No cover image available

The Oxford Handbook of Educational Psychology

Angela O'Donnell (ed.) et al.

https://doi.org/10.1093/oxfordhb/9780199841332.001.0001

Published: 2018 **Online ISBN:** 9780190881283 **Print ISBN:** 9780199841332

CHAPTER

Formative Assessment and Assessment for Learning: Distinctions, Mileposts, and a View to the Future

Jeffrey K. Smith, David Berg, Anastasiya A. Lipnevich

https://doi.org/10.1093/oxfordhb/9780199841332.013.27

Published: 18 August 2022

Abstract

Formative assessment and assessment for learning are two of the most powerful ideas related to learning in classrooms to be developed over the past half century. Used properly, they can enhance learning, build self-efficacy, and increase motivation in students. This chapter explores the history and fundamental ideas of formative assessment; assessment for learning; and, to a lesser degree, feedback—as it relates to assessment. The goal of the chapter is to give the reader a solid idea of how these ideas developed, how they are similar and different, how they relate to classroom practice and instructional theories, and what the future appears to hold for them. It begins by looking at the seminal work of Scriven (1967), and in particular Bloom (1968, 1969; Bloom, Hastings, & Madaus, 1971) who was the first scholar to talk about what eventually became formative assessment (called formative evaluation at the time). It then traces the development of the ideas and how they related to other research on classroom assessment (e.g., Crooks, 1988), and then to the work of the Assessment Reform Group (1999) who developed the notion of assessment for learning, and Black and Wiliam (1998, 2012, 2018) who greatly expanded and elaborated upon the ideas behind assessment for learning. It concludes by analyzing similarities and differences in the ideas presented, and by looking to some promising areas of current research and future possibilities.

Keywords: formative assessment, assessment for learning, feedback, classroom assessment, summative assessment, formative feedback

Subject: Educational Psychology, Psychology

Series: Oxford Library of Psychology

An aspiring archer draws a breath and a bowstring as she takes aim at her target. She walks through practiced steps in the shooting process and then releases the bowstring and watches the flight of the arrow. Slightly high and right. Maybe this is her last arrow in a competition and it has caused her to place second instead of first. Or maybe she is practicing all by herself in a field. If it is the former setting, perhaps she wins a small trophy. She might be pleased with her performance as it was better than she anticipated doing, or perhaps it is a personal best. In any case, it is the end of the day, and our archer returns home. If it is the

latter setting, in the field alone, again she might be pleased if it is an improvement over past efforts that day. But instead of calling it a day, in this setting the archer might be wondering, "What did I do right here, and what did I do wrong? How can I improve on this performance?" The particulars of the setting need to be taken into consideration.

This small vignette shows the same behavior and immediate goal, but with different contexts, and consequently, with differing outcomes and reactions. What we are looking at here is the difference between summative assessment, or assessment *of* learning (the competition grounds); and formative assessment, or assessment *for* learning (the practice field). In the first realization of the vignette, the archer is at a competition, seeing how good she is, how well her skills measure up to those of her fellow competitors, and against her previous efforts. In the second realization, she is alone, working on her skill, and wondering how she can improve. Perhaps she is wishing that her instructor were on hand to provide some feedback and suggestions. An interesting aspect of this story is that, if a camera were focused just on our archer, we could not tell whether she was competing or practicing: the behavior is the same. This is often the case for assessments; it is necessary to understand the context to understand whether this assessment is *of* learning or *for* learning, or pehaps a bit of both.

The purpose of this chapter is to explore the concepts of formative assessment and assessment for learning: the history, the hows and whys of the concepts, theory and research concerning the concepts, applications in instructional settings, and where the ideas of formative assessment and assessment for learning seem to be heading. This chapter is not intended to be a review of the literature on these ideas—excellent reviews already exist (as this chapter will show). Instead, it will be an explanation and exploration of the concepts and is directed toward readers who are relatively new to the ideas. As with almost any popular concept in educational psychology, scholars differ somewhat on what they think formative assessment and assessment for learning are, and where they should be going. We take a historical or developmental approach in this chapter, probably offending all of our professional colleagues instead of just some of them, but at the same time, as this is our chapter, we will also present our points of view. Our goal is to have the reader finish the chapter understanding where the fundamental ideas of formative assessment and assessment for learning come from, how they are similar and different, what the key milestones are in the development of these ideas, what important research findings are, and what the future appears to hold. Along the way we will spend time lookingat the related ideas of feedback and summative assessment.

An Overview

To fully understand the ideas behind formative assessment and assessment for learning, it is helpful to take a step back from the present and look at the development of these concepts, along with two companion ideas: summative assessment and instructional feedback.

Formative (and Summative) Assessment

The ideas behind all of the ideas presented in this chapter began well over a half century ago, with the work of Michael Scriven and Benjamin Bloom. Scriven, writing about the evaluation of curricular and instructional programs, is generally acknowledged to be the first person who used the terms "formative" and "summative" to label differences between efforts designed to improve a process and those designed to judge its overall quality (Scriven, 1967). He coined the terms "formative evaluation" and "summative evaluation" in his work describing efforts to improve programs and efforts to judge their overall quality. Shortly after Scriven, Bloom took the concepts of formative and summative and applied them to tests that might be given as part of instruction (Bloom, 1969). With this conceptual move, "formative" and "summative" became part of the lexicon of assessment of students. In 1971, Bloom edited a book on this topic, *Handbook on Formative and Summative Evalution of Pupil Learning*. At that time, "evaluation" was a term that was frequently used in much the same way that "assessment" is used today—not so much as a test of student ability as an approach to understanding where a student is in an instructional journey. "Measurement" referred to any process that resulted in assigning a number to a characteristic or trait of an individual, and "tests" were ... pretty much what they are today. Bloom's notion of formative evaluation of a student was to administer a short test after a certain amount of instruction had taken place to see how well students were doing and what areas they needed to work on more. The information provided by such an assessment was intended for the teacher as much as for the student (e.g., how to alter subsequent instruction on the topics/skills involved). Bloom incorporated this notion into his work on mastery learning theory (Bloom et al., 1971; also see Guskey, 2007), arguing that teachers should learn about their students' strengths and weaknesses, and allow them sufficient time to shore up weaknesses before moving on with instruction.

Summative assessment, in contrast to formative assessment, was assessment that had as its primary goal knowing where a student (or students) stood with regard to instructional objectives and goals. In summative assessment, there is no notion of using the assessment results to help students improve or advance in their learning. Examples of summative assessment would be national or statewide end-of-year testing, or international assessments such as the TIMMS, PISA, and PIRLS programs. In Bloom's mastery learning model, summative assessment would occur at the end of a course of instruction and often be used as part of the grade assignments for the course.

By the 1980s, the term formative *evaluation* had gradually been replaced with formative *assessment*, although as late as 1987–88, Natriello (1987) and Crooks (1988) still used the term evaluation in their reviews of evaluation in classrooms. (Crooks, however, would have preferred to use the term "assessment," but the editor of the journal Crooks was publishing in insisted upon "evaluation" as he thought that was the common parlance of the time; T. Crooks, personal communication, Dec. 5, 2005). Just why this shift took place is somewhat cloudy, but it may simply be that educators felt that evaluation had connotations of "values" inherent in it, whereas assessment had a softer, more student-centered tone (but that is just speculation on our part).

Assessment for Learning

As will be examined later in the chapter, the ideas of formative and summative assessment gradually became popular in educational research and practice, with formative assessment becoming a staple in instruction (in the ideal, if not always in reality). Assessment for learning is a more recent term, developed by the Assessment Reform Group in the United Kingdom in 1989. They distinguished assessment *for* learning and assessment *of* learning. The two ideas are roughly equivalent to formative and summative assessment, respectively. In 1998, the Assessment Reform Group set forth a set of characteristics for assessment for learning that they argued better promoted what ought to happen with assessment practice:

- It is embedded in a view of teaching and learning of which it is an essential part;
- It involves sharing learning goals with pupils;
- It aims to help pupils to know and to recognize the standards they are aiming for;
- It involves pupils in self-assessment;
- It provides feedback which leads to pupils recognizing their next steps and how to take them;
- It is underpinned by confidence that every student can improve; and
- It involves both teacher and pupils reviewing and reflecting on assessment data (Assessment Reform Group, 1999, p. 7).

