Chapter 1

Our Evolving Understanding of Formative Assessment and the Challenges of Widespread Implementation

Valerie L. Mills, Marilyn E. Strutchens, and Marjorie Petit

> *The power of formative assessment for learning, when well done, is firmly established in the research.*
>
> *Black and Wiliam, 1998*

> *I know of no other school improvement innovation* [referring to formative assessment practices] *that can claim effects of this nature or size.*
>
> *R. Stiggins, 2002*

> *It's really not surprising that formative assessment works so well. What is surprising is how few U.S. teachers use the process.*
>
> *J. Popham, 2013*

These three quotations from well-respected educational researchers, when juxtaposed, leave us scratching our heads and wondering *why*? Why has formative assessment not been used more regularly in the decades since Paul Black and Dylan Wiliam published *Inside the Black Box* (1998)? How is it possible that an instructional innovation with such potential to improve student learning has not been more widely adopted? For those interested in accessing the potential of the formative assessment process, these are vital questions worth exploring. Therefore, we open *A Fresh Look at Formative Assessment in Mathematics Teaching* with a brief look back—first, to review formative assessment's research base, and then to consider why formative assessment may not have been widely adopted. This discussion can then serve as both foundation and motivation for a *new* approach to understanding and implementing formative assessment for mathematics education researchers and practitioners alike.

The Case in Support of Formative Assessment

In the 20 years since Black and Wiliam first published their meta-analysis of the effects of formative assessment practices on learning, *Inside the Black Box* (1998), the topic has received much attention. As a measure of this attention, a non-specific *Google Scholar* search for "formative assessment" on June 8, 2017, yielded nearly 600,000 results, along with a list of 16 related searches such as formative assessment tools, formative assessment strategies, and so forth. This active research stream focused on formative assessment has generally affirmed the findings of effect sizes ranging from 0.40 to 0.70, which were first suggested by Black and Wiliam in 1989. For example, we have additional large meta-analysis studies from Ehrenberg's et al. (2001) and Hattie (2009). Ehrenberg's et al. (2001) found that the impact of formative assessment on student achievement is four to five times greater than reduction of class size. Hattie's (2009) work synthesized approximately 800 meta-analyses covering a wide range of educational programs, policies, and innovations on academic achievement. He found formative assessment and the related strategies of self-assessment and feedback to be ranked in the top 10 of the 138 interventions examined with effect sizes of 0.90, 1.44, and 0.73, respectively. Further, the literature suggests that not only are the gains associated with formative assessment use larger than most instructional innovation strategies (Hattie, 2009), this process has also been found to be particularly helpful for students who have previously struggled (Wiliam, 1989), and appears to produce learning that is sustained over extended periods of time (Wiliam, 2005). Finally, these findings are consistent across developed countries (i.e., the United States, Canada, England, Israel, and Portugal), across age brackets, and across content areas. Indeed, the evidence base associated with formative assessment's potential to support significant growth in student learning is compelling.

This robust research base has led to a growing consensus across educational communities that formative assessment is a powerful instructional tool worthy of every teacher's attention and an important element of effective instruction. Evidence for this consensus can be seen with the inclusion of formative assessment in three of the most influential contemporary frameworks for mathematics teaching. First, NCTM's 2014 publication *Principles to Actions: Ensuring Mathematical Success for All* lists "Elicit and use evidence of student thinking" as one of eight named research-informed Mathematics Teaching Practices. The authors write, "Effective teaching of mathematics uses evidence of student thinking to assess progress toward mathematical understanding and to adjust instruction continually in ways that support and extend learning" (NCTM, 2014, p. 10). Second, Schoenfeld (2015) includes formative assessment as one of only five essential dimensions of mathematics classroom practice in his mathematics education framework, *Teaching for Robust Understanding* (TRU). Regarding this dimension, the framework notes that we are likely to develop powerful student thinkers based on "The extent to which classroom activities elicit student thinking and subsequent interactions respond to those ideas, building on productive beginnings and addressing emerging misunderstandings. Powerful instruction 'meets students where they are' and gives them opportunities to deepen their understandings" (p. 163). A third influential framework for effective teaching is Deborah Loewenberg Ball's 19 High-Leverage Practices (Ball & Forzani, 2010). This framework digs deeply into the work of teaching to elucidate component skills needed to facilitate learning. In this schema, formative assessment components are teased apart for close examination, including the following:

- Eliciting and interpreting individual students' thinking
- Coordinating and adjusting instruction during a lesson
- Checking student understanding during—and at the conclusion—of lessons

- Interpreting the results of student work, including routine assignments, quizzes, tests, projects, and standardized assessments
- Providing oral and written feedback to students

Each of these frameworks represents an attempt to synthesize the knowledge and practices needed to promote learning for each and every student. Whether it is one of eight, five, or 19 practices, formative assessment is well represented in each.

