

## Formative assessment: Stalled by too few right-size reports

William James Popham <sup>1,\*</sup>

<sup>1</sup>University of California, Graduate School of Education and Information Studies, Los Angeles, USA

### ARTICLE HISTORY

Received: Jun. 16, 2022

Accepted: Aug. 10, 2022

### Keywords:

Assessment,  
Formative assessment,  
Reporting,  
Right size reporting.

**Abstract:** After providing key definitions well as substantial supportive evidence for the instructional process under consideration, this analysis identifies a serious shortcoming in the way that many U.S. educators are currently encouraging teachers' adoption of the formative-assessment process--a teaching approach informed by students' en route test performances during instruction. After identifying the basics of the formative assessment process, and the manner in which reports of students' en route assessment performances should be built, it is claimed that formative assessment will attain its much-lauded learning payoffs only when short reports, easily used by both teachers and students, are routinely employed.

## 1. INTRODUCTION

Formative assessment works. When teachers routinely test their students' emerging mastery of instructionally emphasized knowledge or subskills, then make any needed adjustments based on students' measured progress, such adjustments typically pay off. Indeed, classroom formative assessment might well represent education's most successful example of properly conceived ends-means thinking. When en route tests indicate that a teacher's instructional procedures (the means) aren't satisfactorily moving students toward mastery of designated curricular aims (the ends), then different instructional tactics are employed.

### 1.1. In Praise of Formative Assessment

Not only is formative assessment a potent analytically identifiable instructional strategy, but for more than 25 years, we have possessed heaps of hard, empirical evidence indicating that formative assessment works—and works well.

In 1998, Paul Black and Dylan Wiliam, two British researchers, published a comprehensive review of almost 10 years' worth of empirical research dealing with classroom assessment (Black and Wiliam, 1998). The following conclusion from their in-depth review succinctly sums up that analysis: "The research reported here shows conclusively that formative assessment does improve learning (Black and Wiliam, 1998)."

---

\*Corresponding Author: W. James POPHAM ✉ [wpopham@ucla.edu](mailto:wpopham@ucla.edu) 📍 University of California, Graduate School of Education and Information Studies, Los Angeles, USA

---

But were those demonstrable learning improvements substantial or, perhaps, merely modest? Well, Black and Wiliam (1998) concluded from a welter of empirical studies that the student gains in learning triggered by formative assessment were “amongst the largest ever reported for educational interventions.” That’s high praise indeed.

Meanwhile, in the U.S., the enactment of the federal No Child Left Behind Act of 2002 called for American educators to promote substantial—and definitely measurable—improvements in their students’ performances. Circumspect educators realized that formative assessment might well be the chief means for promoting the federal call for improved student test scores. Summing up, then, more than a decade’s worth of research focused on instructional uses of students’ testing—plus a congressional mandate demanding higher test scores—meant that many American teachers seemed nearly certain to hear about, and even install this relatively new assessment-rooted strategy advocated by early proponents of formative assessment.

By the late 1990s, then, many American educators began learning about the nature of formative assessment, while also encountering numerous research reports indicating that formative assessment was a sure-fire instructional winner. In addition, a major federal law had been enacted urging American educators to employ instructional tactics capable of bringing about substantial improvements in their students’ test scores.

Rarely, in the history of U.S. public schooling, have research evidence and legislative demands meshed so fortuitously. Educational commentators of that era often opined that formative assessment, an empirically demonstrable assessment-rooted strategy for instructional improvement, would soon be seen in most American schools.

Yet, many of those optimistic prophecies were issues more than 20 years ago. And, although it was widely believed back then that formative assessment would be installed in many U.S. schools, this oft-voiced prophecy simply failed to flower. What went wrong?

The following analysis will identify one repairable shortcoming in our thinking about formative assessment that represents an important reason formative assessment, apart from the few early years of interest it drew from U.S. educators, has fallen far short of widely foreseen usage hopes. It will be argued that if this single shortcoming were to be rectified, the long-promised learning dividends of properly formulated formative assessment will have a far better chance of being realized.