Although the Assessment Reform Group attempted to distinguish the two ideas, in the literature on formative assessment and assessment for learning, it is clear that the two terms are used more interchangeably than distinctively. And, as Kingston and Nash (2011) point out, Mittler (1973) actually used the term "assessment for learning" back in the early 1970s. The question arises as to whether two terms are needed, or if the field could march on just using one of them and explaining what was meant by it. The problem, from our perspective, with the Assessment Reform Group's characteristics list is that it provides a desiderata rather than a definition. This is a good list of characteristics, even if it is a bit redundant at times; also, it is a creature of its time. That is, there are now other characteristics that are felt to be desirable for optimal formative assessment such as building a sense of self-efficacy in students (Lipnevich & Smith, 2009), or even allowing the student a say in the determination of learning goals. And on the other side of the equation, if an assessment is used to help a lecturer determine how to best present a follow-up lecture on a particularly difficult topic, is that then *not* assessment for learning because it does not provide feedback to students? Unfortunately, the Assessment Research Group draws distinctions between what they call "teacher assessment" and "assessment for learning," but not directly between "formative assessment," and "assessment for learning." We return to this issue later in the chapter, but for now we will consider formative assessment to be any assessment approach that has the improvement of student learning as its goal, and assessment for learning as being a highly learner-focused approach to formative assessment.

Feedback

Another factor has to be entered into the conversation about formative assessment and assessment for learning, and that is the notion of feedback. Feedback is a term that comes from engineering and physics from somewhere back in the 19th century, and relates how a system can provide information in a loop that influences subsequent performance. We are all familiar with what happens when a microphone gets too close to a speaker. To this end, one of the most cited feedback articles is the one by Ramaprasad (1983), who defined feedbck as information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way. Interestingly, Ramaprasad was a professor in information and decision sciences, with a bachelor's in engineering and a PhD in business administration, and he in no way meant to apply his feedback ideas to education. Hence, he was left entirely flabbergasted when he discovered exponentially growing citations on his paper years later, but from the field of education (Ramaprasad, personal communication, Nov. 2019). It was ultimately Sadler (1989) who adopted Ramaprasad's ideas to instructional contexts and who spoke of feedback as the vehicle to close the gap between students' current and desired level of performance. Sadler also discussed three necessary conditions for feedback effectiveness. The first condition has to do with a standard, toward which students aim during specific tasks they are working on. The second condition requires students to compare their actual levels of performance with the standard, and the third emphasizes student engagement in actions that eventually close the gap. Sadler suggested that these conditions had to be satisfied simultaneously for feedback to be most effective.

Around the same time, Kulhavy and Stock (1989) worked on developing ideas about feedback correctness. The central idea of Kulhavy and Stock's studies was that of response certitude, defined as a degree to which the learner expected his or her response to be correct. In addition to response certitude, the authors discussed response durability, defined as the likelihood that an instructional response would be available for the learner's use at some later point in time. Therefore, in situations where feedback is unavailable, the magnitude of certitude increases, and the probability of selecting the same response (often incorrect) increases also. Interestingly, the authors also noted that with increasing complexity, feedback will become indistinguishable from instruction.

Fast forward to 1996, we have Kluger and DeNisi who examined research over almost 100 years, presenting synthesis of how feedback works in instruction. The main and frequently cited finding of their meta-analysis was that feedback interventions increased individuals' performance by.4 standard deviations. At the same time, there was a great deal of variability of results, with one-third of studies showing a negative influence on performance. Wiliam (2018) provided an excellent updated review of the literature on feedback in instruction. Both Kluger and DeNisi (1996) and Wiliam (2018) show a generally strong positive influence of feedback on learning, but also that in some instances, it appears to have a negative effect. That is particularly the case when the feedback consists primarily of a grade given on a classroom assessment. So, decades ago (see, e.g., Kulhavy, 1977), the idea of feedback was closely linked to behaviorism and the development of programmed learning. Today, however, feedback has evolved into a term associated with a highly student-centered, constructivist approach to instruction and learning (see Hattie & Timperley, 2007; Lipnevich, Berg, & Smith, 2016; and Winstone et al., 2017).

Feedback differs from formative assessment and assessment for learning in that the "assessment" part of the process does not really have to be there. The notion of an assessment involves a process that is at least somewhat deliberate and planned, with the intention of gathering information that can be used for improving the learning of the student. Feedback, while it can be part of the process of formative assessment, does not require that it be planned, and indeed, can actually be spontaneous. A person can be walking down a sidewalk and trip on a section of sidewalk that is uneven. That individual will subsequently be checking the sidewalk to avoid another instance of that occurrance. An art student may smudge a part of a chalk drawing by accident and decide that it results in a desirable effect. She decides to keep that move in her repertoire.

At the other end of the spectrum, a teacher may decide to use a standardized measure such as an Advanced Placement test (in a US context), or other course credit assessment as a way to help students in her course with their learning. This would be a very carefully constructed assessment meeting demanding criteria, but could still be the genesis of instructional feedback to the student. Finally, the term feed-forward has entered into the literature on instruction. It is used to emphasize that the information being presented to the student is not about the past, but the future. That is, the emphasis should not be on where the student has been, but where he/she is going. So, instead of saying, "You forgot to reduce this fraction to its lowest terms," you would say, "It's best to take this one step further, and reduce it to its lowest terms in the future." Or to say, "On your next essay, let's focus on really thinking about the reader and how he/she is processing the story that you are telling." Whether one wishes to frame it as feeding it back, forward (or laterally), the idea is the same: Students should be able to act upon information presented to them for the purposes of improving performance.

Formative Assessment, Assessment for Learning, and Feedback: Comparing and Contrasting

Although formative assessment, assessment for learning, and feedback are similar terms, and sometimes (mistakenly) used interchangeably, it is important to delineate their similarities and differences. To begin, formative assessment and assessment for learning both involve the notion of assessment, which can be thought of as an intentional effort to learn about where a student is in the learning process. Feedback can involve assessment, but it does not have to. So, formative assessment and assessment for learning both necessarily involve assessment, whereas feedback does not. Next, we look at providing the learner with information about the learner's progress. This is a definite part of assessment for learning and feedback, but not necessarily for formative assessment. Although formative assessment frequently involves communication to the learner/student, it may be the case that a formative assessment is used solely to provide information to the teacher regarding next steps in the instructional process.

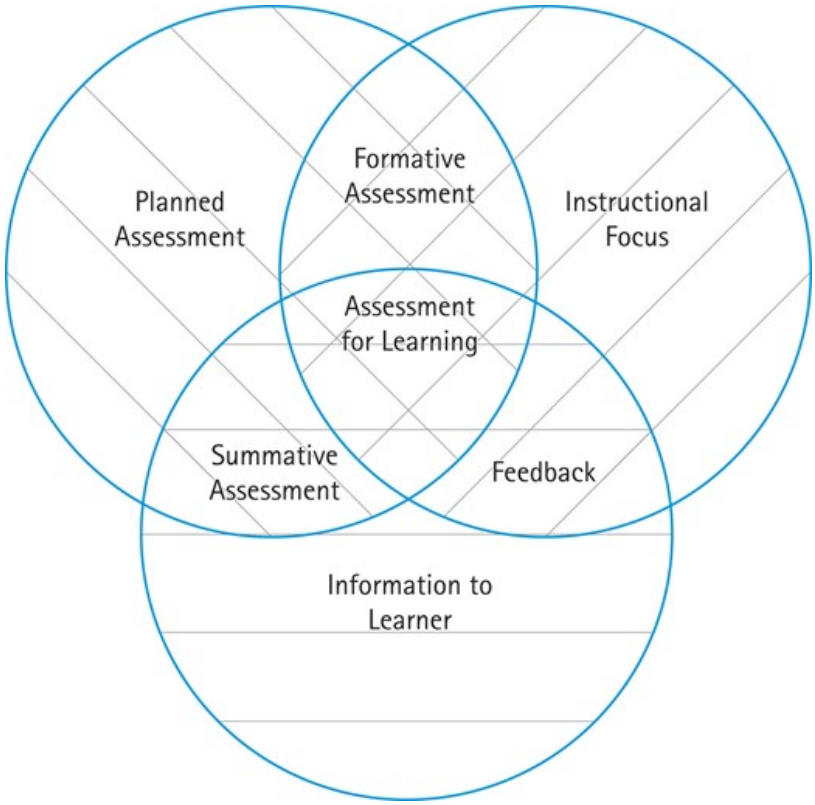
Feedback comes from psychology by way of physics and is a concept that extends well beyond instruction, particularly instruction in classrooms (Ramaprasad, 1983; Kluger & DeNisi, 1996). Formative assessment comes from measurement theory and classroom instruction, and involves issues like reliability and validity, test construction, bias, and measurement error (Bloom, 1968; Bloom, Hastings, & Madaus, 1971; Bloom, Madaus, & Hastings, 1981). Assessment for learning has evolved from more of a philosophical stance on how instruction should take place, and comes from a strong constructivist, and even social constructivist perspective. (Assessment Reform Group, 1999; Broadfoot et al., 2002).