Formative assessment research since 1998 has not only supported the Black and Wiliam findings but it has also added important detail and depth to our understanding of the process and the instructional strategies that contribute to that process. This maturing understanding of formative assessment is illustrated in the work of the Formative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Student Standards (SCASS). The FAST SCASS is a working referent group for educators employed primarily by the various state-level education departments from across the country. This special interest group is organized under the Council of Chief State School Officers (CCSSO) for the purpose of providing guidance and resources to state-level policy makers on formative assessment. When the group initially formed in 2006, they created a working definition for formative assessment. The group, led by Margaret Heritage, adopted an updated definition designed to reflect their evolved understanding of the process and the associated instructional strategies in 2017. It is instructive to compare the two definitions found side by side in Table 1.1 below, attending to the ways in which the two descriptions of formative assessment are alike and different. The similarities and differences described below, particularly the differences, exemplify the ways in which our thinking about this instructional innovation has grown and matured over the past 15 or 20 years.

Table 1.1. Similarities and differences between FAST SCASS 2006 definition and FAST SCASS 2017 definitions of formative assessment

FAST SCASS 2006 Definition	FAST SCASS 2017 Definition
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.	Formative assessment is a planned, ongoing process used by all students and teachers during learning and teaching to improve student understanding of intended disciplinary learning outcomes, supporting students becoming more self-directed learners. Effective use of the formative assessment process requires students and teachers to integrate the following practices: • Clarifying learning targets within a broader progression of learning; • Eliciting and analyzing evidence of student understanding; • Engaging in self-assessment, self-reflection, and peer assessment; • Providing actionable feedback; and • Using evidence and feedback to move learning forward by adjusting either learning strategies or next instructional steps.

While both definitions carefully describe formative assessment as a *process*, compared side-by-side, one notices that in the updated version the description of the *process* gains detail becoming a *planned ongoing process*. Further the subjects of the process shift from *teachers and students* to *all students and teachers*, emphasizing the role that students play in the process. The role of students in the process gets an additional lift as the "when we employ" formative assessment evolves from "during *instruction*" to "during *teaching and learning*." This change is again intended to signal that students as well as teachers play a role in the process. Readers will also notice that the learning targets of formative assessment are broadened from *intended instructional outcomes* to become *improve student understanding of intended disciplinary learning outcomes, supporting students becoming more self-directed learners*. This change alerts us to both the critical role of individual subject areas within the process, and the opportunities formative assessment affords learners to cultivate skills we associate with life-long learners. Finally, the revised definition assigns a list of explicit strategies to the process description that adds clarity and depth to the reader's understanding of the formative assessment process. While each adjustment is relatively small, the changes and additions substantively add to our thinking about who, what, when, and how the process is intended to be implemented.

The addition of explicit strategies to better describe the formative assessment process has been part of the effort to understand and advance formative assessment from the beginning of this educational storyline. Four years after the U.S. publication of *Inside the Black Box (1998)*, these researchers, with others, published *Working Inside the Black Box*. These authors summarized findings from the study they designed to unpack the components of this process. This work identified and described four component strategies: Questioning, Feedback through Grading, Peer and Self-Assessment, and Formative Use of Summative Assessments. Three years later, Leahy et al. (2005) in *Minute by Minute* described the following five formative assessment tools that added greater detail to our understanding of the strategies that fuel this instructional process:

- clarifying and sharing learning intentions and criteria for success;
- engineering effective classroom discussions, questions, and learning tasks
- providing feedback that moves learners forward;
- activating students as the owners of their learning; and
- activating students as resources for one another.

In the years that followed the Leahy et al. list, other groups have offered either more detailed or focused lists of strategies, or both, to better support users at varying grade-levels in their efforts to understand and implement this productive instructional process. These lists have been further developed as the basis for a wide variety of books and resources with illustrations and suggestions for using the strategies in classrooms to support teachers working to move formative assessment into the components of their practice.

