

thinkpoint

Deforming the formative: How a summative mindset thwarts the aims of formative assessment

by Arthur Chiaravalli

In the past couple decades in education, few concepts have received such universal approval as *formative assessment*. Broadly defined as “information collected and used by teachers and students during instruction to improve teaching and learning as it is occurring” (Michigan Assessment Consortium, 2018),

cycle, such as a unit, term, or school year. Although few teachers would argue that summative assessment does not have a role in education, it has limited use for guiding real-time instructional adjustments. True to their names, formative assessment is designed to form new teaching and learning; summative

Formative Assessment: Essential features and some obstacles to classroom integration

Most writers on the topic of formative assessment identify three essential features, all of which can be phrased as questions:

1. Use of learning targets: Where am I going?
2. Evidence of student understanding: Where am I now?
3. Plan for improvement: How can I close the gap?

Ever since Michael Scriven coined the phrase “formative evaluation” in 1967, research has continued to show the beneficial effects of the formative assessment process. In their landmark survey of this research, Paul Black and Dylan Wiliam (1998) stated that formative assessment “is at the heart of effective teaching,” noting that “improved formative assessment reduces [achievement gaps] while raising achievement overall.” Part and parcel of the method is an understanding that students will themselves become “assessment capable,” using teacher, self, and/or peer



formative assessment has been roundly recommended as an essential component of good teaching.

Formative assessment is most often contrasted with summative assessment, which is designed to measure educational outcomes at the end of a learning

assessment is designed to sum up (usually in the form of a final grade or score) what has been learned. In an ideal world, formative and summative assessment work seamlessly together within a balanced assessment system to raise student achievement.

feedback to adjust learning goals and tactics. John Hattie (2009) asserted that self-assessment—a critical piece of the formative assessment process—tops the list of educational interventions with the highest effect size.

Few would disagree with the effectiveness of formative assessment, especially when understood as an intentional process skillfully integrated with instruction. Its continued and widespread popularity stems in part from the ease with which formative assessment activities can be incorporated into the flow of teaching. A quick search of the word “formative” on Edutopia returned 594 results, along with countless other articles on adjacent topics such as feedback, self-assessment, metacognition, and differentiation. And while this enthusiasm often lacks awareness of formative assessment as a “systematic and planned process” with necessary elements (Greenstein, 2010, p. 170), many teachers seem to show a genuine appreciation for formative assessment and its potential to improve teaching and learning in their classrooms. Most intuitively grasp assessment expert Bob Stake’s distinction: “When the cook tastes the soup, it is formative; when the guests taste the soup, it is summative” (cited in Hattie, 2015, par. 12).

Yet, obstacles still abound in normative classroom practice. Black and Wiliam’s assertion that enhanced formative assessment will likely require “significant changes in classroom practice” still rings true more than 20 years later.

In theory, formative assessment still makes sense in an era dominated by high-stakes accountability tests, both internal and external. All things being equal, why wouldn’t a teacher want to know where students stand vis-à-vis the learning targets tested on these summative assessments? And if a gap remains, what teacher or student wouldn’t want the opportunity to use feedback from formative assessments to make improvements?

Despite this commonsense appeal, practices with demonstrably negative cognitive and motivational effects still dominate in American schools. Again, some of this stems from an incomplete understanding of formative assessment as an intentional planned process. But it can also be traced to failures in one or more of formative assessment’s three phases. I would further argue that these failures frequently have a common source: the overriding dominance of high stakes summative assessments that prevent teachers and students from embracing the promising practices, priorities, and dispositions that support the formative assessment process.

Although these summative exams may be relatively infrequent, they have had an outsized impact on the quality of formative assessment, often all but eliminating their beneficial effects. Moreover, this negative impact is not visited equally on all segments of the population, as minority and low-income schools often devote vastly more time toward instruction

and assessment that mimics the summative exams (Lazarín, 2014, p. 21).