The links among these concepts are presented schematically in Figure 1. There are three basic ideas that combine in different ways to form the concepts of formative assessment, assessment for learning, and feedback. They also provide a definition of summative assessment, or assessment *of* learning. The first idea is whether the assessment is something that is planned, or intended. This does not mean that it has to be some sort of paper and pencil activity. If a teacher is systematically walking around a classroom observing the efforts of the students on a mathematics objective, seeing how well they are doing, then this would be considered a planned assessment. But a planned assessment can be, and often is, a more formal assessment. It might be an essay, a science laboratory, a mathematics test, or a peformance of some sort. Assessment for learning, formative assessments, and summative assessments all involve a notion of planning and intentionality. But feedback does not have to involve a planned activity.

A second idea that differentiates these concepts is whether there is an instructional focus to the activity. By this, we mean that improving instruction for the learner is a central purpose of the assessment. Assessment for learning and formative assessment share this characteristic. Summative assessment does not necessarily include a notion of instructional focus. Summative assessment might be for general

improvement of the school system, or determining whether one should receive a driver's license (or a trophy in archery). The degree to which an assessment has an instructional focus is essentially the degree to which it is formative assessment or assessment for learning. Feedback is a bit tricky on this one. Although one might get feedback in life that does not have an instructional focus (results of a blood test, or a motor vehicle inspection, or even a grade), feedback in schools and learning, which is our concern here, does have an instructional focus.

Figure 1:



Relationships among formative assessment, summative assessment, feedback, and assessment for learning.

Figure courtesy of Jonathan Gutterman.

The third idea has to do with whether information is directed to the individual learner. Here we see that feedback would be directed to the individual learners, and assessment for learning would be as well, but formative assessment might not be. Formative assessment might take place where the information is solely for the benefit of the teacher to guide subsequent instruction (perhaps for the class as a whole). From Figure 1, we can see that the essential characteristics that differentiate formative assessment, assessment for learning, and feedback are whether there is a planned assessment or not, whether the activity is instructionally focused, and whether the activity focuses on the individual learner. Now, in truth, the idea of instructional focus really differentiates summative assessment from the other three, but we include it here for the sake of comprehensiveness.

Having looked at similarities and differences, the question arises as to how to examine somewhat different ideas within one paper. Our approach is based primarily on history. Prior to the work of the Assessment Reform Group, assessment for learning did not exist as a concept, but since the work of Scriven and Bloom, formative assessment has existed. Feedback is much older than both concepts. We are going to examine formative assessment up to the point where assessment for learning came into being. We will then explore the notion of assessment for learning and how it is similar to and different from formative assessment, and

once having done that, we will talk jointly about formative assessment and assessment for learning. When one examines this literature, it is clear that formative assessment and assessment for learning are used pretty much interchangeably. Indeed, if a formative assessment that provides information to individual students, and embodies a belief on the part of the instructional system that the learner can be successful, then it *is* assessment for learning. And feedback that comes from an assessment is part of assessment for learning. Thus, for us, we feel that too strong a concern on the differences and subtleties among these terms is not particularly helpful. We will make distinctions where we feel they are beneficial. We will focus somewhat less feedback as a general concept as it takes us down a rather different, but fascinating path. We strongly encourage Wiliam's (2018) *tour de force* on the topic for interested readers.

The Evolution of Formative Assessment and Assessment for Learning

As mentioned in the introduction to the chapter, there have been a number of extensive reviews of the literature on formative assessment and assessment for learning. Our purpose here is not to summarize or critique that work, but rather to provide what we hope is a clear understanding of the terms, the issues involved, and what the research has to say about practice. We start with Bloom and his pioneering work on formative evaluation (which became formative assessment, and mastery learning theory).

Bloom's Ideas of Formative and Summative Assessment and Learning for Mastery

In the 1960s Benjamin Bloom was working on what he would call Learning for Mastery, which would develop into Mastery Learning Theory (Bloom, 1968, 1971). Bloom was a great synthesizer of ideas; he took Carroll's (1963) notion that learning could be conceptualized as a ratio of the time needed to learn something (the denominator) and the time spent on learning (the numerator). If the time spent learning matched the time needed to learn, learning would occur. Now, this may seem tautological, but it held the important advantage of breaking learning into two distinct factors: time needed to learn and time spent in learning. Time needed to learn something could be attributed to general aptitude to learn as well as prior learnings in the area under consideration. It could also be related to the quality of instruction that a person received. Time spent in learning could be attributed to how much time was allocated to different instructional areas, and how much time could be spent in a given learning task (Bloom, 1968). Bloom brought into this mix the notion that testing could be used for helping a teacher understand just what a student had and had not learned on a given learning task. Armed with an objective sense of what had and had not been learned, the teacher could devise subsequent instruction to address the learning needs and build on the strengths of students. Bloom adapted the notion of formative and summative evaluation of instructional programs recently developed by Scriven (1967) and applied it to evaluating the learning of individual students in a class. Bloom's notion was that initial instruction on a unit (or chapter) in a course could be followed immediately by short, objective tests (either multiple choice or short answer) that would assess the students' knowledge and skills related to the instruction (Bloom, 1969; Bloom, Hastings, & Madaus, 1971; Bloom, Madaus, & Hastings, 1981). The teacher could then score and analyze the performance of the students, and could devise a round of remedial instruction. This would continue until all students had reached a sufficient level of mastery, and then the next unit would be introduced. This naturally would cause instruction to go more slowly at the beginning of a course, but would have the distinct advantage of having all students ready to proceed to unit two instead of having some who were still struggling with unit one. At the end of the course, a summative evaluation would be given that would be used for course grades.

Thus, formative evaluation, and summative evaluation, were part of an overall approach to instruction, one that was a fascinating combination of a behaviorist notion of how people learned with a social constructivist notion that all students could learn what currently only the most advanced learned. It also involved students helping one another and building community, and the employment of testing for informing and guiding

learning instead of solely measuring it for purposes of grading. Bloom et al. (1971) laid out the underlying ideas behind formative and summative assessment in an extensive book on the topic. Guskey (Guskey, 2007; Guskey & Pigott, 1988) reviewed the literature on the efficacy of mastery learning, showing its application in a wide variety of settings.

Criterion and Norm-Referenced Testing and Assessment

At roughly the same time that Bloom was promoting the notion of formative evaluation, Robert Glaser and others (Glaser, 1963; Glaser & Klaus, 1962; Popham & Husek, 1969), were introducing another important distinction, between *criterion-referenced* and *norm-referenced* testing and assessment. Nitko (1980) provided an excellent history and review of the evolution of this idea. The problem that Glaser addressed was how to make meaning out of a test score. If, for example, Marcus receives a 68 on his final examination, how good a score is that? How does he make sense out of it? If he receives information that 70 was the highest score in the class and that the average score was 56, he might feel quite good about his performance. If he was told that he received the lowest score in the class, he would then likely be somewhat distressed. But the score was the same in both cases. By referring (or "referencing") his score to a norm group (in this case, the class), his score is given meaning. Without that information, his score is a rather abstract number.

A second way that Marcus might get meaning out of his score is to refer it to an absolute criterion. If this were Marcus's driving test, and he was told that a 68 was a pass, then he would feel great. It wouldn't really matter to him how well others had done; he had passed the "criterion." The people who were in charge of issuing driver's licenses had determined that his driving ability was sufficiently good to get his license. Presumably, their determination was based on years of experience, and perhaps even research. They had set a criterion which had to be exceeded in order to pass. It didn't matter how many people passed or failed on the day that Marcus took the test. There was no norming group here, but rather a criterion. It's important to note that using either a norm- or criterion-reference approach does not guarantee accuracy or objectivity in a score. If Marcus's 68 was his final grade in a high school course, what would that mean? In many American grading systems, a 68 would be a failure; in others, it might be a high D. In other countries, it might be a B or a second level distinction of honor. Different grading systems give different meanings to the same score.

Notwithstanding any notion of absolute objectivity, the idea of criterion-referencing scores played an important part of the early work in formative assessment and still does today. In Bloom's mastery learning notion, the formative evaluations that he included at the end of each instructional unit were criterion-referenced in nature. An acceptable level of performance was to be set by the teacher, and then all students who met or exceeded that level were deemed to be ready for the next unit. Not only was a norm-referenced interpretation of the scores not necessary, in Bloom's view, it was not desirable. The purpose was not to see who was doing better than whom; it was to see when all students were ready to proceed to the next unit. Thus, although the notion of formative versus summative assessment, and criterion- versus norm-referenced assessment bear some similarities, they are essentially independent frameworks. For example, a criterion-referenced summative assessment might be one where course grades are determined by an absolute standard set before the assessment was given. A criterion-reference formative assessment would be a mastery learning end of unit assessment. Again, an absolute standard was set based on the teacher's determination of what level of mastery was required to proceed (criterion-referenced), but with the intention of guiding future instruction (formative) rather than giving end of course grades (summative).