Based on confirming research gathered over the past twenty years, a maturing explication of formative assessment, and a growing catalog of resources to support users in implementing the process, the education community has signaled its confidence in this process and its important place in our understanding of teaching practices that are likely to support greater learning for each and every student.

Challenges and Barriers to Implementation

Having briefly explored the evolving research and practice supporting the significance of formative assessment as an effective practice, we turn our attention to the last of the opening quotations for this chapter and the question it prompts: Why has formative assessment not been taken up more consistently in the decades since the publication of *Inside the Black Box*? Further, which challenges or barriers might we hypothesize have hampered implementation thus far, but could be addressed going forward? One place to begin this discussion is to wonder if our early, less articulate descriptions of the process may have undermined teachers' understanding of the formative assessment process and their ability to translate it into their day-to-day practice. We find some evidence for this in studies that have questioned the magnitude and range of effect sizes described above. Several researchers have challenged the accuracy of the estimated effect sizes for student achievement gains based on concerns about the original studies used to develop these projections (Wininger & Norman, 2005; Dunn & Mulvenon, 2009; Apthorp et al., 2016; and Briggs et al., 2012). Many of these concerns are connected to variability in the way that different researchers/studies chose to define formative assessment and their sometimes small sample sizes (Wininger & Norman, 2005).

From the point of view of the practitioner, anecdotal evidence suggests that early, less complete descriptions of formative assessment have contributed to at least two prevalent misconceptions. First, it seems that many educators, including teachers, principals, policymakers, and other leaders have had difficulty distinguishing between summative assessments and the formative assessment process. Confounding these two processes is the word *assessment* that appears in both. The challenge has been to help educators broaden their current understanding of assessment as *events* (chapter tests, quizzes, large-scale assessments) that occur *outside* day-to-day lessons to also include assessment as a *process* that takes place *inside* the daily stream of teaching and learning for the purpose of adjusting instruction. For many this conflated thinking about formative and summative assessment has meant that to adopt formative assessment practices is to layer additional assessment events into the school calendar, rather than incorporating a process of eliciting and using evidence of learning to adjust instruction into daily practice. The additional assessments, erroneously labeled formative, take many forms such as benchmark assessments or pre-tests administered prior to teaching but not used to adjust initial instruction. Assessments used in this way cannot deliver the promised growth enabled by the formative assessment process that is intended to occur within the daily stream of planning and instruction.

A second common misconception occurs when educators understand that formative assessment is intended to be embedded in instruction, but use only one or two of the individual strategies. In this situation, strategies that could be productive are left to stand alone rather than being connected to a completed process. For example, many teachers use exit tickets to gather evidence of student understanding, or carefully post daily student learning goals, or ask students to critique one another's work, but these practices are not then used to advance learning as part of the process. That is, exit tickets are used to efficiently collect evidence of student's current learning, but are rarely summarized and then used to adjust instruction. Similarly, while student learning goals are posted, they may not connect to the day's instructional tasks, and students are not given the time or support needed to analyze their work in relation to the learning goals. Although individual strategies can be useful, each constitutes only a portion of the process that includes identifying learning goals, eliciting and collecting evidence of learning around the goal, and then

critically, using the evidence to adjust the instruction and engage students in ways that advance learning toward the goal *during instruction* (Heritage, 2008). Individual strategies that contribute to the overall process are important elements within the process, but alone, they may not be sufficient to advance learning.

Prevalent misuses and misunderstandings about formative assessment as a process suggest that there is still a great deal of work to be done. Despite the maturing descriptions, additional resources, and growing consensus, the education community has yet to fully grasp and effectively use this powerful instructional practice. Clearly the field is working to advance understanding for practitioners and researchers alike, but this work has not been sufficient, and the need still exists to help greater proportions of the education community to understand and use formative assessment productively.

