POLICY PAPER

Recommendations from the National Panel
on the Future of Assessment Practices

The Future of Assessment Practices:

COMPREHENSIVE AND BALANCED ASSESSMENT SYSTEMS

LSI
LEARNING® SCIENCES INTERNATIONAL

Table of Contents

Introduction

About LSI

Learning Sciences International® (LSI) empowers schools and districts to transform core instruction and leadership practices, resulting in rapid gains in student learning.

At the center of this transformation is the company's Schools for Rigor partnerships, which are proven to raise student performance through strengthening core instruction and leadership practices and meet Every Student Succeeds Act (ESSA) requirements for evidence-based interventions.

LSI empowers each student and educator to meet the new challenges of a new economy (in which today's students and educators must prepare for a future in which new jobs, skills, functions, and disciplines are necessary) by transforming traditional core instruction and leadership practices with research-based, results-driven strategies, products, and services. By combining the most effective elements of traditional pedagogy, such as the strong social bonds forged by impassioned educators, with the advancements of new technology at a student's fingertips, LSI is at the forefront of this educational evolution and transformation for the better.

About The Panel

The United States spends $130,000 to educate each student from K through 12 – yet lags behind many other countries in academic achievement and is slipping further behind. Now is the time to fix our classrooms. Our students have waited long enough.

In 2018 the National Panel Charting the Future of Assessment Practices in the U.S. began as a movement where student success takes center stage. In that same year, at the 2nd annual Formative Assessment National Conference, leading educational experts on formative assessment—Susan Brookhart, Rick Stiggins, Jay McTighe, and Dylan Wiliam—participated in a fervent panel discussion. In the end, they all agreed a lack of a comprehensive and balanced assessment system is at the very heart of our challenges.

In that discussion Dr. Susan M. Brookhart exclaimed that we have seen an absence of implementation despite the many assessment systems which have been written and developed over the years.

While Dr. Dylan Wiliam lamented, "It is hard for me to imagine how it could be any worse." He went on to expound that teacher education needs to be treated as a process of habit change.

In 2019 at the 3rd annual Formative Assessment National Conference we tackle the elephant in the room - grading.

Susan M. Brookhart, Jay McTighe, Tom Guskey, and Dylan Wiliam will continue to discuss this important shift which can ripple into a far-reaching effect on how students ultimately think and behave.

In fact, Dr. Wiliam maintains, "Grading is essential in American schools. We have to have measures of how much the students have learned. The trouble is the way it's done in many schools, grading gets in the way of learning."

Join us in our effort to give each and every one of our students a shot at a better life. Let's start by raising awareness with this thought-provoking policy paper, "Comprehensive and Balanced Assessment Systems."

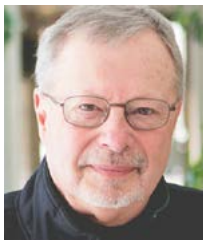
Author Bio

Susan Brookhart

Dr. Susan M. Brookhart is Professor Emerita at Duquesne University and an expert consultant with an extensive background working with schools, districts, universities, and states. She studies the role of both formative and summative classroom assessment in student motivation and achievement, the connection between classroom assessment and large-scale assessment, and grading. She is author or co-author of 18 books and more than 70 articles and book chapters and has served as editor for academic journals.

Jay McTighe

Jay McTighe brings a wealth of assessment experience from leading classroom formative performance assessment efforts with the Maryland Assessment Consortium, from his work on large-scale performance assessments with the Maryland State Department of Education, and from his many other projects at state and district levels. He is co-author of 15 books, including the award-winning and best-selling Understanding by Design® series with Grant Wiggins, and has written more than 40 articles and book chapters.



Rick Stiggins

Dr. Rick Stiggins founded the Assessment Training Institute (ATI) to help teachers, school leaders, policy makers, and communities develop assessment literacy. He guides practitioners to assess accurately and use the classroom assessment process to support, not merely monitor, student learning. ATI's approach to assessment has been used productively by hundreds of thousands of teachers and school leaders for the past three decades. He is the author of dozens of articles, books, and training programs.

Dylan Wiliam

Dr. Dylan Wiliam has helped to successfully implement classroom formative assessment in thousands of schools all over the world. A BBC documentary tracking his work showed how his formative assessment strategies empower students, significantly increase engagement, and shift classroom responsibility from teachers to students. He has written over 300 books, chapters, and articles; his latest book breaks down the gaps between what research tells us works and what we actually do in schools.

Executive Summary

Educational assessment is the process of eliciting, gathering, and interpreting evidence of student learning to describe student learning and/or inform educational decisions. School district assessment systems should serve to improve student learning and to document that learning for a variety of stakeholders. Comprehensive assessment systems assess all valued learning outcomes, not just those that are easy to test, and assess learning at all levels of the system: individual learners, classrooms, schools, and districts. Balanced assessment systems provide meaningful, relevant, and sufficient information for each stakeholder, with information quantity and quality commensurate with the uses to made from it: more detailed information for individual learners and their teachers in the classroom, where the learning takes place, and proportionally less (more general, and more aggregated) information available as the distance from the learning increases. Comprehensive and balanced assessment systems include a variety of types of assessments, producing evidence that can be used formatively, to improve learning, and evidence that can be used summatively, to certify, report on, or evaluate learning. Comprehensive and balanced assessment systems pay attention to the quality of assessment information; the process used to gather, interpret, and use assessment information; and the people who participate at all levels of the system, including students.

To be blunt, most district assessment systems are neither comprehensive nor balanced. This white paper describes the components of an ideal comprehensive, balanced assessment system that includes classroom formative assessment (within and between lessons), medium-cycle formative assessment (within and between instructional units), classroom summative assessment (grading), long-cycle formative assessment (several times during the school year), and district and state-level accountability assessment. It suggests ways these components should work together to provide the information needed at all levels to support teaching and learning and support a view of student learning consistent with current theories of student learning and motivation. The paper ends with some suggestions for districts interested in moving forward toward this vision, and advocates for doing so.

To be blunt, most district assessment systems are neither comprehensive nor balanced.