The Reviews of Crooks (1988) and Natriello (1987), and the Work of Stiggins

At roughly the same time, Crooks (1988) and Natriello (1987) produced extensive and influential reviews of the literature concerning the impact that classroom assessment practices had on achievement in schools. Interestingly, although their articles had almost identical titles, of the over 300 publications that they reviewed (summed across the two articles), only 9 of those publications were mentioned in both works (Black & William, 1998). In the 20 years between Bloom's introduction of the notion of formative evaluation (assessment), and the reviews of Crooks (1988) and Natriello (1987), there had been a focus on assessment in classrooms that had not existed before that time. Prior to that time period, classroom assessment was seen as the domain of instructional theory, not assessment theory. Crooks's review focused on the research that had been conducted on classroom assessment practices, and carefully analyzed a host of different factors that influence the practice and consequence of classroom assessment. Among other findings, Crooks reported that:

- Students pay close attention to classroom assessments if they feel that the assessments represent what the teacher feels is important. Furthermore, the frequency and timing of assessments structure students' activities in their learning, such as frequency of study and review for assessments.
- Classroom assessments typically focus on lower-level objectives, and Crooks called for an increased emphasis on what he called deep learning.
- Assessment should be focused on assisting learning rather than be focused on grading. Although the term "formative assessment" does not appear in the article, clearly this is what Crooks was calling for.
- Feedback is important in focusing student learning and the need for it to be delivered in a timely fashion.
- Setting high but attainable standards increases student achievement, particularly in mastery learning settings (Kulik & Kulik, 1987).

Natriello (1987) took a rather different approach to the question from Crooks (1988), focusing his review on a more structural and prescriptive approach to evaluation in classrooms. He presented a model of a desired approach to classroom assessment, and called for more research that took a wholistic view of the evaluation process. Natriello's review emphasized process issues more than the effects of those processes. An interesting aspect of both reviews is that they do not mention "formative assessment." Although Crooks discussed the notion of assessment designed to assist learning, most of the studies he (and Natriello) reviewed do not differentiate between formative and summative assessments. This is particularly remarkable because of the strong influence that the consequence of an assessment has on the behaviors of those being assessed (Wolf & Smith, 1995; Wolf, Smith, & Birnbaum, 1995). Formative assessment, particularly in Bloom's approach to it, does not count toward a student's grade and therefore is not as consequential to the student (with regard to a grade), whereas a summative assessment would be. This lack of focus on the fundamental nature or purpose of the assessment is fairly commonly seen in the 1970s and 1980s. This changes in later work, as will be seen. It should be noted however, that some scholars were taking the student perspective on assessment as early as the mid-1970s. Easley and Zwoyer (1975) stated:

> If you can both listen to children and accept their answers not as things to just be judged right or wrong but as pieces of information which may reveal what the child is thinking you will have taken a giant step toward becoming a master teacher rather than merely a disseminator of information." (p. 25)

Into this zeitgeist, Stiggins and his colleagues (1991, 1985; Stiggins & Bridgeford, 1985; Stiggins, Conklin, & Bridgeford, 1986) repeatedly made the case for greater attention to be paid to how we assess student growth

in classroom settings. They argued early on for a student-centered approach to classroom assessment practice with a special emphasis on the needs and growth of the student. Stiggins has maintained this concern over 40 years of work. Basically, Stiggins makes the point that the student is where the learning resides; it is where the growth takes place, and therefore assessment, which is intended to facilitate that growth, should always be closely aligned to the student's needs and interests. Although this seems like something new, the same sentiment can also be found in Dewey (1902). Stiggins illustrated this perspective beautifully in a personal description of how he came to understand that he could be a successful learner after years of learning difficulty (Stiggins, 2018). He explained that after repeated failure in traditional schooling, he found success in training to become a US Air Force mechanic because his instructors were not interested in rank ordering the students in the class. Their goal was simple: train *all* the students to become excellent mechanics. Once a teacher takes that perspective and works on what that means, assessment takes on a new perspective, one whose sole purpose is to facilitate learning.

One final review should be mentioned here, and that is the work of Fuchs and Fuchs (1986), who presented an excellent meta-analysis (quantitative summative analysis of original empirical research) on the effects of feedback to learners with learning disabilities. They found that frequent presentation of feedback had a strong positive effect on these learners, and that presenting learners with visual displays of their learning progress was highly effective.

The Assessment Reform Group and The Black Box Series

In 1989, the British Educational Research Association formed the Assessment Reform Group, consisting of a set of highly respected scholars in the field of educational measurement, and charged them with looking at how to reform assessment practice in the United Kingdom. As mentioned in the outset of the chapter, they took on the task of looking carefully at how assessment can be effective in promoting learning (Assessment Reform Group, 1999; Black, Harrison, Lee, Marshall, & Wiliam, 2003; Broadfoot, Daugherty, Gardner, Harlen, James, & Stobart, 2002). Their work was in part an effort to embed assessment into a view of learning in which (among other ideas) teachers were confident of their students' ability to learn; that saw pupils not soley as recipients of feedback on their efforts, but critically engaged in the processes and outcomes of the assessment as partners with their teachers; and which provided explicit information on how the learner could close the gap between current status and desired goal. Interestingly, these ideas intersect with Bloom's (1968) mastery learning theory, althought there are other aspects of both that do not align. The work of the Assessment Reform Group and its members working individually and in smaller groups, resulted in something of a "sea change" in thinking about the use of assessment to enhance learning. They coined the term "assessment for learning," and were instrumental in bringing about a more direct student focus on assessment in classrooms.

Particularly influential was a piece by Assessment Reform Group members Black and Wiliam (1998) called, "Inside the Black Box: Raising Standards Through Classroom Assessment." Part of a series called "The Black Box" series by the Assessment Reform Group, this piece brought the ideas behind formative assessment to practicing educators, and is particularly valuable for getting a sense of the state of play in the development of the ideas around formative assessment at that time, as well as foreshadowing the introduction of *assessment for learning* as an alternative to *formative assessment*. Other work by the Assessment Reform Group presented the results of studies conducted looking at the efficacy of assessment for learning efforts (see, e.g., Wiliam, Lee, Harrison, & Black, 2004). Although later the argument would be made that there are important differences between assessment for learning and formative assessment, the Wiliam, Lee, Harrison, and Black article uses the phrase, "... increased use of formative assessment (or assessment for learning) ..." (p. 49). Although the Assessment Reform Group disbanded in 2010, the members have been leading scholars writing in this field for decades. Their work has explored peer and self-

assessment, the use of summative measures as formative measures, and teacher questioning as formative assessment (Black, Harrison, Lee, Marshall, & Wiliam, 2004; Black & Wiliam, 2012; Wiliam, 2018).

More Recent Reviews

Reviews of formative assessment and assessment for learning have continued apace in the past 20 years. In addition to general reviews (Bennett, 2011), there are a number of specialized reviews, such as reviews of prerequisites for implementing assessment for learning (Heitlink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016), computer-based assessment for learning (Shute & Rahimi, 2017), subject area-based reviews (Hodgson & Pyle, 2010), and reviews at different levels of instruction (e.g., Nusche, 2008). One of the changes that can be seen in these and other reviews is that there is a shift from thinking about formative assessment as an activity or document (a test) toward thinking of it as a process of collecting information about a learner, interpreting it, and then providing feedback to the learner. That is, it becomes more a process and less a "thing" or one-off event.

Shute (2008) employed the somewhat novel term "formative feedback" to discuss information that is provided to learners for the purpose of improving their learning. It is interesting to note that she does not include the process of collecting information in that definition, although most of the research she reviews involves some level of systematic data collection. This is more than a passing concern as the development, administration, and interpretation of formative assessments are key components of such assessments. It is also worthwhile noting that Shute was concerned with task-level feedback as opposed to more general feedback. This is important in two respects: first, it is a simpler undertaking to work toward growth on a particular task than on a broader skills; and second, when looking at growth on a task that is currently being directly taught, the impact of feedback is more likely to be clearly observed than when looking at broader issues. Wiliam (2018) makes the important distinction between performance (on the current task) and learning, which is generalizable to related tasks. Shute (2008) found that formative feedback is most effective when it is "non-evaluative, supportive, timely, and specific" (p. 153). In a particularly powerful metaphor, she cleverly likens formative feedback to a good murder:

"Formative feedback" might be likened to "a good murder" in that effective and useful feedback depends on three things: (a) *motive* (the student needs it), (b) *opportunity* (the student receives it in time to use it), and (c) *means* (the student is able and willing to use it)" (p. 175).