## **1.2. Definitions: A Pair**

Although, these days, most educators possess a general notion of what formative assessment is, and many of those educators understand that this strategy represents a measurement-spurred instructional approach, it is always useful to define the central focus of any analytic commentary. Accordingly, then, bedecked in boldfaced italics below, is a formal definition of what most educators mean these days when they employ the descriptor “formative assessment:”

Formative assessment, an ongoing process seeking intermittent evidence of students’ emerging learning, is used by teachers to adjust their instructional procedures and/or by students to adjust their current learning tactics.

But there’s one more label that needs defining. This is because what’s being defined above overlooks a missing ingredient in most formative-assessment dissemination strategies. It is an ingredient that, if lacking, decisively limits the expanding implementation of formative assessment. At least in the U.S., regrettably, formative assessment is rarely accompanied by “right-size reporting.” So, in a bow to even-handed definitions, what’s meant by “right-size is presented below—predictably, in boldfaced italics.

Right-size reporting describes efficient methods of describing students' test performances so that report-users can easily arrive at defensible decisions regarding next-step instructional actions consonant with the test's intended use.

The advocacy of right-size reporting's use during formative assessment usually stems from a belief that the more teachers who employ formative assessment, the better taught will be those teachers' students and, therefore, the better those students will learn. Although properly conceived formative assessment can be employed by students as well as teachers, the following remarks apply chiefly to teachers' needs for right-size reporting.

Nonetheless, ask any teacher who has made a serious commitment to employing formative assessment for an extended period to comment on that experience. What you'll often hear in response from the teacher is that (1) the formative-assessment process was effective and (2) it required too much work from the teacher to frequently implement it. We are not surprised by the "effective" response, of course, but the "too much work" replies often come as a surprise. Yet, when we think hard about the most distinctive feature of the entire formative-assessment process, it is the use of ongoing tests to collect evidence indicating whether instructional modifications are needed.

Although, depending on the curricula aims being pursued and, of course, the particular students being taught, teachers typically determine how often to measure their students' progress. Typically, there will be one or two short-duration assessments (called en route tests) used during a week or so of instruction. If the instructional period at hand is at all lengthy, for instance, five or six weeks, this quickly translates into a hefty number of en route tests that must be administered, scored, and then employed to arrive at appropriate instructional decisions regarding instructional next steps. Where do those tests come from?

Putting aside for the moment the who-creates and who-scores issues, what attributes should en route tests possess if they provide right-size reports and, therefore, contribute to improved instructional decisions by teachers? Here, then, are three features that, if present, optimize the instructional contributions of en route tests employed during the formative-assessment process.

- **Balanced Representation.** The evidence reflecting the content, i.e., the knowledge and/or skills assessed, provides an accurate representation of this content.
- **Suitable Numbers of Items.** For whatever knowledge and/or skills are tested, sufficient but not excessive numbers of items are present.
- **Actionability.** Content of each item on a formative assessment's en route test, depending on a student's responses, suggests next step(s) for teachers.

Let's briefly consider these three attributes of the en route tests used during formative assessment because, as we will see, the pressures on teachers to incorporate truly first-rate tests have, surprisingly, led many educators to completely abandon use of the formative-assessment process. More about this shortly.

### **1.2.1. Content representativeness**

First, students' responses to formative assessments' en route tests supply teachers with the evidence needed to make any necessary adjustments in ongoing instruction. Such evidence is, arguably, the essence of the formative-assessment process. It is clearly necessary, therefore, for formative assessment's en route tests to be accompanied by evidence, perhaps judgmental in nature, indicating the degree to which students' responses to a test's items will provide a sufficiently representative reflection of students' status regarding the en route targets being sought. Accordingly, credible evidence of some sort—perhaps gathered from a teacher's colleagues—should be routinely provided to indicate the representativeness of an en route test's items.

Through the last few decades, many teacher-review panels have been employed to judge the adequacy of test-items' content for the intended use of test's results—particularly for high-stakes tests. We have learned that formatively focused en route tests should also, if possible, have their items reviewed for content representativeness. For any particularly important en route tests, a suitable rating form plus a systematic orientation for reviewers' use is normally required. For less significant tests, a content-representativeness judgment from one or two content-knowledgeable colleagues is often sufficient.