Comprehensive and Balanced Assessment Systems

Educational assessment is the process of gathering evidence of student learning to inform educational decisions. Assessment systems should serve both to improve student learning and to document that learning for a variety of stakeholders. An assessment system is composed not only of assessment tools and processes, but also the people who use them. Many school districts use collections of assessment tools and processes that either do not serve to improve student learning, miss important learning outcomes, or under-serve one or more stakeholder groups. The purpose of this white paper is to describe ideal comprehensive and balanced assessment systems for school districts. We will address the system concept as a school district matter because this is the context in which the educational decisions are made that impact student learning. Districts may use this description to evaluate their own assessment system and set goals for improvement. The paper is organized into three sections: an overall vision for comprehensive and balanced assessment systems, the components of a comprehensive and balanced assessment system, and recommendations for enacting such a system.

A Vision for Comprehensive and Balanced Assessment Systems

If an assessment system is to help improve student learning and document that learning for a variety of stakeholders, it must be both comprehensive and balanced. Comprehensive assessment systems assess all valued learning outcomes, not just those that are easy to test, and assess learning at all levels of the system, with results and analyses describing learning for individual learners, classrooms, schools, and districts. Comprehensive and balanced assessment systems include a variety of types of assessments to serve a variety of purposes and uses, producing some evidence that can be used formatively, to improve learning, and some evidence that can be used summatively, to certify or report learning. Balanced assessment systems strike a balance in the assessment such that the available information is appropriate and useful for the information needs at the various levels of the system. Following this logic, a balanced assessment system does not provide an equal amount of assessment information available to each level of the system, but rather offers more detailed information to individual learners and their teachers in the classroom, where the learning takes place, and proportionally less (more general, and more aggregated) information available as the distance from the learning increases.

Learning outcomes are the foundation of a comprehensive, balanced assessment system and the reference against which assessment information should be interpreted. An important feature of a comprehensive and balanced assessment system is coherence among the learning outcomes, attendant assessment and instruction, and the views of learning they imply, at all levels of the system (Wilson, 2004). State standards are broad statements of

learning goals measured by district and state level assessments. Curricular and unit goals are smaller in scope, and typically a state standard will encompass more than one curricular or unit goal. Measurement of learning goals at this level is typically accomplished by both medium-cycle formative assessment and classroom summative assessment. Each unit learning goal typically encompasses several daily learning targets for individual lessons, and classroom formative assessment garners information keyed to lesson-sized learning targets. A critical aspect of a comprehensive assessment system is that these learning outcomes are coordinated; they work together to guide students' learning and teachers' instruction; they describe all the valued learning outcomes necessary for students to ultimately reach the standards; and they are framed by compatible understandings of learning, instruction, and assessment.

A balanced assessment system prompts educators to collect data in grain sizes that are appropriately actionable at each level of the system. Balanced assessment systems generate a great deal of classroom formative assessment information, varying in length from a few seconds to a week, because the resulting actions are more immediate and smaller in scope—typically actions taken by learners and their teachers during lessons. These small outcomes are often not recorded—although they can be—but rather are the basis for student and teacher action. As the assessment information increases in aggregation and distance from the classroom, or is collected periodically, the resulting actions are more distant and larger in scope—typically resource allocation or policy decisions made by administrators for district planning. Such information should be less frequent and less detailed. A comprehensive and balanced assessment system should attend to both the assessment tools (tests, skill checks, performance assessments, classroom questions) and processes (the methods by which students and teachers participate in assessment activities, and the classroom climate in which they do so) that are currently presented in other descriptions of assessment systems, and also to the assessment literacy and information needs of the actors at each level of the system (Michigan Assessment Consortium, 2017; Stiggins, 2017).

The process of evaluating and improving local systems should be guided by a set of key questions:

- Are the learning goals to be assessed clear to all stakeholders, including students?
- Is the purpose of each assessment clear: What is the decision to be informed and who will make it (them)?
- Are the assessment tools capable of providing the needed information?
- Do the assessment processes deliver the needed information into the hands of the intended users in a timely and understandable form?
- Do assessment users at all levels of the system have the skills they need to gather, interpret, and use assessment information?

This last question focuses on the assessment literacy of the teachers and school leaders who manage assessment at all levels; that is, the level of their mastery of the basic principles of sound

assessment practice. Without this foundational professional competence in place, development of a quality local assessment system is highly unlikely.

One of the current problems with assessment systems in many districts is that this balance is backward, with more resources spent on the less frequent and summative components of the system.

Figure 1 on the next page, identifies the components of a comprehensive and balanced assessment system. The locus of assessment administration and use moves from closest to the learning on the left to closer to administrative and policy decision-making on the right. The frequency of assessment is greater and grain size of information is smaller on the left and increases toward the right. Arguably, then, the amount of time and other assessment resources invested should be largest on the left and decrease toward the right. One of the current problems with assessment systems in many districts is that this balance is backward, with more resources spent on the less frequent and summative components of the system. The result is more information to inform the periodic instructional decisions made by administrators and less information to inform those made continuously day to day in the classroom by learners and their teachers.

Figure 1. Components of a Comprehensive Assessment System

Comprehensive Assessment System Components				
Short-Cycle Classroom Formative Assessment	**Medium-Cycle Formative Assessment**	**Classroom Summative Assessment (Grading)**	**Long-Cycle Formative Assessments**	**District-Level Summative Assessments and Annual State Accountability Assessments**
Evidence of learning of lesson-sized learning target(s), generated and used by both students and teachers during the course of learning	Evidence of learning across related lessons or a unit (e.g., weekly diagnostics), for short-term instructional and learning adjustment	Evidence of student achievement at a point in time, for reporting (e.g., unit tests, performance assessments)	Evidence of student learning, typically 2 to 3 times a year, for longer-term instructional planning	Evidence of student achievement of curricular learning outcomes and/or state standards, for reporting (e.g., end-of-course exams, state accountability assessments)
High Utility to Teachers and Parents			**High Utility to Central Office Administrators**	
				High Utility to Policy Makers
Have students learned the lesson content? What do they think the learning target is, where are they now, and what should they do next?	Have students retained their learning (learned curriculum)?		Is the retained learning (learned curriculum) aligned with the accountability system?	Does the retained learning (learned curriculum) meet district and state expectations?
Appropriate to answer questions such as:				
• How are students thinking about lesson-sized chunks of content (daily learning target concepts/skills)? • What next steps do the students need to take in their understanding? • Was the planning of my lesson effective? • Did the students learn the lesson learning targets? Which students struggled (and why)? Which students need enrichment (and why)? • How will I adjust my planning of tomorrow's lesson for those students	• How are students thinking about unit-sized chunks of content (unit goal concepts/skills)? • What next steps do the students need to take in their understanding? • Did the students retain what they learned in previous lessons? • Which students are still struggling with the content, and which students need enrichment? • How will I adjust my planning in the next few lessons in this unit?	• What are students' current status/ achievement levels on the learning goal(s) assessed? • How should we report students' current achievement to parents/guardians and to the reporting/ record-keeping system?	• Are the standards being taught and learned? • Does our curriculum have gaps between learning expectations and assessment? • What structural or instructional changes might be helpful?	• Does the curriculum cover the standards in appropriate breadth and depth? • How does each tested grade level, subject, and school perform in regard to the standards? • Which curricular area(s) may need more resources?
NOT appropriate to answer questions such as:				
• Which students "got it"/"didn't get it"?	• Which students "got it"/"didn't get it"?	• Which students are the best/smartest? • Which teacher is more effective?	• Which teacher is more effective? • Which school is more effective?	• Why did students perform the way they did? • Why did schools perform the way they did?