A New Look at Formative Assessment

In January of 2013, a small group of practitioners and researchers from the National Council of Supervisors of Mathematics (NCSM) and the Association of Mathematics Teacher Educators (AMTE) came together to begin exploring the questions raised in the sections above and to wonder if a joint NCSM/AMTE effort might be developed to promote greater understanding and use of the formative assessment process. The group wondered why current research and professional learning efforts have not been more successful, despite maturing descriptions of the process and additional resources. The group also wondered if there was a different way into the formative assessment conversation that might be productive for practitioners and researchers alike. That is, how might the treatment of the formative assessment process in preservice and in-service learning be adjusted to better enable teachers to use these practices as intended? Based on the robust research base suggesting the importance of this instructional process, and the need for additional work, given the challenges associated with implementing these practices, the group agreed that a joint NCSM/AMTE task force should be formed. To ground this work, the task force began by clarifying, for our audience and ourselves, what is meant by formative assessment and included descriptions of the five component strategies outlined by Leahy et al. (2005) in the publication of a joint NCSM/AMTE position paper, *Improving Student Achievement in Mathematics Through Formative Assessment in Instruction*.

Beyond the grounding of task force efforts with the development and publication of the joint position paper, the task force also launched efforts in two other directions. The first was to commission a small group of practitioners and researchers led by Jeane Joyner and Mari Muri to develop a collection of formative assessment professional learning modules. These modules are designed for preservice and in-service leaders working with those who are new or inexperienced with formative assessment practices as they relate to mathematics teaching and learning. The *Jump-Start Formative Assessment* professional learning resources are now available to NCSM and AMTE members and can be accessed from the respective association websites.

The second effort of the joint task force was far less well defined and rested on an insight that emerged early-on as members explored causes and opportunities to impact the current state of formative assessment in America. This insight was prompted as the group began to speculate on possible connections between the formative assessment process and other well-researched approaches in mathematics education. What followed this early discussion was an exciting, enlightening, and ultimately productive journey. The journey grew to include practitioners and researchers from across the country and to produce evidence suggesting that formative assess-

ment plays a far more significant role in mathematics teaching and learning than was previously understood or explicitly described. The chapters that follow describe the rich results of this exploration including—

- a study developed and implemented to better understand how leaders in the field currently use or do not use the relationship between formative assessment and other mathematics frameworks, tools, and approaches;
- a working meeting with experienced users to initially explore formative assessment connections to each of five different instructional frameworks, tools, or approaches;
- thorough analysis and description of formative assessment connections in six widely used mathematics instructional resources written by practitioner and researcher teams deeply experienced with the resource;
- discussion of additional connections to NCTM's Effective Instructional Practices and issues of equitable achievement, and finally; and
- discussion of the important implications for changes in how we organize professional learning for formative assessment.

In closing, consider a final quotation from Paul Black and Dylan Wiliam. Their words foreshadow both our evolving understanding of the complexity associated with implementing formative assessment effectively and the work of the NCSM/AMTE task force to make explicit the role of formative assessment within many of mathematics education's supporting instructional frameworks, tools, and approaches.

> The evolution of effective teaching. The research studies . . . show very clearly that effective programs of formative assessment involve far more than the addition of a few observations and tests to an existing program. They require careful scrutiny of all the main components of a teaching plan. Indeed, it is clear that *instruction and formative assessment are indivisible*. (Black & Wiliam, 1998)

References

Apthorp, H., Klute, M., Petrites, T., Harlacher, J., Real, M., & Society for Research on Educational Effectiveness. (2016). *Valuing a More Rigorous Review of Formative Assessment's Effectiveness*. Paper presented at The Society for Research on Educational Effectiveness Conference, Washington, DC.

Ball, D. L., & Forzani, F. M. (2010). Teaching skillful teaching. *Educational Leadership 68*(4), 40–45.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

Briggs, D.C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice Winter, 31*(4), 13–17.

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation, 14*(7), 1–11.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Williams, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest, 2*(1), 1–30.

Hattie, J. (2008). *Visible learning: A Synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment.* Washington DC: Council of Chief State School Officers.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63*(3), 19–24.

National Council of Teachers of Mathematics (NCTM). (2014). *Principles to actions: Ensuring mathematical success for all*. Reston, VA: NCTM.

Schoenfeld, A. H. (2015). Thoughts on scale. *ZDM: The International Journal on Mathematics Education, 47*(1), 161–169.

Wininger, S. R., & Norman, A. D. (2005). Teacher candidates' exposure to formative assessment in educational psychology textbooks: A content analysis. *Educational Assessment 10*(1), 19–37. doi:10.1207/s15326977ea1001_2