Kingston and Nash (2011) addressed the issue of just how effective formative assessment is and under what conditions it is most effective using a technique called meta-analysis. Meta-analysis is basically a procedure for quantitatively synthesizing empirical studies. It takes the results of a group of studies looking at the same research question, and then estimates how much better the group getting the treatment (in our case, formative assessment or assessment for learning) did compared to the control group. The difference in the means between the two groups is transformed into a standard deviation metric. That is, the difference is expressed in how many standard deviations better the treatment group did than the control group. A standard deviation is a standardized metric for looking at the spread of scores in a distribution of scores. In a prototypical normal curve, 96% of the scores would fall between two standard deviations below the mean and two standard deviations above the mean. The difference between two means in a research study expressed in standard deviation units is called the "effect size" of the study. Kingston and Nash look at the claim in Black and William (1998) "that the effect size for testing feedback is no lower than .7 standard deviations" (Kingston & Nash, 2011, p. 28). Kluger and DeNisi (1996) had found an effect size of .80 and other researchers found effect sizes for feedback of around .40 (e.g., Hattie & Timperley, 2007). Thus, while there is some degree of unanimity that formative assessment/assessment for learning enhances learning, there is wide disagreement on just how strong that effect is. And indeed, different approaches in different settings would be expected to be more or less efficacious.

Kingston and Nash re-examined some of this research using more stringent criteria for including studies into their work and found an overall effect size of .25 was a better estimate of the effect of formative assessment. Also, they noted that Kluger and DeNisi found that in one-third of the studies in their research, feedback had a negative impact on achievement. So, why the discrepancy in the findings? A large part of the differences seen in these meta-analytic studies concerns differences in the studies being examined. Kingston and Nash were much more stringent in determining which studies were of high-enough quality to include in their analyses. Some of differences found have to do with what is being learned (e.g., a study of learning one's times tables in mathematics versus developing a sense of irony in one's story writing ability), and some of the differences have to do with the nature of the formative assessment and consequent feedback being given. There is a world of difference between considering giving a grade or not being feedback (as in some of the studies in the Kluger and DeNisi (1996) analysis), and detailed, personalized recommendations for growth found in other studies. Thus, combining various studies of formative assessment and feedback quantitatively can run into technical concerns and disagreements over what should be included and what should not be.

In their review of feedback (and not specficially formative assessment or assessment for learning), Hattie and Timperley (2007) examined the power that feedback has to influence learning in educational settings. As they point out, this feedback typically follows instruction and provides information that will help learners in their next steps. Thus, the notion that some form of formative assessment has taken place between the instruction and the feedback is reasonable to infer. This extensive and highly influential review argues that the three essential questions that feedback attempts to answer, from the perspective of the learner are the following:

"Where am I going? How am I going? and Where to next? The answers to these questions enhance learning when there is a discrepancy between what is understood and what is aimed to be understood" (Hattie & Timperley, p. 102). Although these questions can be seen in different form in literature going back to at least Ramaprasad (1983), they are restated in Hattie and Timperley with clarity and succinctness.

Hattie and Timperley additionally offered a hierarchy of feedback messages that is particularly useful. Their first level of feedback concerned the task under consideration. Teachers may inform the learner on whether the task has or has not been successfully completed, or may offer more extensive feedback and correctives about task performance (such as, "You seem to have forgotten to carry the 1 here," or "Your subject and verb don't agree."). The second level concerned the processes that underlie the task. This type of feedback provides information on how the learner can improve performance on the task. It might be of the type, "Remember to look to see if the answer seems reasonable to you," or "Try putting the subject right next to the verb to see if they agree in number." The third type of feedback concerned the metacognitive processes associated with working on the task. It would be used to suggest broad, self-reflective engagement that would generally lead to improved performance. Examples of this type of feedback might include, "When solving math problems, always check to see if your answer seems reasonable given the nature of the problem," or "It can be really helpful to put your writing aside for a day or two and then read it afresh to see if you are really saying what you want to be saying." The fourth level of feedback had to do with the individual learners and their characteristics. Examples here would be, "You're really good at mathematics," or, "Writing is really what you are best at." Hattie and Timperley argued that this level of feedback is typically not beneficial to students and is actually more likely to be harmful as it causes a focus on the student and not the learning.

Bennett (2011) provided an excellent analysis of the issues in formative assessment and assessment for learning, critically examining six different concerns related to the topic. He looks at "the definition of formative assessment, the claims commonly made for its effectiveness, the limited attention given to domain considerations in its conceptualization, the underrepresentation of measurement principles in that conceptualization, the teacher-support demands formative assessment entails, and the impact of the larger

educational system" (Bennett, 2011, p. 5). In considering Bennett's concerns, one can see the continued development of formative assessment/assessment for learning as a field. The questions being asked are those of a maturing, if not fully matured, sub-discipline at the intersection of measurement and classroom instruction. Based on his review, Bennett concluded that there is still something of an orthodoxy conflict over the definition of the term, "formative assessment," and that claims for its effectiveness are somewhat suspect. He called for work on defining formative assessment more rigorously and the utilization of strong measurement principles in the development of such assessments. He also argued that teachers need more training in order to become proficient in formative assessment.

Formative Assessment and Assessment for Learning Today

Formative assessment and assessment for learning are very different creatures than they were in the late 1960s and early 1970s when the terms were first introduced. They have changed from being primarily about developing tests that could be used by teachers to guide subsequent instruction to being an integral part of instruction itself, and informing students directly as a primary focus of their existence. In this section we look at the issues regarding these terms and the related concept of feedback. We break these issues down as follows: defintions of the terms, incorporation into classroom instruction, how feedback figures in formative assessment and assessment for learning, and the shift to a learner-centered focus.

Definitions of the Terms

Today, formative assessment and assessment for learning are used, for the most part, interchangeably. When one reads either term, it is not certain what is about to be discussed. The notion that something may be formative assessment, but not assessment for learning, simply does not exist in the literature. There are arguments that perhaps formative assessment is something that looks like a test, and assessment for learning is an ongoing process, but that distinction seems to be fading, and we will not attempt to revive it here. Broadly speaking, formative assessment/assessment for learning is a process through which information (1) is obtained about a learner's level of achievement (ability, knowledge, skill, performance) in a particula area of learning, (2) analyzed and intepreted in terms of where learner is and what he/she needs to do in order to improve and how the teacher can facilitate that learning, and (3) is received and interpreted by the learner who can then act on the feedback in his/her efforts to learn. The idea that these assessments are primaily tests has also given way to a broader set of approaches that might be used to gain information that is helpful to students and teachers. As a process, formative assessment/assessment for learning may involve something that looks like a traditional classroom test, but it does not have to. It could also be a laboratory report, an essay, a dance performance, or even something as subtle as a smile of recognition on the part of a student as a teacher asks, "Does that make sense?" One might argue that we have removed the notion of "assessment" in this definition. Perhaps from a measurement theory perspective, we are guilty, but we would argue that the underlying idea of "assessment" simply does not equate to "test." There are alternatives for gathering information about student learning, and all such approaches are appropriately, assessment.

Incorporation Into Classroom Instruction and Learning

Several important shifts have occurred with regard to formative assessment/assessment for learning as it is realized in classrooms. Perhaps the most important shift is the focus on who receives the information garnered by the process. We can see in Bloom's mastery learning model (Bloom, 1968), which is where formative assessment started, the teacher would analyze the data from the formative assessment and use it to develop feedback and correctives for the students. The teacher might also decide to re-teach certain portions of an instructional unit if a number of students seemed to be having trouble with it. Today, the recipient is much more likely to be the individual student. The teacher is likely to be involved in evaluating the assessment and providing the feedback, but his/her focus is much more likely to be one student at a time rather than the class as a whole.

Sadler (1989) argued for an instructional approach that emphasizes the ability of students to monitor the quality of their work themselves rather than relying continually on the feedback and correctives provided by the teacher. He contended that the instructional system must provide the means for the student to develop these self-monitoring and evaluative skills. Black and Wiliam (2018) laid out a model of instruction wherein formative assessment plays a critical role. Their model included making sure that (1) students understand clearly what is to be learned and what the goals of instruction are, (2) effective instruction takes place, (3) students received formative feedback that progesses their learning, (4) students collaborate in their learning, and (5) students take a sense of ownership in their learning.

This brings us to the next important change in formative assessment/assessment for learning, at least in the research on the topic if not also widespread in practice, and that is who delivers the feedback. Teachers provide feedback for sure, but so do computers (Shute & Rahimi, 2017), peers (Panadero & Alqassab, 2019), and even the individual students themselves (Andrade & Valtcheva, 2009). It could even be an inanimate object such as the arrow in the opening vignette of the chapter. Another important advancement is that assessments are more likely to be focused on delivering useful feedback to students from the design all the way through to the delivery of the information that learners receive. The whole purpose of the assessment is to get the information needed to provide feedback. This is a shift from a focus on providing a summative score that would indicate to the students how well they did. Frequently, what learners will receive may not even have a grade or score on the assessment.