What is being recommended here is that, whenever feasible, the content representativeness of formative assessment's tests be determined so that the teacher (as decision-maker) can determine how much confidence should be based on the evidence garnered by different en route tests.

### **1.2.2. Item numbers**

One of the most vexing requirements facing teachers who use formative assessment hinges on a seemingly small problem, namely, how many items to employ in a teacher's en route tests. The potential mistakes made here are usually "too few" or "too many." If too few items are employed in an en route test, then it is unlikely that teachers can draw a valid inference about a test-taker's mastery of the content represented by such a tiny collection of a test's items. Conversely, if far too many items are included in en route tests, then students' performances on those tests may, in fact, accurately reflect students' content mastery but, because of an excessive number of en route items, such along-the-way testing takes far too much time—time that might otherwise be profitably spent on instruction.

### **1.2.3. Actionability**

Teachers engage in formative assessment to help them discern whether instructional changes should be enacted and, if so, to decide which changes to make. In some instances, of course, students' performances will indicate that no instructional modifications are needed—because the teacher's students are learning wonderfully. But if some students' less than lustrous en route performances make it clear that instructional alterations are necessary, the teacher must then identify what instructional changes to make and determine when to make them.

Realistically, there are three main options to consider when dealing with next-step options based on the rests used as part of the formative-assessment strategy. First, if teachers are working alone, then such teachers will need to come up with—on their own—one or more next-step instructional options. Second, a group of teachers working collaboratively in the same school—or affiliated with the same school district—could also devise a set of potential instructional activities for students that would address test-isolated content or subskills in need of an instructional re-do. But if you were to ask many teachers who have taken part in such collaborative test-building to comment on such endeavors, you're almost certain to learn that these sorts of collaborative instruction-building efforts are, obviously, dependent on the individuals involved, particularly time-consuming and, often, not all that effective.

The third major source of potential next-step instructional activities are the many commercial products now being sold by both profit-making and non-profit organizations. Spurred often by the positive results from long-term formative assessment, numerous formatively oriented systems are currently being marketed so that both en route tests, as well as suggested instructional alternatives aimed at such targets, are now purchasable. However, in the attempts of commercial vendors to market their products to a sufficiently large and heterogeneous array of potential purchasers, almost all such purveyors of these sorts of ready-made materials for formative-assessment materials are obliged to create materials far too general to satisfy the needed accuracy of truly on-target instruction. The effect of educators' adopting such too-

general instructional and/or assessment materials is that much of today's en route testing—and subsequent instructional amelioration—leads to far less successful learning than hoped.

### 1.3. Formative Assessment: An Obstacle and A Solution

Summing up, then, although we now possess ample evidence from many quarters indicating that formative assessment, a potent marriage of ends/means assessment and instruction, is capable of producing substantial improvements in students' learning (Black and Wiliam, 1998) we see far less real-world usage of classroom formative assessment than had been widely prophesied (Popham, 2008, 2011). It was claimed in this analysis that a prominent deterrent to teachers' expanded employment of classroom formative assessment was that its implementation requires a raft of classroom formative practices that are simply too difficult for most teachers to undertake.

Fashioning en route tests so that they do not require Herculean efforts to employ, yet provide evidence needed for making adroit next-step instructional decisions, is what's needed. Although the most significant factor in the formative-assessment process is the quality of the en route assessments being used and the evidence they provide, sufficient attention has simply not been given to how to provide teachers with formative tests, or how to report right-size results. Putting it more tersely, we need to make it easy for teachers to employ formative assessment. That's right, easy.

Scrutiny of the many introductory books devoted to formative assessment reveals scant attention given to the necessity of creating right-size reports and, moreover, little heed to sharpening the way in which right-size reports will be provided so that formative assessment makes a meaningful improvement in students' learning. It is, clearly, time to change our ways.

### Declaration of Conflicting Interests and Ethics

The author declares no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author.

### Orcid

William James Popham  <https://orcid.org/0000-0001-5822-0303>

### REFERENCES

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning, *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-73. <https://doi.org/10.1080/0969595980050102>
- Popham, W.J. (2008). *Transformative assessment*. ASCD.
- Popham. W.J. (2011). *Transformative assessment in action*. ASCD.