Although incontrovertible proof of testing culture’s harmful and distorting effects on formative assessment may be hard to produce, I hope to show how a testing mentality infects and distorts each of the three phases of formative assessment, and provide some preliminary thoughts about how we might find our way toward the practices that will help us reap its proven benefits.

Where am I going?

With its focus on objectivity and reliability, the accountability era has emphasized aligning instruction and assessment around clear learning targets, aiding students in answering the first question of formative assessment, “Where am I going?” As a novice language arts teacher, I appreciated how the concept of learning targets helped organize and sharpen my thinking around instruction and assessment in the classroom. By training my focus on the specific skills students would acquire, learning targets helped me to challenge and ultimately change my department’s widespread practice of testing kids on “who did what when” in the literary works we read.

I even felt that the high-stakes summative assessments awaiting us at the end of each semester (for a time, 50% of my evaluation was based on students’ performance on these exams) was at least preferable to its predecessor: curriculum maps that dictated what content needed to be covered when. The exam’s hands-off approach

allowed me the flexibility to be more responsive to my students' learning needs. Since modern accountability tests are largely agnostic to questions of content, process, and products, I felt free to give my students considerable latitude in the avenues they could use to demonstrate learning. All this was very good.

As time went on, however, I began to notice how summative assessments exerted a distorting influence on this first question of formative assessment: Where am I going? This recognition first occurred to me when my colleagues and I began revising the posttests in our department, shifting them away from their former focus on facts (e.g. who did what when in such-and-such novel) toward one explicitly linked to state and national standards. While it was true that the new exams gave us freedom to differentiate content, instruction, and tasks in the classroom, I noticed that the multiple-choice format of the exam (required due to concerns about "objectivity") dictated we choose those understandings that were easily measured in this way.

Without dismissing the importance of clarity in developing and sharing learning targets, I noticed that the learning targets featured on the exams were of lower quality—vocabulary, terminology, grammar, lower-order cognitive processes. In theory at least, one might say that nothing prevented me from going beyond these measurable minimums in the classroom, emphasizing those that are closer to the heart of language arts: critical thinking, self-expression, literary

analysis, creativity. When push came to shove, however, I noticed that both the students and I intuitively began to privilege these same lower quality targets in the classroom, to the point where the occasional remark "this is going to be on the final," became the cue that would get people off their phones and paying attention.

Why was it that this impoverished counterfeit of my discipline often eclipsed the rich, interconnected, complex learning targets that are the essence of language arts? Why was it that motivation spiked during games of Kahoot! and Quizlet Live preparing for vocabulary or ter-

later demonstrate. This assumption, however, often turned out to be false.

One might argue that I could further advocate for the inclusion of more complex performances, such as a writing prompt, on the posttest. Arguably, this change would lead to increased focus and motivation in learning these skills in the classroom. But leaving aside the fact that very few teachers would fight for the opportunity to rubric-score 100+ student essays at the end of a semester, this approach also has its problems. In short, in order to ensure the reliability and objectivity required on high-stakes summative assessments,

Although summative exams may be relatively infrequent, they have had an outsized impact on the quality of formative assessment, often all but eliminating their beneficial effects. Moreover, this negative impact is not visited equally on all segments of the population, as minority and low-income schools often devote vastly more time toward instruction and assessment that mimics the summative exams.

minology quizzes, but devolved into cynicism during classroom discussions, slam poetry, reader's theater, and passion projects? My own imperfect teaching notwithstanding, it truly seemed that the test had become the target, diminishing the importance of anything not as easily measured. When first developing the tests, I told myself that these lower-order skills were simply necessary prerequisites, proxies for the richer, more challenging understandings students would

I must focus my attention on those elements of writing that are easily and reliably measured. Maja Wilson (2007) observes that measurable aspects can represent "only a sliver of... values about writing: voice, wording, sentence fluency, conventions, content, organization, and presentation" (p. 62-63).

As Linda Mabry (1999) puts it, The standardization of a skill that is fundamentally self-expressive and individualistic

obstructs its assessment. And rubrics standardize the teaching of writing, which jeopardizes the learning and understanding of writing (p. 678).