The Influence of Feedback

An inherent, if often unspoken, component of formative assessment/assessment for learning is that the goal of the process (today) is to provide feedback to the student on his/her learning efforts. This being the case, the idea of feedback, and the research that has been conducted on it, comes into play. And we are just beginning to learn how complex and multfaceted a topic that is. How much feedback should be given on any one assessment; how complex should that feedback be? Is providing students with tools and asking them to generate feedback enough, or are teacher comments still needed? How can we be sure that the student will understand the feedback? What will the student's emotional reaction to the feedback be? Do students really want feedback or do they just want to see their grade? How does the use of praise affect the receipt of feedback? Should feedback address all issues on an assessment, or should teachers be selective in what is delivered at any given point in time, and should different feedback be given to learners depending upon their particular circumstances? Should feedback be focused on what has occurred in the assessment itself, or on the learning trajectory of the students and be focused on the future? Formative assessment/assessment for learning and feedback are inextricable. Essentially, feedback is the benefit that derives from formative assessment activities. Without the provision of feedback, ideally to the learner, but at least to the teacher, formative assessment is pointless. Feedback is why we do it.

Turning the Focus to the Learner

A final aspect of formative assessment/assessment for learning is a turn away from the teacher as the focus to the learner as the focus. Revisiting Bloom's (1968) mastery learning model, we see that the basic purpose of formative assessment was to guide the teacher in developing additional instruction and providing feedback and correctives to learners. Now, to be certain, that feedback and corrective material was intended to vary according to the needs of the individual learner, but the teacher was in control of the situation, and whole-class-based remediation was an important part of the process. Looking at Sadler's (1989) work, we see the emphasis changing to the learner, and Black and Wiliam's (2018) model truly emphasized the centrality of the learner in learning. There are many other examples of scholars and practicioners calling for a renewed focus on the learner in assessment.

More specifically, what students do with the feedback they receive and how different variables affect this process has become a central topic in recent years (Panadero & Jonsson, 2013; Winstone et al., 2017; Van der Kleij & Lipnevich, 2020). Hattie has recently moved from examining characteristics of feedback as proposed by his model with Timperley, to exploring conditions for the students' specific use of feedback (Hattie & Clarke, 2019) and there has been a notable increase in research into the role of student in the process of feedback with two systematic reviews (Panadero & Jonsson, 2013; Winstone et al., 2017), an integrative review (Panadero & Jonsson, 2013) and a scoping review of students' perceptions of feedback (Van der Kleij & Lipnevich, 2020) appearing in the last few years. Jonsson (2013) identified five challenges for the use of feedback in higher education students: (1) feedback needs to be useful; (2) students prefer specific, detailed, and individualized feedback; (3) authoritative feedback is not productive, students may lack (4) strategies for productive use of feedback; and (5) an understanding of academic terminology or jargon.

Similarly, Winstone et al. (2017) presented the idea of *proactive recipience* of feedback. The researchers proposed a model with three factors influencing recipiency. First, feedback interventions should focus on internalizing and applying standards, sustainable monitoring, collective provision of training and manner of feedback delivery. Second, they proposed the SAGE recipiency processes coming from Self-appraisal, Assessment literacy, Goal-setting and Self-regulation, and Engagement and motivation. And, third, they described interpersonal communication variables that attend to the characteristics of the receiver, sender, message, and context.

Finally, in 2018 Jonsson and Panadero synthesized research, taking as central the feedback model by Lipnevich et al. (2016) and discussing factors moderating students' engagement with feedback. They finalized their review summarizing research within three general pedagogical conditions. First, feedback needs to be perceived as useful by the students and this can be achieved via satisfying three conditions: (1) students should have the opportunity to use it, (2) information should be simple and accessible enough that students can act upon it, and (3) feedback needs to be understandable. Second, students need strategies for using the feedback and, third, feedback should be delivered without a grade (to enhance its formative effect). In their model, students are the focus of the feedback process and focusing on their use of information may be the right direction for the field.

In sum, what we see in formative assessment/assessment for learning today is a strong focus on the student, a much broader view of what might constitute an assessment, increasing use of peer and self-assessment, a more social constructivist (as opposed to behaviorist) philosophical perspective, less use of marking (grading), and praise with a concommitant increase in commentary and suggestions for what can be done to improve upon current status.

Looking at the Future of Formative Assessment and Assessment for Learning

Given the remarkable growth in knowledge concerning formative assessment/assessment for learning, and its potential for enhancing learning, there are still mountains to climb. In this closing section, we look at some of the areas of research that we think are most important and most promising.

Performance versus Learning

Wiliam (2018), in discussing the literature on feedback, argued that there is a critical difference between performance on a given task and learning. Much of the research on feedback and formative assessment, as well, tends to focus on the immediate task at hand—such as an essay—and improving that particular task as much as possible. This is what Wiliam calls "performance." Improvement on a given essay is good, but it does not necessarily mean that it will result in improved performance on the subsequent essay produced by that student. And in all honesty, probably only the teacher and the student will be the readers of the essay under consideration. Thus, argues Wiliam, should we not be more concerned about future performance than that of the recent past? This leads to the notion of feed-forward instead of feedback. We have only mentioned feed-forward briefly as to not overly complicate the discussion. We note that the term did not start in educational research (see, e.g., Dunn, Gillig, Ponsor, Weil, & Utz, 1986). But, the underlying idea is worthy. When we provide feedback to a student, shouldn't we be concerned about what happens next with the student—what permanent changes can be encouraged—rather than making sure all the subjects and verbs agree in the essay at hand? This is what Wiliam (2018) is arguing for and we agree. Thus, one of the issues that is gaining in importance, appropriately we think, is an emphasis on long-term gains as opposed to more temporal ones.

How Do We Say What We Say?

It is one thing to say that we should be student-centered in providing feedback to students on their formative assessments, and another thing to know how to do that. Johnson (2004) makes a convincing argument that the particulars of how we engage learners verbally can have dramatic effects on how they hear what we say. This goes not simply to the question of whether learners understand what teachers are saying (Chanock, 2000), but also how they respond from an affective perspective (Hyland, 2013). We think it is important to consider that in many educational settings, especially in secondary and tertiary institutions, written feedback on an assignment may be the primary, or even sole, communication that a student receives from an instructor. Thus, it is critical that we better understand how those messages are received (Ivanic, Clark, & Rimmershaw, 2000). Consider three possible comments that could be made on the same grammatical error in an essay:

1. "SUBJECT/VERB AGREEMENT!!,"
2. "Read this sentence aloud to yourself and see if you can find an error here," and
3. "This is a tricky one, Joshua, but say the subject and the verb without the phrase in between them."

These comments vary in tone, personalization, the level of the feedback (see Hattie & Timperley discussion earlier), and the attitude that the instructor takes toward the student. They all concern the same error, but they are likely to be received differently. How do we know what to say to whom? This is an area where we think much more research will be highly beneficial.

Issues of Reliability and Validity

It is important to not lose sight of the fact that formative assessments and assessments for learning are assessments, and should be of high quality. With the loosening of what might be considered an assessment, it is possible to lose sight of the fact that we need the information used to be valid. Or, more properly, as Cronbach (1971) noted, "One validates, not a test, but *an interpretation of data arising from a specified procedure*" (p. 447, emphasis in original, taken from Black and Wiliam, 2018). The argument that it is the interpretation of the data that is the critical factor is no less true today than it was half a century ago: if we are to make an inference about a student's learning, we want that to be a correct inference. The importance of the quality of the assessments that we use to draw those inferences has been emphasized by Black and Wiliam (2018), Bennett (2011), Stiggins (2008), and Kane (2006), among others. Although much has been written about the reliability and validity of measures, it is worth noting that these terms take on somewhat special meaning when considering formative assessments/assessments for learning. To begin, validity is the key issue here. Reliability is basically a necessary but not sufficient condition for validity. If a measure is to be valid, it has to be reliable. It is analogous to "antique" and "old." But, whether one's assessment is a multiple-choice test, an essay exam, or the observation of a series of free throws taken in basketball, Cronbach's admonition applies. We might translate his words into, "Are we drawing the right conclusions about this student based on this information?" Thus, our information has to be the right information for the setting. It has to lead us to the proper conclusions. Reliability in this context might be simplified to the notion of "Do we have enough information to draw a valid conclusion?" (For example, have we seen enough free throws by this player to be able to see what is being done right and wrong? What should we say to the player, or demonstrate, to promote improved shooting?) Smith (2003) provided a useful discussion of the issue of reliability with regard to classroom assessment, arguing that "sufficiency of information" is the essence of reliability in classroom assessment. Validity would then ask the question, "And is that information the right information for drawing proper inferences?"