The Components of Comprehensive and Balanced Assessment Systems

In this section, we discuss the following assessment components in turn: daily classroom formative assessment (sometimes called short cycle formative assessment), formative assessment within and between instructional units (sometimes called medium-cycle formative assessment) and interim/benchmark assessment (sometimes called long-cycle formative assessment), assessment for classroom grading, and district- and state-level assessments. Each component is defined and its purposes are specified. Then a brief discussion explains how the component should function in the system, what research says about the component, and what questions its information can and, perhaps more importantly, cannot answer. Next, we describe the responsibilities of the various parties involved. In most cases, people from several role groups share joint responsibility in order to coordinate assessment practices and information throughout the system. Finally, for each component the current state of practice is compared with how the component should function in an ideal comprehensive and balanced assessment system.

Short-cycle Classroom Formative Assessment

Short-cycle formative assessment occurs in the classroom, is on-going, and serves only to support student learning. It takes place during—and as part of—instruction, which typically means during a lesson or practice. It helps student/teacher teams make incremental decisions focused specifically on what they are trying to teach and learn, where they are in the process, and what they need to understand or do next to improve. Formative assessment helps teachers make incremental decisions about what they are trying to teach, how students currently are thinking about the concepts, and what immediate next instructional adjustments would help move students along. Wiliam (2010, p. 31) lists five key strategies that comprise short-cycle formative assessment:

1. Clarifying, sharing, and understanding learning intentions and criteria for success
2. Engineering effective classroom discussions, questions, and tasks that elicit evidence of learning
3. Providing feedback to teachers and students to inform instruction and improve learning
4. Activating students as instructional resources for one another
5. Activating students as the owners of their own learning

When formative assessment is intended, designed, and used to support students as they make the decisions that promote their learning, it helps them understand their learning target, participate in the collection of evidence of their own level of attainment, and collaborate with their teacher in deciding what comes next in their learning.

Research. There is evidence that formative assessment, when done well, improves student learning (Black & Wiliam, 1998; Graham, Hebert, & Harris, 2015). In a well-functioning system, short-cycle formative assessment includes both informal methods, like classroom questioning and observation, and more formal methods, like homework and practice work that, while not graded, helps inform students and teachers of learning progress during instruction while there is still time to address learning before reporting time (Ruiz-Primo & Brookhart, 2018). Importantly for the concept of an assessment system, classroom formative assessment is the component that most involves the students and is most directly connected to their learning process as it is happening. When formative assessment is absent, weak, or poorly implemented in an assessment system, the system's major link to the focal stakeholders—the learners—is weakened or broken. This disenfranchises learners from a system that should be designed to benefit them and, essentially, washes out the foundation of the system itself.

When formative assessment is absent, weak, or poorly implemented in an assessment system, the system's major link to the focal stakeholders—the learners—is weakened or broken. This disenfranchises learners from a system that should be designed to benefit them and, essentially, washes out the foundation of the system itself.

Questions addressed. Information from short-cycle formative assessment helps students and teachers know how students are thinking about lesson-sized chunks of content from their daily learning targets and what next steps they need to take, for students to enhance their understanding and/or for teachers to adjust their instruction. Done well, it focuses on uncovering student thinking as opposed to evaluating or scoring student performance. A common but shallow understanding of formative assessment is that it helps teachers know which students "got it" or "didn't get it." This view of formative assessment is not only impoverished; it can lead to evaluative judgments of students by teachers and students themselves about their own learning. Such thinking robs students of the confidence they need to continue striving for success and works against student learning, especially for students who struggle (Stiggins, 2017).

In contrast, interpreting information from well-designed formative assessment as evidence of student thinking and current place in learning progressions helps learners and teachers figure out next steps. So, for example, the more useful formative assessment information from an incorrect answer to a two-step mathematics problem is not that the student got the problem wrong, but what thinking was in evidence (e.g., was confused about when to divide and when to multiply). This kind of information is immediately actionable, both to focus the student's attention and intentions and to inform the teacher's immediate next instructional decisions. It is detailed at a fine grain size (e.g., not "mathematics" or even "numbers and operations," but "distinguishing multiplication and division").

Thus, classroom formative assessment information is the foundation from which a comprehensive, balanced assessment system is launched; it is foundational in the sense that if the overarching purpose of the assessment system is to support learning, that support begins and is based in this level of the system. It involves and informs the most vulnerable and the most important stakeholders, students. It supports a view of learning that understands students as the agents who regulate their own learning (Zimmerman & Schunk, 2011). Although students are the primary stakeholders – school districts exist primarily for the purpose of educating students – they are often overlooked in assessment systems, which are typically designed to meet the needs and desires of the adult stakeholders. Formative assessment also empowers teachers, who should be key players in assessment systems but, in current practice, often feel like assessment is something done to them rather than for them. Comprehensive, balanced assessment systems include a solid foundation of high-quality formative assessment, in every lesson, by every student and teacher.