According to Connie M. Moss and Susan Brookhart (2012), learning targets help students “sharpen their aim in pursuit of essential understandings” (p. 47). The fact that essential understandings are so often obscured and obstructed by the values of summative assessment suggests that they undermine this first phase of formative assessment, blocking students from meaningful, coherent, accurate answers to the question, “Where am I going?”

Where am I now?

Other factors prevent students from benefiting from feedback or learning how to self-assess. Feedback, commonly seen as fundamental to the formative assessment process, has frequently been shown to be ineffective. Kluger and DeNisi (1996) found that in 38% of well-designed studies, feedback actually made performance worse (p. 258).

Perhaps the most common and detrimental of these practices is the widespread conflation of assessment with grading. This feature, which likely takes its cue from the summative impulse to provide summaries of achievement for reporting purposes, has been shown to short circuit students’ ability to engage in the formative assessment process. Although the reasons for this may not have been entirely clear in his time, even Bloom (1969) noted that formative assessment is “much more effective...if

it is separated from the grading process and used primarily as an aid to teaching” (p. 48).

Ruth Butler (1987, 1988) examined 3 types of feedback: scores alone, comments alone, and scores with comments. She found that students who received scores alone showed no subsequent improvement in learning. Interestingly, scores with comments were just as

tend to become discouraged. Diagnostic comments without scores were effective, she reasoned, because they represented “task-involving” feedback. In short, the kind of feedback students receive make them either interested or uninterested in the diagnostic comments that can help them appraise their progress and move forward in their learning.



ineffective in that students focused entirely on the score and ignored the comments. Only students who received comments alone demonstrated improvement. Butler also found that both low- and high-achieving students’ motivation and achievement declined significantly when graded, compared to those who received only diagnostic comments.

Butler concluded that scores with or without comments were equally ineffective because both contain the “ego-involving” feedback of scores. Students who receive high scores tend to become complacent, whereas students receiving low scores

These findings confirmed my own experience. Too often when I returned papers or assignments to students, they would often flip immediately to the rubric to view their scores, ignoring any feedback I’d painstakingly provided. This remained true even though, for most of my career, I have allowed students to revise and redo assessments. Something about the letter grade causes learning to stop.

Of course, subsequent studies have shown how extrinsic motivators like grades actually inhibit the development of a “learning orientation,” critical to the formative assessment process. Carol Dweck’s research

(2002, 2006) has since shown us how feedback to children can have motivational consequences. Inasmuch as grades are often seen as a judgment of intelligence, they contribute to what Dweck termed an “entity’

students are given a percentage or grade, they can do little else besides compare it to others around them, with half or more deciding that they are not as good as others. This is

character. As Ruth Butler and later researchers have shown, diagnostic comments can do little to mitigate this overriding impression, one that effectively prevents students from answering the question, “Where am I now?”

Inasmuch as grades are often seen as a judgment of intelligence, they contribute to what Dweck termed an “entity’ theory of self,” otherwise known as “fixed mindset.” Students with a fixed mindset, in addition to avoiding academically challenging tasks, were more likely to lie about their performance, effectively preventing them from assessing their current level of understanding, the second phase of the formative assessment process.

theory of self,” otherwise known as “fixed mindset.” Students with a fixed mindset, in addition to avoiding academically challenging tasks, were more likely to lie about their performance, effectively preventing them from assessing their current level of understanding, the second phase of the formative assessment process. Students exposed to more process-oriented, diagnostic feedback were more likely to develop an “incremental’ theory of self,” or “growth mindset,” one that produced more positive responses to learning challenges.

Jo Boaler (2016) lamented the ubiquity of this “performance culture” in mathematics:

The testing regime of the last decade has had a large negative impact on students, but it does not end with testing: the communication of grades to students is similarly negative. When

known as “ego feedback,” a form of feedback that has been found to damage learning. Sadly, when students are given frequent test scores and grades, they start to see themselves as those scores and grades. They do not regard the scores as an indicator of their learning or of what they need to do to achieve; they see them as indicators of who they are as people. The fact that U.S. students commonly describe themselves saying “I’m an A student” or “I’m a D student” illustrates how students define themselves by grades (p. 142-143).

