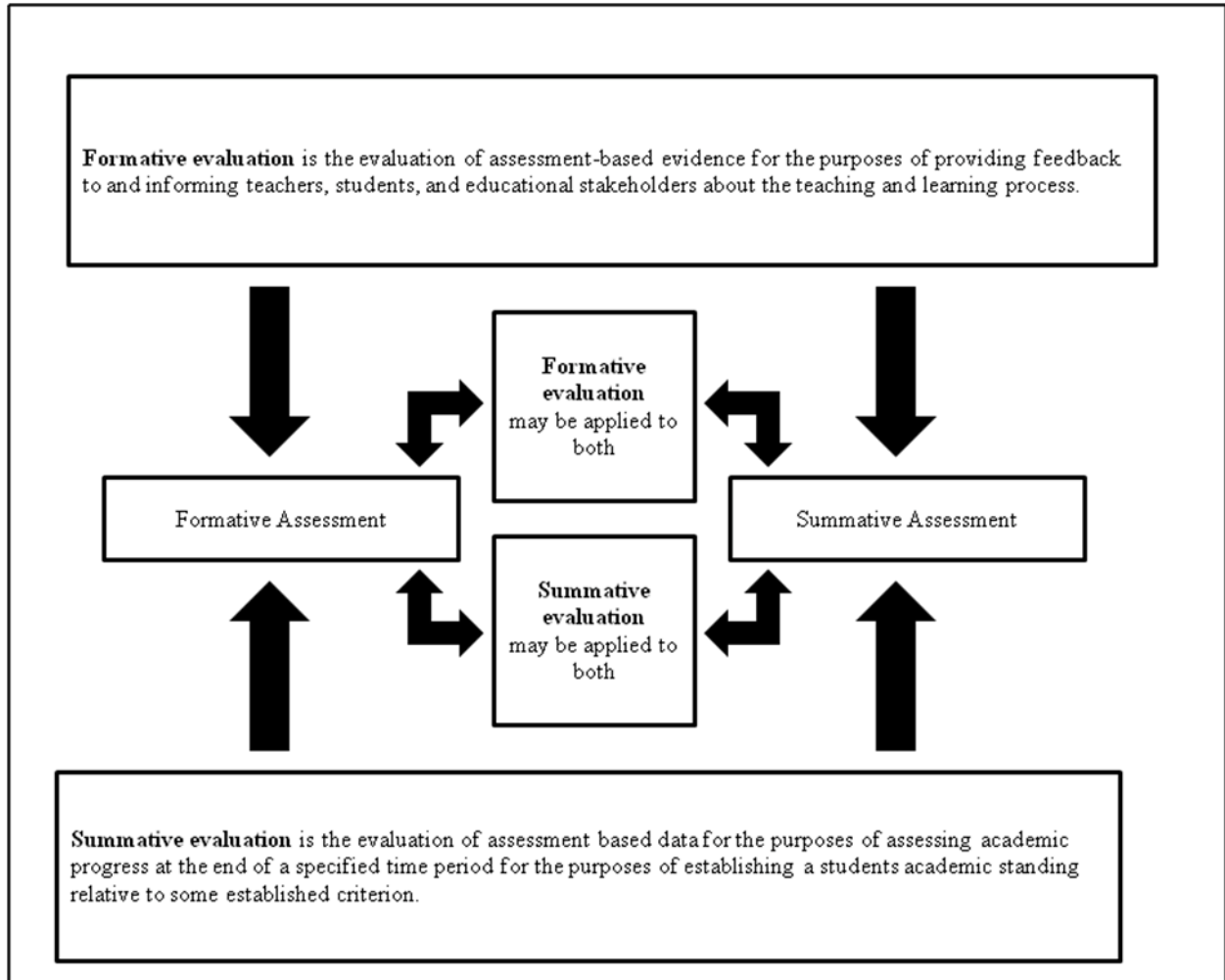


Figure 2. Practical Model of Assessment and Evaluation System (Dunn & Mulvenon, 2009)

## Footnotes

<sup>1</sup>The partnership between the University of Arkansas and the Arkansas Department of Education is facilitated by the National Office of Research on Measurement and Evaluation Systems (See NORMES homepage: <http://normesapps.uark.edu/newsite/>). NORMES is a research office and part of the University of Arkansas. The goal of NORMES is to provide a readily accessible source for reliable educational achievement data in Arkansas for educators, policymakers, parents and other educational stakeholders, and to answer questions related to public school systems. NORMES seeks to extend the current best practices model of a student-centered assessment system developed at the Office of Research, Measurement and Evaluation (ORME) for use in a national center for schools.