Responsibility and system coordination. Responsibility for this component of the system rests, in different ways, with students, teachers, and school leaders. While it may seem odd to give students responsibility for a part of the assessment system, research has shown that when students take responsibility for their own learning and assessment, assessment does support learning—the purpose of the assessment system—and when they don't, learning is less well supported, for students across the achievement range (Zimmerman & Schunk, 2011). Similarly, teachers improve in their formative assessment effectiveness when they begin to look at learning and assessment through students' eyes and approach their assessment practices from that perspective, which is a sea change for most educators (Brookhart, 2017). Finally, school leadership (building and district) and support is critical for formative assessment to function effectively and systematically within a school (Noyce & Hickey, 2011; Schneider & Randall, 2010). Building principals should take overall responsibility for instructional quality in their building.

Similarly, teachers improve in their formative assessment effectiveness when they begin to look at learning and assessment through students' eyes and approach their assessment practices from that perspective, which is a sea change for most educators (Brookhart, 2017).

Current status vs. ideal functioning. Despite its position as the foundational component in a system whose major purpose is to support student learning, classroom formative assessment typically is the weakest component in most districts' assessment systems. This is due in part to the lack of assessment literacy training both for teachers and their supervisors in their pre-service preparation—training that should develop assessment knowledge and skills as well as the realization that assessment is part of their professional responsibility and the disposition to do it well. Accordingly, professional development in this arena is clearly needed and strongly recommended.

Another issue needing attention is the status of the student, the least powerful stakeholder in systems run by adults. Presently, the students function as examinees who respond to assessments rather than as proactive learners who are actively involved in the assessment process (Stiggins, 2014a). Modern learning theory holds that students actively construct learning (Shepard, 2001; Zimmerman & Schunk, 2011), and one important aspect of coherence is that assessment of learning be underpinned by similar theories of learning (Pellegrino, Chudowsky, & Glaser, 2001; Wilson, 2004). The mismatch between treating students as active constructors of knowledge for short-cycle classroom formative assessment and as passive examinees for district tests creates a lack of coherence in the system. Many teachers and even more administrators have yet to realize the sea change described above, looking at learning from the students' point of view. On the contrary, many educators and others still hold associationist theories of teaching and learning and a traditional view of assessment merely as something adults do to students, in which students are respondents (examinees) rather than active participants in the learning process (Brookhart, 2017; Shepard, 2001).

Research suggests that this change can be difficult, re-orienting classrooms and building cultures from primarily adult-centered to primarily student-centered, and is more a matter of habit change than knowledge acquisition.

To move toward a comprehensive and balanced assessment system, a district should begin with intensive development of knowledge, skills, and practice in formative assessment, for all teachers and administrators (Black & Wiliam, 2004). Research suggests that this change can be difficult, re-orienting classroom and building cultures from primarily adult-centered to primarily student-centered, and is more a matter of habit change than knowledge acquisition. The authors are very aware that calls for the improvement of formative assessment are common, and often not successful. District policy makers who do not know which part of an accountability system most supports learning, and how that happens, mistakenly prioritize large-scale testing over classroom formative assessment. Often, good-faith efforts to improve formative assessment in classrooms, schools, and districts are misdirected or misunderstood (e.g., formative assessment presented as a list of "techniques" such as an Exit Ticket), underfunded, or under-prioritized (e.g., despite formative assessment initiatives, more attention still rests on large-scale accountability tests and teacher evaluation). Only when radical shifts in beliefs about learning and teaching and in classroom and school culture are made will comprehensive, balanced assessment systems be possible.

Medium-cycle Formative Assessment

Typically accomplished with more formal formative assessment (Ruiz-Primo & Brookhart, 2018), medium-cycle formative assessment occurs within and between instructional units,

typically in intervals of from one to four weeks (Wiliam, 2010) to inform students' decisions about studying and teachers' decisions about adjusting larger, longer-term lesson plans. For example, in Philadelphia, the year is divided into six-week blocks, with essential standards being taught in the first five weeks, on which students are tested, with the test performance used by teachers do determine whether week six is spent on extension or review (Goertz, Oláh, Nabors, & Riggan, 2009).

Another example is the common assessments used by teams of teachers in the context of professional learning communities (DuFour, 2004). In this case, teams devise assessments reflective of the intended outcomes units of instruction offered by all team members across classrooms. Results are analyzed by the team to discern which team members achieved the best results so as to instruct others about how to improve their instruction.

Medium-cycle formative assessment typically involves assessment of student work on quizzes or performance tasks that encompass one or more instructional objectives, as opposed to the smaller grain-sized daily learning targets referenced in short-cycle formative assessment. Thus, the main actors in this component of the system are also students and teachers, but the purpose is somewhat broader. Medium-cycle formative assessment shows how students are synthesizing the bite-size chunks of content from their lessons into more general understandings often summarized as unit goals derived from state standards.

Research. Research on medium-cycle, formal formative assessment has been mixed, largely because of problems in implementation (Furtak et al., 2008). However, there have been some exceptions. Saunders, Goldenberg, and Gallimore (2009) reported on a five-year study of work with grade-level teams in Title I schools. The first two years of work with principals only produced no changes in achievement, but the second phase, which included training for both principals and teacher leaders, increased both achievement and growth.

Questions addressed. Medium-cycle formative assessment answers questions about how students are thinking about unit-sized chunks of content, how they are able to apply what they are learning to build up larger understandings, and where they should go next. The focus of such periodic formative assessment should be on identifying what students are thinking, where they are in a learning progression, and what student or teacher instructional moves might be most likely to increase progress.

While short-cycle formative assessment informs adjustments the teacher or students make during live instruction, medium-cycle formative assessment provides more formal evidence on which teachers can base more general instructional planning, for example lesson planning, adjusting lesson pacing, grouping or regrouping students for remediation or enrichment, tutoring, providing additional practice, and so on. In the context of ongoing classroom formative assessment, the actionable information comes from insights about individual student thinking and performance that assessment results permit. But in the periodic assessment context, actions are suggested by

patterns of student performance detected over time and across classrooms and/or instructional approaches.

Responsibility and system coordination. In larger school districts, the responsibility for medium-cycle formative assessment may lie with district curriculum leaders. Teachers, working alone or in teams, and building principals should share in this work. Teachers and building principals are responsible for implementing the curriculum for students, that is, for mediating the written curriculum into the taught curriculum. As for classroom formative assessment, principals have supervisory responsibilities toward the teachers and coordinating responsibilities toward the rest of the system, as well.