The fact that summative scores and marks are continually logged, averaged, and reported out via our 24-7 online gradebooks inevitably infects the culture of the classroom, causing even purely formative activities to lose their essential

How can I close the gap?

As mentioned in the previous section, online gradebooks can have the effect of turning every assessment into a summative one. As Black and William remark, “the collection of marks to fill in records is given higher priority than the analysis of pupils’ work to discern learning” (p. 5). A related factor seems to blame for students’ inability to answer this final question, “How can I close the gap?”

While teachers have seemingly been willing to embrace ungraded formative activities like exit tickets, fist-to-five, thumbs-up/thumbs-down, or audience response-style clickers, they have been less willing to consider anything that’s already in the gradebook as formative. As a first-time administrator, I notice how in a majority of classes these electronic scores, percentages, and letters are often as indelible as figures chiseled in stone. Why this is so stubbornly the norm is not entirely clear. Improving on the class record book and pencil, electronic gradebooks can not only erase and replace grades, but also *automatically recalculate the total grade*.

One would think that this development would open the door to considering everything up to the end of the term, semester, or year formative. This situation would also be vastly enhanced

if these assessments were ungraded so as to prevent students from developing defeating performance mindsets that sap motivation. But even if that is too radical an idea, redos and retakes seem like an absolute requirement for fostering a learning orientation in schools.

As Moss and Brookhart note, “When classroom lessons consist of do-or-die tasks or assignments—one-time-only chances to demonstrate mastery—students have little chance or reason to learn how to assess their own work and to value the process” (pp. 79-80). Over time, as poor grades compound, struggling students find less and less value in the question “How can I close the gap?” because the gap between them and success becomes a yawning chasm. Not even a month into the semester, I witness students dropping out of classes because their calculated final grade—an idiosyncratic amalgamation of homework, classwork, quizzes, and tests, graphic organizers, outlines, drafts, and papers—has solidified into an immovable mass. These immutable numbers, already calculating in the first week of school, betray yet another characteristic of summative assessment: finality. Summative assessments, in their role of assessing and reporting the final result of a learning cycle, rightly possess a certain finality to their pronouncements. Students may retake the class or do summer school, but institutions are not able to postpone these kinds of evaluations indefinitely.

While the class is still in “mid-stream,” however, it seems

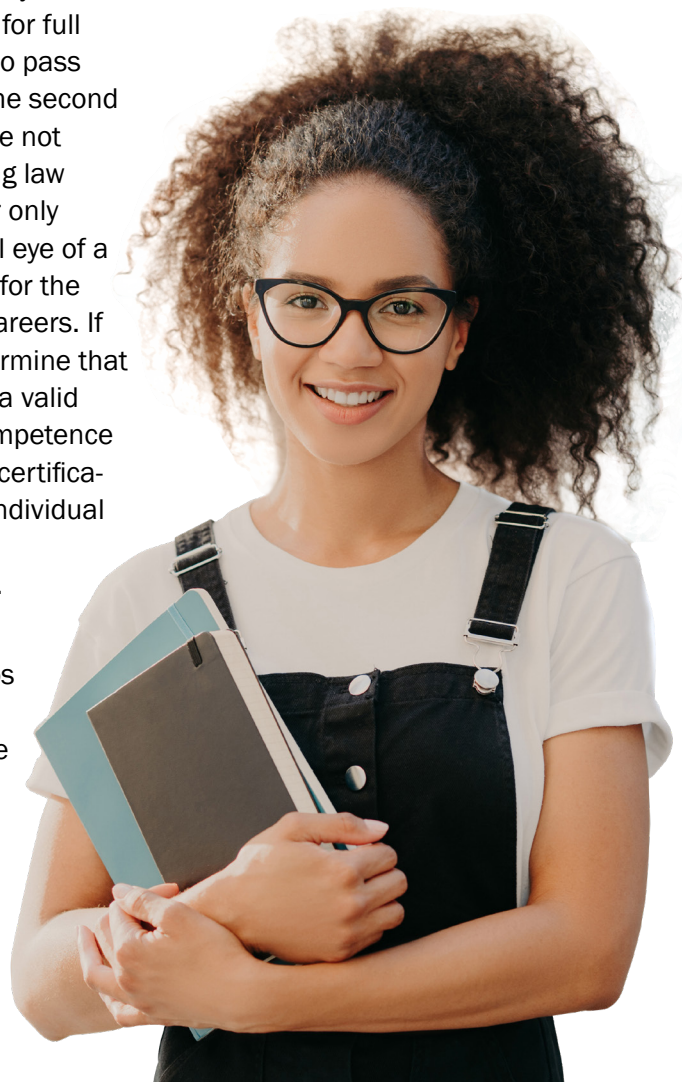
excessively harsh to prevent students from having frequent opportunities to engage fruitfully in this third stage of the formative assessment process from the very beginning. Some teachers argue that they are forming students, preparing them for the “real world” where you don’t get second chances. Rick Wormeli (2018) capably exposes this assertion as a myth:

LSAT. MCAT. Praxis. SAT. Bar exam. CPA exam. Driver’s licensure. Pilot’s licensure. Auto mechanic certification exam. Every one of these assessments reflects the adult-level, working-world responsibilities our students will shoulder one day, and all can be retaken for full credit. Lawyers who pass the bar exam on the second or third attempt are not limited to practicing law only on Tuesday or only under the watchful eye of a seasoned partner for the duration of their careers. If professionals determine that a certifying test is a valid assessment of competence in their field, then certification qualifies the individual for all rights and privileges (p. 211).

Teachers rightly point out that allowing redos and retakes can be potentially much more time consuming than a “one-and-done” approach to assessment. With burnout among American teachers reaching unprecedented levels, this topic

requires more research. As Dylan Wiliam points out,

My final concern in all this is that many, if not most, research efforts on supporting teachers in the use of formative assessment represent a “counsel of perfection.” There is a focus on meeting the needs of all students that is laudable, but simply unlikely to be possible in most American classrooms. American teachers are some of the most hard-working in the world, with around 1,130 contact-hours per year compared to the OECD average of 803 hours for



primary and 674 for upper secondary (p. 287).

That said, if we are unable to find ways to go beyond a mindset of finality and toward a formative one, the promise of formative assessment will be largely squandered. Working around the margins of a mindset that values finality over formation, our efforts will bear little fruit. With no available avenues forward, students will lose the appetite and ability to answer the question, “How do I close the gap?”

Conclusion

“Unfortunately, the arrival of formative assessment in America was ill timed. This potentially powerful classroom-based learning and teaching innovation was overshadowed almost immediately by the No Child Left Behind Act (January 2002) with its intense pressure to raise scores on external accountability tests.”
– Lorrie A. Shepard, 2007, p. 279

While I’ve mostly considered ways our obsession with high-stakes summative assessments hinder efforts to promote the formative assessment process, reasons for optimism still exist. First among these is the continued common-sense appeal of formative assessment. Especially in this age of accountability, no teacher or student wants to be blindsided by poor performance. Instead, we see that providing clear targets, ascertaining student understanding, and providing ongoing opportunities for improvement are, if nothing else, acts of self-preservation.

The metaphor of the coach, who prepares players for “game time,” seems especially apt. Formative assessment is analogous to the process used by coaches to prepare players for an upcoming game. Coaches don’t put a

holding the space for students to pursue meaningful learning goals (Where am I going?).

This last task cannot help but involve a major shift from the

Working around the margins of a mindset that values finality over formation, our efforts will bear little fruit. With no available avenues forward, students will lose the appetite and ability to answer the question, “How do I close the gap?”

score on the scoreboard during practices; that only happens during the game. Up until that “moment of truth,” coaches do everything they can to develop players in the skills and concepts they will need to succeed.