Murray, Gasson, and Smith (2018) examined the grades and comments given by a set of tutors in a nursing course on an essay that was assigned to help develop the writing skills of the students. Five tutors each marked the same set of five essays in the course in which they were teaching. Since the purpose of the assignment was primarily formative, extensive comments were written on the essays; however, marks were also given as was the practice in the course. Consistent with general research on essay marking, the grades assigned to students varied substantially by tutor (ranging from 54 to 85 out of 100 on one of the essays). But, more pertinent to the discussion here, the number of comments made on the papers varied dramatically (ranging from 19 to 63 on one of the essays) as did the nature of the comments. On one paragraph, one tutor said, "This is a better paragraph"; a second said, "This could be explained more clearly —it is a bit confusing"; and a third said, "Muddly." To a large degree, what was received by the students was more a function of which tutor marked their paper than what they had written. We know there are issues here, but how to address those problems is not obvious. Teachers have limited time and are typically not well-trained in the development and use of formative assessments.

Considering Effiency in Assessment

There are only so many hours in a teacher's day. One of the authors of this chapter, after having presented a workshop on how to glean the most information out of a student's assignment, received the following question from one of the teachers in attendance, "You do realize I have 120 students, don't you?" The approach that had been delivered would have taken about 15 minutes per pupil. With 120 students, that would have required 30 hours of work—with no breaks—to interpret and provide feedback on one assignment. Providing such detailed feedback on an ongoing basis is simply not feasible. And so, it is imperative that more efficient forms of providing feedback be explored. Efficiency is generally considered something of a dirty word in education, but if our goal is to maximize learning, we must take into consideration how to most wisely use the time and resources of both teachers and students. Lipnevich, McCallen, Miles, & Smith (2014) examined the use of providing students with detailed rubrics and exemplars of good, average, and poor work in enhancing the students' work in writing research design proposals. The students preferred receiving exemplars of good work over rubrics (and over average and poor work). Analysis of their second drafts of their research proposals showed that while exemplars improved performance by about a half a standard deviation, provision of rubrics improved performance by a full standard deviation. Although arguably neither the exemplars nor the rubrics are formative assessments nor feedback, they have a tendency to serve in a feedback role. Students receive the materials and then compare their work to the exemplars, or to the scoring rubrics. In doing so, the students engage in self-assessment of their work, evidently quite effectively, and generate inner feedback (Nicol & McCallum, 2021). Andrade and her colleagues have also seen the benefits of providing rubrics to students in their learning efforts (see, e.g., Andrade, Du, & Wang, 2008), as have Price, Berg, and Smith (2017). Other alternatives for making formative assessment more efficient is the use of computers, peers, and self-assessment. For example, computers can provide worthwhile feedback on student work in a variety of settings, including multiple-choice responses, computer-guided instructional programs, and computer-generated feedback on essays.

Looking at the Process of Assessment and Feedback

The role of formative assessment/assessment for learning in instruction is complex. It affects not only learning directly, but affective considerations as well. For example, Beatson, Berg, and Smith (2018) found that level of mastery on a midterm examination was related not only to performance on the final examination, but also on the students' sense of self-efficacy. A number of authors have tried to build formative assessment/assessment for learning into a model. Several of those efforts (e.g., Bennett, Wiliam) have alread been discussed. Recently, the authors of this chapter have developed a process model of how feedback is generated, delivered, received, and worked on by students (Lipnevich, Berg, & Smith, 2016). Prominent in the model was the concern for the affective response to the feedback that is received. If the student feels threatened, not respected, offended, patronized, discouraged, or any of a host of other negative responses, even the best-intended feedback message will simply not be received. The affective dimension of feedback is a critical aspect for future research. It is important to have data on this issue, and not simply opinions of scholars.

Concluding Remarks and Where to From Here?

It is an exciting time for those scholars interested in questions of formative assessment/assessment for learning. It is also a somewhat dangerous time as well. Formative assessment, assessment for learning, and feedback have been promoted as perhaps the most important aspect of the educational process in terms of improving student learning (Hattie, 2012). But other aspects of learning have been promoted as panacea before, and ultimately fallen short. It is without question that the formative assessment/assessment for learning/feedback triad has a much stronger research base than what has gone on before, but at the same time, it is clear that it is not a silver bullet. Formative assessment/assessment for learning has to be employed intelligently, with due concern given to the environment in which students are learning, the nature of the subject area, the developmental level of the students, etc. And, as can be seen in the Murray, Gasson, and Smith (2018) study, not all feedback is the same.

So, where to from here? As researchers who have spent many years working on these issues and discussing them with colleagues, we believe that three ideas will best lead us into a productive future: (1) more experimental research in real instructional settings, (2) more emphasis on learning as opposed to performance, and (3) a concern for the vicissitudes of classroom life and the lives of teachers and learners.

The first idea comes from looking at a wide variety of research studies and seeing that the ones that really progress the field are those that combine the rigor of experimental methods (internal validity) with the generalizability of ecological verisimilitude (external validity). That is, we think the best studies are those that can eliminate alternative hypotheses to the findings effectively, and that take place in real classrooms with real teachers on topics that are of importance to students and teachers alike. Even if those studies occasionally suffer from small sample sizes, we think it is more effective to accumulate research over those studies than to rely on work that may not be reliable from an internal validity perspective—where we cannot be sure what caused the results. This will require researchers and practitioners to work together, each bringing their special expertise to research projects while respecting the expertise of the other. As a corollary to this (or perhaps a precursor), we strongly argue for research over advocacy. There is no substitute for knowing what works as opposed to "knowing" what works.

The second idea is that we need more research that examines learning as opposed to performance (see Wiliam, 2018). That is, what is important is not whether one's current lab report, math homework, or competitive dive is successful; what is important is whether something had been learned that will go with the learner into the future. If a student produces a piece of work, gets that work evaulated, is provided with recommendations on how to improve that piece of work, and then does so, that is great. But it does not necessarily mean that the student will produce improved work on his/her next effort. They might, but unless we see that future performance, we won't know if that will occur, or to what degree.

The third idea is to appreciate real classroom life. This means taking into consideration teacher expertise, how much time is available, competing interests in the classroom, and the lives of teachers and students. Although having extensive feedback from a trusted source provided honestly yet compassionately is perhaps an ideal, how often can it be realized? And if the student receiving it has recently been bullied, or has a difficult home life, or unrealized learning challenges, then even the best program of formative assessment/assessment for learning might not be effective. Researchers must figure out how to take such issues into consideration in developing our work in understanding formative assessment/assessment for learning. In working with teachers on the issues under discussion in this chapter, we need to take into account what great teachers are great at, and what they might be less skilled at. In thinking about the best teachers we have ever had and what they were like, almost no one says, "She could really make a good formative assessment." This is not a mystery: great teachers are great at working with students.

One final thought. In addition to strong research, the field needs strong thinking. It needs ideas that can advance our understanding and lead to better practice. Ideas could come from the chalkface or from the ivory tower. Whence doesn't matter; wherefore does.

References

Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, *27*, 3–13. doi:10.1111/j.1745- 3992.2008.00118.x
Google Scholar WorldCat

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, *48*(1), 12–19, doi: 10.1080/00405840802577544
Google Scholar WorldCat

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. London, UK: Qualifications and Curriculum Authority.
Google Scholar Google Preview WorldCat COPAC

Beatson, N. J., Berg, D. A. G., & Smith, J. K. (2018). The impact of mastery feedback on undergraduate students' self-efficacy beliefs. *Studies in Educational Evaluation*, *59*, 58–66. doi: 10.1016/j.stueduc.2018.03.002
Google Scholar WorldCat

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25.
Google Scholar WorldCat

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning*. New York, NY: McGraw-Hill.
Google Scholar Google Preview WorldCat COPAC

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in classrooms. *Phi Delta Kappan*, *86*(1), 9–21.
Google Scholar WorldCat

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

Black, P. J., & Wiliam, D. (2012). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning* (pp. 9–25). Thousand Oaks, CA: SAGE Publications.
Google Scholar Google Preview WorldCat COPAC

Black, P. J., & Wiliam, D. (2018). Classroom assessment and pedagogy. Assessment in *Education: Principles, Policy & Practice*, *25*(6), 551–575. doi: 10.1080/0969594X.2018.1441807
WorldCat

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment (UCLA-CSIEP)*, *1*(2), 1–12.
Google Scholar WorldCat

Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Educational evaluation: New roles, new means* (National Society for the Study of Education Yearbook, Vol. 68, Part 2, pp. 26–50). Chicago, IL: University of Chicago Press.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (Eds.). (1971). *Handbook on the formative and summative evaluation of student learning*. New York, NY: McGraw-Hill.
Google Scholar Google Preview WorldCat COPAC