For all types of formative assessment, those who devise, conduct and use it must be assessment literate.

Current status vs. ideal functioning. For all types of formative assessment, those who devise, conduct and use it must be assessment literate. They must understand and be able to apply basic principles of sound assessment. Specifically, this means they must be masters of the learning goals to be assessed, able to select a proper method for the goal(s), able to build quality assessments and scoring schemes and able to anticipate and minimize any sources of bias that can distort results. These requirements apply regardless of the formative assessment context. We already have established that many teachers and building principals would benefit from skill development in these two areas, including involving students in the formative learning cycle and reasoning from evidence of learning.

Programs that have embedded periodic formative assessment in curriculum materials without attention to these principles have not had much success (Yin et al., 2008). Once these principles are in place and teachers and administrators begin to develop skills in using them, medium-cycle formative assessment tools such as quizzes and short performance tasks can be incorporated into the process.

Classroom Summative Assessment (Grading)

Classroom tests and performance assessments are the most common tools used to assess (evaluate) student achievement at a point in time, typically at the end of a series of related lessons and at the end of a unit. These are scored in different ways, most commonly as percent correct or by matching performance to levels on a rubric, sometimes translating the result into grading symbols (e.g., ABCDF) for communication. These individual components are aggregated for reporting at regular intervals, for example, for report cards issued at the end of a 9-week quarter or other intervals specified by district policy. The purpose of grades is to judge the sufficiency of student learning given pre-set achievement expectations. We seek to inform students and parents of a student's current status on either a subject or standard,

depending on the type of reporting used, in effect creating "punctuation" points in a student's learning trajectory to take stock of learning in a formal way. A secondary purpose is to inform administrators and future teachers of a student's performance, for potential use in administrative or placement decisions. For older students, grades are entered into their permanent records. These are summative functions, although it is possible to use summative assessment results for formative purposes, as well, as for example when a teacher reviews test results to prompt further studying and assessment (Black et al., 2003). [Note that some states "grade" schools as part of the state's accountability system. This is not a district function. In this paper, we use the term "grades" to mean the grades students receive on classroom assessments or report cards, not ratings of schools by states.]

Research. Research on grading has identified several problematic issues (Brookhart, Guskey et al., 2016). Certain teacher grading practices, for example, counting surface features of an assignment that are unrelated to the standard it is designed to assess, or counting class participation in a grade intended to assess content learning, threaten the quality of information about learning that grades provide. Variability in grading practices and inconsistent application of criteria also threaten the reliability of grades. Nevertheless, grades can predict important educational outcomes like dropping out of school and being admitted to and successful in college. They also serve an administrative function in schools by summarizing student learning with a simple indicator that has utility especially in large schools and districts.

Questions addressed. Done well, grades should answer questions about students' current achievement status on important learning goals, to inform students, parents and guardians, and the school and district. For standards-based or standards-referenced grading, those important learning goals are expressed as reporting standards. Grades should not be used to compare students with one another (norm-referencing). The actionable information grades provide for students is less about learning specific concepts and skills—every 9 weeks is a bit late for that—and more about broader questions of whether students' learning needs are being met. They can serve as a way in to discussing learning and school more generally with students and parents. For standards-referenced grading, grades are intended to represent students' current status on learning standards and should not include attendance, motivation, or effort. However, these non-cognitive qualities can be brought in as part of the conversation as students, parents, and teachers interpret and discuss students' grades. Because grades are sometimes difficult to interpret, this component often represents a weak spot in district assessment systems. Grades stand at the transition point in a comprehensive assessment system, between assessment of learning for direct student and teacher consumption and use and assessment of learning for evaluative and administrative purposes.

Responsibility and system coordination. The state legislature empowers the local board of education to establish local policies for their operations, including grading (McElligott & Brookhart, 2009). Therefore, the local school board and district administrators bear responsibility for grading and can be sued

in court for perceived abdications of this responsibility. Suits mostly focus on due process or equal protection concerns under the 14th Amendment of the U.S. Constitution (McElligott & Brookhart, 2009). However, in practice, shared responsibility for grading rests with the teachers who assign the grades, building principals who oversee and, in many districts, have the authority to change grades if deemed appropriate, and district administrators.

These responsibilities must be coordinated. Classroom teachers' grading practices and classroom-level policies should be as consistent as possible with other teachers' practices and policies. At the classroom level, the policies are usually about details of what counts as evidence for various grades and how evidence may be collected (e.g., due dates and late policies). That means teachers are responsible for the match between their classroom assessments (e.g., tests and performance assessments), intended learning outcomes, and the approach to learning supported by the system. Teachers are also responsible for weighting and aggregating classroom assessment information into a report card grade that communicates about students' current status on those learning outcomes. At the building level, principals are responsible for seeing that teachers carry out meaningful grading practices, and also for reviewing due process and equal protection concerns. The district is responsible for seeing that students receive due process and equal protection in grading issues, and that grades are accurately recorded into the district database.

Current status vs. ideal functioning. Similar to formative assessment, grading is at present a weak spot in most districts' assessment systems. To begin with, the dependability of any report card grade depends of the quality of the evidence on which it is based. It is impossible to combine low-quality test scores and get a meaningful representation of a student's level of achievement. We have already mentioned our concerns about the lack of assessment literacy in the classroom. This concern generalizes from classroom formative to medium-cycle formative to classroom summative assessment (report card grading). Professional development may be needed, depending on local circumstances.

Second, in many cases, grading relies on a banking model. Once students have demonstrated their proficiency on a specific standard (once it's "in the bank"), graded work pays no attention to whether what was assessed is retained. However, students often do forget. In some cases, forgetting occurs because learning was not deep enough to begin with, for example, topics were touched on but not completely understood, or skills were not practiced to fluency.

… many current grading policies hurt students rather than support learning.