Most teachers I speak with intuitively understand this analogy. Even so, more work is needed to challenge a summative system responsible for pushing both teachers and students to their breaking point. As I see it, the road out of this degrading and demotivating situation involves working backward through the three questions of formative assessment. Teachers should:

1. find workable ways of providing multiple opportunities to demonstrate learning (How can I close the gap?),
2. minimize or eliminate grades that obscure diagnostic feedback (Where am I now?), and
3. challenge the reductive, narrowing imperatives of summative assessments,

policies and priorities of the accountability age, toward a new paradigm in which formative assessment can fully flourish.

Author



Arthur Chiaravalli is the co-founder of *Teachers Going Gradeless*, a global group of educators convinced that teaching and learning is better when we grade less and emphasize feedback and growth. He has had an extensive career as a high school teacher and currently serves as Dean of Students for a mid-Michigan high school.

Reflection Points

- 1 How have you seen a summative mindset infect the formative assessment process?
- 2 Which recommendations for mitigating this mindset can you see yourself incorporating into your own classroom or school?
- 3 Which of the recommendations do you think would be too difficult or impractical? What needs to change in education for these to become feasible?
- 4 How can we help teachers develop their knowledge of formative assessment as an intentional, planned process without becoming a “counsel of perfection”?

References

- BLACK, P., & WILIAM, D.** (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- BLOOM, B. S.** (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Educational evaluation: new roles, new means: the 63rd yearbook of the National Society for the Study of Education (part II)* (Vol. 69(2), pp. 26-50). Chicago, IL: University of Chicago Press.
- BOALER, J.** (2016). Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages, and innovative teaching. San Francisco, CA: Jossey-Bass.
- BUTLER, R.** (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474-482.
- BUTLER, R.** (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58(1), 1-14.
- DWECK, C. S.** (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 37-60). San Diego, CA, US: Academic Press.
- DWECK, C. S.** (2006). *Mindset: The new psychology of success*. New York: Ballantine Books
- GREENSTEIN, L.** (2010). *What teachers really need to know about formative assessment*. Alexandria, Va: ASCD.
- HATTIE, J.** (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- HATTIE, J.** (2015, October 27). We aren't using assessments correctly: There's a distinction between formative and summative assessments. Retrieved from <https://www.edweek.org/ew/articles/2015/10/28/we-arent-using-assessments-correctly.html>.
- KLUGER, A. N., & DENISI, A.** (1996). The effect of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- LAZARÍN, M.** (2014, October 16). Testing overload in America's schools. [PDF file]. Retrieved from <https://cdn.americanprogress.org/wp-content/uploads/2014/10/LazarinOvertestingReport.pdf>
- MABRY, L.** (1999). Writing to the rubric: Lingering effects of traditional standardized testing on direct writing assessment. *Phi Delta Kappan* 80(9), 673-79. Michigan Assessment Consortium. (2018, February). *Assessment literacy standards glossary* [PDF file]. Retrieved from <http://michiganassessment-consortium.org/wp-content/uploads/ALS-Glossary-Feb2018.pdf>
- MOSS, C. M., & BROOKHART, S. M.** (2012). *Learning targets: Helping students aim for understanding in today's lesson*. Alexandria, VA: ASCD.
- SHEPARD, L.** (2007). Formative assessment: Caveat emptor. In C. A. DWYER (Ed.), *The Future of Assessment: Shaping Teaching and Learning*. New York: Taylor and Francis Group.
- WILIAM, D.** (2006). Formative assessment: Getting the focus right. *Educational Assessment*. 11(3), 283-289.
- WILSON, M.** (2007). Why I won't be using rubrics to respond to students' writing. *English Journal*, 96(4), 62-66.
- WORMELI, R.** (2018). *Fair isn't always equal: Assessment and grading in the differentiated classroom*. Portland, Maine: Steinhouse.

To learn more

The Michigan Assessment Consortium website contains a rich library filled with sharable assessment resources as well as online learning modules designed to increase assessment literacy among all education stakeholders. Explore these resources, then spread the word! Visit www.michiganassessmentconsortium.org.