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). *Evaluation to improve learning*. New York, NY: McGraw-Hill.
Google Scholar Google Preview WorldCat COPAC

Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge, UK: University of Cambridge School of Education.
Google Scholar Google Preview WorldCat COPAC

Carroll, J. (1963). A model of school learning. *Teachers College Record*, *64*(8), 723–733.
Google Scholar WorldCat

Chanock, K. (2000). Comments on essays: Do students understand what tutors write? *Teaching in Higher Education*, *5*(1), 95–105.
Google Scholar WorldCat

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 443–507). Washington, D.C.: American Council on Education.
Google Scholar Google Preview WorldCat COPAC

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438–481.
Google Scholar WorldCat

Dewey, J. (1902). *The child and the curriculum*. Chicago, IL: University of Chicago Press.
Google Scholar Google Preview WorldCat COPAC

Dunn, T. G., Gillig, S. E. Ponsor, S. E., Weil, N., & Utz, S. W. (1986). The learning process in biofeedback: Is it feed-forward or feedback? *Biofeedback and Self-Regulation*, *11*(2), 143–156.
Google Scholar WorldCat

Easley, J. A., & Zwoyer, R. E. (1975). Teaching by listening—Toward a new day in math classes. *Contemporary Education*, *47*, 19–25.
Google Scholar WorldCat

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation—A meta- analysis. *Exceptional Children*, *53*(3), 199–208.
Google Scholar WorldCat

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes (1963). *American Psychologist*, *18*, 519–521.
Google Scholar WorldCat

Glaser, R., & Klaus, D. J. (1962). Proficiency measurement: Assessing human performance. In R. Gagne (Ed.), *Psychological principles in systems development* (pp. 419–474). New York, NY: Holt, Rinehart & Winston.
Google Scholar Google Preview WorldCat COPAC

Guskey, T. R. (2007). Closing achievement gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery." *Journal of Advanced Academics*, *19*(1), 8–31.
Google Scholar WorldCat

Guskey, T. R., & Pigott, T. D. (1988). Research on group-based mastery learning programs: A meta-analysis. *The Journal of Educational Research*, *81*(4), 197–216.
Google Scholar WorldCat

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York, NY: Routledge.
Google Scholar Google Preview WorldCat COPAC

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.
Google Scholar WorldCat

Hattie, J., & Clarke, S. (2019). *Visible learning: feedback*. Oxfordshire, UK: Routledge.

Google Scholar Google Preview WorldCat COPAC

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, *17*, 50–62.
Google Scholar WorldCat

Hodgson, C., & Pyle, K. (2010). *A literature review of assessment for learning in science*. Slough, UK: Nfer.
Google Scholar Google Preview WorldCat COPAC

Hyland, D. F. (2013). Student perceptions of hidden messages in teacher written feedback. *Studies in Educational Evaluation*, *39*, 180–187. doi: 10.1016/j-stueduc.2013.06.003
Google Scholar WorldCat

Ivanic, R., Clark, R., & Rimmershaw, R. (2000). What am I supposed to make of this? The messages conveyed to students by tutors' written comments. In M. Lea, & B. Stierer (Eds.), *Student writing in higher education: New contexts* (pp. 47–65). Buckingham, UK: Open University Press.
Google Scholar Google Preview WorldCat COPAC

Johnson, P. H. (2004). *Choice words: How our language affects children's learning*. Portland, ME: Stenhouse Publishers.
Google Scholar Google Preview WorldCat COPAC

Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, *14*(1), 63–76.
Google Scholar WorldCat

Kane, M.T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education/Praeger.
Google Scholar Google Preview WorldCat COPAC

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, *30*(4), 28–37.
Google Scholar WorldCat

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention Theory. *Psychological Bulletin*, *119*(2), 254–284.
Google Scholar WorldCat

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, *47*(1), 211–232.
Google Scholar WorldCat

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*(4), 279–308.
Google Scholar WorldCat

Kulik, C-L. C., & Kulik, J. A. (1987). Mastery testing and student learning: A meta-analysis. *Journal of Educational Technology Systems*, *15*, 325–345.
Google Scholar WorldCat

Lipnevich, A. A., Berg, D., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. L. Brown and L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 169–185) New York, NY: Routledge.
Google Scholar Google Preview WorldCat COPAC

Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science*, *42*(4), 539–559.
Google Scholar WorldCat

Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, *15*(4), 319–333. https://doi.org/10.1037/a0017841
Google Scholar WorldCat

Mittler, P. J. (1973). Purposes and principles of assessment. In P. J. Mittler (Ed.), *Assessment for learning in the mentally handicapped* (pp. 1–16). London, UK: Churchill Livingstone.
Google Scholar Google Preview WorldCat COPAC

Murray, J., Gasson, R., & Smith, J. K. (2018). Towards a taxonomy of written feedback messages. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 79–96). Cambridge, UK: Cambridge University Press. doi: 10.1017/9781316832134.006
Google Scholar Google Preview WorldCat COPAC

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, *22*(2), 155–175.
Google Scholar WorldCat

Nicol, D., & McCallum, S. (2021). Making internal feedback explicit: Exploiting the multiple comparisons that occur during peer review. *Assessment & Evaluation in Higher Education*, *47*(3), 424–433.
Google Scholar WorldCat

Nitko, A. J. (1980). Distinguishing the many varieties of criterion-referenced tests. *Review of Educational Research*, *50*(3), 461–485.
Google Scholar WorldCat

Nusche, D. (2008). Assessment of learning outcomes in higher education: A comparative review of selected practices, OECD Education Working Papers, No. 15, OECD publishing, © OECD. doi:10.1787/244257272573

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education*, *44*(8): 1253–1278. doi:10.1080/02602938.2019.1600186.
Google Scholar WorldCat

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, *9*, 129–144.
Google Scholar WorldCat

Popham, W. J., & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement*, *5*, 1–9.
Google Scholar WorldCat

Price, D., Smith, J. K., & Berg, D. A. G. (2017). Personalised feedback and annotated exemplars in the writing classroom: An experimental study in situ. *Assessment Matters*, *11*, 122–144.
Google Scholar WorldCat

Ramaprasad, A. (1983). On the definition of feedback. *Behavioural Science*, *28*(1), 4–13.
Google Scholar WorldCat

Sadler, R. D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144.
Google Scholar WorldCat

Scriven, M. (1967). *The methodology of evaluation*. Washington, DC: American Educational Research Association.
Google Scholar Google Preview WorldCat COPAC

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189.
Google Scholar WorldCat

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning*, *33*(1), 1–19.
Google Scholar WorldCat

Smith, J. K. (2003). Reconceptualizing reliability in classroom assessment. *Educational Measurement: Issues and Practice*, *22*(4), 82–88.
Google Scholar WorldCat

Stiggins, R. (2008). *An introduction to student-involved assessment FOR learning* (5th ed.). Upper Saddle River, NJ: Pearson-Merrill/Prentice Hall.
Google Scholar Google Preview WorldCat COPAC

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, *72*(7), 534–539.
Google Scholar WorldCat

Stiggins, R. J. (1985). Improving assessment where it means the most: In the classroom. *Educational Leadership*, *43*(2), 69–74.
Google Scholar WorldCat

Stiggins, R. J. (2018). The emotional dynamics of feedback from the student's point of view. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 519–530). Cambridge, UK: Cambridge University Press.
Google Scholar Google Preview WorldCat COPAC

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, *22*, 271–286.
Google Scholar WorldCat

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice*, *5*(2), 5–17.
Google Scholar WorldCat

Van der Kleij, F. M., & Lipnevich, A. A. (2020). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation and Accountability*, *33*(2), 345–373.
Google Scholar WorldCat

Wiliam, D. (2018). Feedback: At the heart of—but definitely not all of—formative assessment. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 3–28). Cambridge, UK: Cambridge University Press.
https://doi.org/10.1017/9781316832134.003
Google Scholar Google Preview WorldCat COPAC

Wiliam, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles Policy and Practice*, *11*(1), 49–65.
Google Scholar WorldCat

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, *52*(1), 1737. doi: 10.1080/00461520.2016.1207538
Google Scholar WorldCat

Wolf, L. F.,& Smith, J. K. (1995). The consequence of consequence: motivation, anxiety, and test performance, *Applied Measurement in Education*, *8*(3), 227–242. doi: 10.1207/s15324818ame0803_3
Google Scholar WorldCat

Wolf, L. F., Smith, J. K., &. Birnbaum, M. E. (1995). Consequence of performance, test, motivation, and mentally taxing items, *Applied Measurement in Education*, *8*(4), 341–351. doi: 10.1207/s15324818ame0804_4
Google Scholar WorldCat