In addition, many current grading policies hurt students rather than support learning. For example, some classroom grading schemes result in students realizing halfway through a unit that they have no chance of passing, causing them to give up and sometimes see themselves as

stupid or worthless. Change may be required so that grades report current levels of student achievement of intended learning outcomes after students have had sufficient formative (learning and practice) opportunities and that the classroom assessment climate supports and motivates students to participate to the best of their ability in the formative learning cycle. Grades should convey to students where they are on learning outcomes they understand and what they are on track to do next. These changes require better description of student work across a continuum for each learning outcome, matched closely to standards and supportive of an active view of student learning.

Changes in grading policies and practices like these may run into some resistance. Some parents and others in our communities see grades as positional goods, whereby higher grades for some students convey status that relies on lower grades for other students. Such attitudes will need to change, although the assessment system we are proposing is possible even if we cannot stop some parents from regarding grades as positional goods. In addition, some new policies and practices will need to be worked out, to deal more appropriately with diversity in student abilities in a learning-referenced grading system, such that helpful and accurate reporting of learning can happen without hurting students. Such policies will be critical to ensuring that standards-based grading does not exacerbate the problems inherent in current and traditional grading systems.

Long-cycle Formative Assessments

Many districts use interim or benchmark assessments, both of which are typically purchased from commercial vendors, although some larger districts develop their own. Interim assessments usually are parallel test forms for an external accountability test; they cover an entire year's worth of content and are administered two or three times during the school year to track student learning and achievement growth. Benchmark assessments usually are non-parallel test forms covering a portion of the year's content (e.g., the first report period) and are intended to be administered at a specified point in the school year and curriculum (Ferrara, Maxey-Moore, & Brookhart, in press). However, some educators use the terms interchangeably. Both interim and benchmark assessments are intended to identify students who need more support to succeed and to inform curriculum planning and resource allocation. At present, some teachers see interim and benchmark tests as simply "test prep" practice for the state accountability tests; this is not the use for which these tests were designed.

Instructional and grouping decisions based on long-cycle assessments are not the fluid, in-class adjustments and groupings based on short- and medium-cycle formative classroom assessment, but rather grouping for pull-out interventions and other more structural purposes. At this point in the system, students become secondary stakeholders, involved only to the extent that decisions by teachers and administrators ultimately affect their experiences.

The primary stakeholders for interim and benchmark tests are administrators and teachers. Interim and benchmark tests primarily inform educators, not students, and the decisions made on the basis of their results often affect students other than those who took the assessment (for example, resulting in better curriculum alignment for next year's students). In fact, when benchmark assessments are used to monitor students' progress toward state accountability test performance, they are functioning summatively.

Research. To date there is very little research evidence that using interim/benchmark assessments helps improve student achievement. One study showed no effects of using interim/benchmark data on student achievement in grades K to 2 and very small effects in grades 3 to 8 (Konstantopoulos et al., 2011). There is some evidence that when data teams in schools use interim/benchmark assessment data, they focus more on internal teaching issues than external forces not under their control (Gallimore et al., 2009), although it is worth reporting that this study reported a significant impact on student achievement. However, a study of teachers' use of mathematics interim/benchmark assessments found teachers mostly used results to group students or reteach procedural knowledge, rather than making sense of students' conceptual understanding (Oláh, Lawrence, & Riggan, 2010). Reviewing these and other studies, Abrams and McMillan (2013) concluded that interim assessment data influenced topic selection as teachers decided to teach or reteach, but not cognitive considerations about how to reteach. Thus the value of devoting resources to interim and benchmark assessments, as they are currently used, can be questioned.

... the value of devoting resources to interim and benchmark assessments, as they are currently used, can be questioned.

Questions addressed. Interim/benchmark data can answer general questions about student achievement in different areas in the curriculum, and sometimes the standards, depending on the test. However, large-scale assessments like this are much better for raising questions than answering them. Rather than collecting diagnostic information on every student, these monitoring assessments are best used to figure out which students need help; then, a separate assessment is needed to figure out what help to get them. For example an interim assessment might raise the question, "Why are my students not performing in mathematics at the level I expected?" Deep answers to these questions require looking at classroom-level assessment information. For example, a look at students' classroom work over time might find that they are better at computation than problem-solving using fractions and would also identify what kinds of mathematics work they had been asked to do (and perhaps, what they had not been asked to do but should have been). Effective action plans can be made based on these answers, and they cannot be made based on state test results alone.

Responsibility and system coordination. Interim and benchmark assessments are a relatively new addition to the components of a comprehensive and balanced assessment system. They arose in response to a perceived need for more instructional, predictive, and evaluative information, at more frequent intervals, than the once-a-year state accountability tests that preceded them (Perie, Marion, & Gong, 2009). To date, responsibility for purchasing and administering interim and benchmark tests has rested with district administrators, and responsibility for interpreting results has been delegated to building principals and school data teams (Gallimore et al., 2009), with the not altogether satisfactory results reported above.

Current status vs. ideal functioning. As currently practiced, interim and benchmark assessment is the component of an assessment system with the least research support. It may be that, with enhanced short- and medium-cycle formative assessment and improved grading practices, this component can be eliminated or at least have its use radically transformed. When schools primarily use long-cycle interim or benchmark assessments to determine interventions instead of using quicker-acting systems (e.g., classroom formative assessment), they squander the power of formative assessment to prevent learning gaps in the first place. One of the goals of a balanced system weighted heavily on the side of classroom short-cycle and medium-cycle formative assessment is to strengthen core instruction and eliminate over-reliance on interventions.

If interim/benchmark assessments were to be reformed and not eliminated, this component of the assessment system should be conceived and designed in connection with classroom formative assessment (privileging the curriculum as it is taught), and not large-scale accountability assessment as is the case currently, where it is common for interim/benchmark tests to be built from the same item banks that are used in state accountability tests. Ideally interim/benchmark assessments, if used at all, should be less about mimicking state tests and more about reflecting standards and learning goals within standards more closely than they do now. Tracking systems for reconceived interim/benchmark assessments should track learning in concert with classroom formative assessment and should include students as partners. As one of the authors observed, "The state test is a snapshot, and what we need is a photo album."

Ideally interim/benchmark assessments, if used at all, should be less about mimicking state tests and more about reflecting standards and learning goals within standards more closely than they do now.

Finally, if interim/benchmark tests are reinvented, the quality of the teacher learning communities or data teams that deal with the data must be improved. This will require teacher efforts and principal and teacher leadership.

District-level Summative Assessments and Annual State Summative Assessments.

District-level summative assessments are typically end-of-course exams for various subject areas in the curriculum, sometimes for final course assessment and sometimes for high school graduation. They should be keyed to the district course curriculum expectations.

Annual state summative assessments have been much in the news since the reauthorization of the Elementary and Secondary Education Act as the No Child Left Behind Act in 2002 and the current Every Student Succeeds Act in 2015. Annual state assessments are typically keyed to state standards, but at a very large-grain-size level, so that the results speak to aggregated standards (for example, Reading, Mathematics, Writing) rather than to different individual standards within subject areas.

Research. Because the information is so broad in scope, state summative assessment results are best suited for informing policy decisions, not instructional decisions. However, policies affect schools (Au, 2007) and indirectly affect instructional decisions by creating various pressures on teachers and other aspects of the school system. Supovitz (2009) reviewed research on the use of high-stakes, test-based accountability in the United States and concluded that testing does motivate teachers to change, but the changes are mostly (p. 211) "superficial adjustments in content coverage and test preparation activities rather than promoting deeper improvements in instructional practice." Current teacher evaluation practices that use value-added models based on state summative assessment pressure teachers to change, but the effectiveness of these practices remains, on balance, unproven (Darling-Hammond, 2015). Value-added estimates for individual teachers are not very precise (Jacob & Lefgren, 2005), vary from year to year (McCaffrey et al., 2009), and depend heavily on statistical assumptions made in the different models (Goldhaber, Goldschmidt, & Tseng, 2013). For these reasons, the use of value-added modeling for making decisions about individual teachers' effectiveness is not recommended (American Statistical Association, 2014; Baker et al., 2010; Wiliam, 2016).

Questions addressed. End-of-course exams can answer questions about whether students are learning and retaining information they were supposed to learn in the course. This information can be aggregated to answer similar questions at the course, school, and district levels. End-of-course exams typically are not designed to be diagnostic or answer questions about why students performed the way they did.

State level accountability tests can answer questions about general performance in different subject areas. They can, if the tests are well-constructed, be used to describe the performance of different districts in teaching state standards. They cannot answer questions about the reasons for different performance from district to district.

For a variety of practical and technical reasons it is unacceptable to evaluate teacher performance based on change in annual standardized test scores analyzed using value-added models. For example, when tests sample broad domains of

achievement limitations in testing time require that many important learning outcomes go untested or are covered in a very superficial manner. Therefore, a fundamental mismatch could arise between what is tested and some teachers' assigned instructional responsibilities, rendering the test incapable of detecting the mismatched teacher's impact. Over and above the problems with the tests, there is the problem of the year-long time span between pre and post testing during which a wide variety of school and personal factors beyond the control of teachers have been shown to exert profound impacts on student learning success. Finally, there are the problems of the unstable estimates of teacher effects that have been revealed when using value-added analyses of scores. There is a role for the consideration of student growth in teacher evaluation but not using these scores or this kind of analysis. (Stiggins, 2014b).

Responsibility and system coordination. Responsibility for district-level summative assessments rests with district administrators, including curriculum coordinators, and is shared by building principals and teachers, especially the respective subject-area departments in which the assessments are used. This responsibility includes both quality control issues for the assessment tools (tests or performance assessments) as well as policy issues (e.g., whether and to what degree a student's results will count in a final grade).

The state, of course, is ultimately responsible for the quality, utility, and effectiveness of its state accountability testing program. District administrators are responsible for administration and reporting in accordance with the state's requirements. Because administering the state accountability test reaches down into school and classroom schedules, both building administrators and teachers share responsibility for implementation (e.g., following prescribed administration guidelines when giving the test).

Current status vs. ideal functioning. Three issues must be addressed to move current state accountability tests to more ideal functioning.

First, state accountability tests need to move more in the direction of testing applications of knowledge and problem-solving and away from testing discrete facts, as called for by many next-generation learning standards. There is some evidence that this is happening slowly, but it has not gone far enough fast enough.

Students must feel like the state accountability assessments are helpful, or in some way support their learning, in order to be motivated to do their best.

Second, there is the issue of student motivation. We learn little about students' achievement or understanding when they are not performing at their best, which can happen if students do not believe the assessments are important. Students must feel like the state accountability assessments are helpful, or in some way support their learning, in order to be motivated to do their best. At present this is not always the case. Most districts approach state accountability tests as something students must "do," and not only do once but prepare for weeks, in order

to make their school proud. Some school walls sport posters to that effect. Before student motivation about accountability tests really changes, the relevance of state test results for their own learning and for their school must be demonstrated to them. Current state accountability "school report cards" and other uses are not likely to advance this agenda, nor do they fit with a student-centered view of learning.

Third, assessment design for accountability needs to move from testing discrete knowledge of a large amount of content to testing for the application and transfer described in most contemporary learning standards. Then assessment reporting for accountability needs to be redesigned to encourage and support interpretation and use of assessment results for instructional and policy applications beyond emphasizing low-scoring subjects, to include more information about thinking, problem solving, and transfer. In fact, this is a consequence of the more general point that the assessment system should serve the curriculum, which in turn should be based on contemporary standards that include using knowledge, not just accumulating it.

Further Thoughts on Getting There

Four major conclusions follow from comparing typical district accountability systems with the ideal comprehensive and balanced assessment system described here.

1. Almost every district in the country needs to increase time, money, and professional development resources to raise both the quantity and quality of formative assessment in classrooms and to make appropriate use of this vital information. This may involve reducing the amount spent on other aspects of assessment: grading a smaller percentage of classroom assessments and increasing ungraded formative work with feedback, and transferring some of the resources now spent on large-scale assessment to classroom assessment.
2. Almost every district in the country needs to increase time, money, and professional development resources to improve teachers' grading practices and district grading policies that enable those practices. As above, this means a shift in the use of assessment resources.
3. Almost every district in the country needs to reduce the amount of time and energy spent on interim/benchmark tests and/or increase the amount of actionable information drawn from them.
4. At all levels of the system, from the classroom to the state, assessment tools and practices need to be broadened to include more assessments that call for students to apply what they know in more realistic (authentic) contexts (McTighe, 2018). At the classroom level, this calls for a change in classroom questioning and student discourse, an increase in the use (and quality) of performance assessment, and improvement in the interpretation and use of the results. At the large-scale level, this calls for assessment design changes so that evidence of student learning matches standards at a deeper level than at present.

Rebalancing districts' comprehensive assessment systems, with more focus and weight on short- and medium-cycle formative assessment, and with appropriate systems and professional development including on how to use the evidence with and for students, is a moral imperative. When teachers and administrators take actions, grounded in sound assessment, for the support of learning, and when students can understand and track their learning, the achievement of all students will rise, and the differences between different groups of students (e.g., minority status, EL status) will diminish. This will reduce the persistent reliance on intervention programs to make up learning deficits that should be a function of strong teaching in core instruction. Investments in short- and medium-cycle systems that strengthen core instruction will be offset with savings in the reduced need for interventions over time.

Evidence for the effectiveness of an ideal comprehensive and balanced assessment system should be collected and used. Such evidence should include evidence of student learning (did it improve? in what way(s)?) and evidence of the student self-efficacy for learning and self-regulation of learning that a student-centered view of learning entails. Additional academic evidence, such as students' understanding of their learning goals, and academic-related evidence, such as student conscientiousness, perseverance, and collaboration, should also be monitored. A comprehensive and balanced assessment system will be ideal to the extent that it supports student learning on outcomes that matter most, does not hurt students, comports with current understandings of how students learn, and contributes to a well-functioning learning culture in classrooms, schools and districts.

Assessment literacy. Assessment literacy is a term with a quarter-century of history at this point (Stiggins, 1991). Originally referring to educators' understanding of how to produce and interpret high-quality student achievement data, the term has broadened to include the understanding of other stakeholders, including students, parents, and policy makers, needed to participate in a comprehensive assessment system. Assessment literacy is a well-studied academic phenomenon; Xu and Brown (2016), for example, reviewed 100 studies of teacher assessment literacy. Less obvious to the authors of this white paper is evidence of systematic pursuit of assessment literacy as a regular practice in districts across the country. One big step in "getting there" must be continued professional development for teachers and other educators, and continued education about assessment evidence and results for students, parents, and policy makers like school board members.

Allocation of responsibility for various parts of the system. The authors of this white paper agree with Shepard and Penuel (2018, p. 54) that School districts are the most appropriate locus for the design and development of coherent curricular activity systems because control of curriculum most often rests with districts. School districts are also responsible for teacher professional development, grading policies, and interim testing mandates.

For these same reasons, the ideal comprehensive and balanced assessment system described in

this paper is intended as a district system, not a state system. States do not control curriculum and, while they do control state achievement standards, those standards describe end points or outcomes and not the learning needed to get there. State accountability tests are only one part in the system, over which districts have little or no control. Designing a comprehensive and balanced assessment system remains in the hands of the district.

Within the district's assessment system, allocation of responsibility has been described above and is summarized here. Notice that each component has several layers of responsibility (for implementing the assessment, for supporting and monitoring that the assessment is done well, for interpreting and using results, for communicating with other levels of the system). This multi-layer responsibility is reflected in the fact that each component implies responsibilities for more than one category of stakeholders.

Most responsible parties at each level include:

- Short-cycle classroom formative assessment – students, teachers, and building principals
- Medium-cycle formative assessment – teachers and building principals (and sometimes district administrators)
- Classroom summative assessment (grading) – teachers, building principals, and district administrators
- Long-cycle interim/benchmark assessments [if used] – district administrators, building principals, school teacher teams
- District assessments and state accountability assessments – district administrators (including curriculum coordinators), building principals, and teachers, especially the respective subject-area departments

Improvements in assessment systems and increases in assessment literacy that must accompany them cannot be accomplished by the states. Although constitutional authority for education falls to the states, state education policies and Education Department staff tend to change frequently, making for an unstable state assessment landscape. Moreover, state education agencies are too far from the classroom to design and support systems whose main purpose is to support student learning. Neither can the solution be left solely to universities, as studies have documented the inadequacies of preservice teacher and administrator education in assessment literacy (Stiggins, 1991; Xu & Brown, 2016). The last best hope for improving assessment systems and increasing the assessment literacy of the responsible parties resides at the district level. That is where the main responsibility for the parts of the system lie, and where the benefits and consequences—and thus, presumably, the motivation—accrue.

Alignment of the system. The previous section described issues of shared responsibility so that all stakeholders are responsible for important parts of one or more of the components of the assessment system. These actors will be the means by which the system is aligned. Thus, an important part of their work will be checking that all parts of the system are based on, and give information about, the appropriate standards at the appropriate grain size. The alignment should be deep and based on more than categorization of topics from assessment to assessment. Rather, conceptions of the learning standards and theories of student learning underlying

their instruction and assessment should be coordinated. Wilson (2004, p. 276) calls this "systemic coherence."

Interplay must exist among the components so they work as a system.

Conclusion. Most current district assessment systems are not comprehensive or balanced. At best, the results include less than optimal information for supporting student learning and less than optimal assessment climates in schools, and at worst, can harm students and their teachers. The most vulnerable, especially students who struggle, students of color, and students in poverty, are disproportionately harmed. It will take the concerted efforts of all stakeholders in the district, and a major shift in many educators' understanding of the role of the student in learning and assessment, to improve this situation. This white paper has laid out some issues, described components of an ideal comprehensive and balanced assessment system, and offered some thoughts about getting there. These thoughts are based in research, some of which was cited here, practical experience in teaching and assessing, and a great deal of care and concern about the systems now in place and their harmful effects. The treatment here was brief, as befits a white paper, and needs to be expanded and informed by the work of model and pilot districts willing to take on the challenges of improvement. The authors are convinced this can be done. It will not be easy, but it will be worthwhile.

References

References

Abrams, L. M., & McMillan, J. H. (2013). *The instructional influence of interim assessments: Voices from the field.* In R. W. Lissitz (Ed.), I*nforming the practice of teaching using formative and interim assessment* (pp. 105-133). Information Age Publishing.

American Statistical Association. (2014). *ASA statement on using value-added models for educational assessment.* Alexandria, VA: American Statistical Association.

Au, W. (2007). *High-stakes testing and curricular control: A qualitative metasynthesis.* Educational Researcher, (36)5, 258-267.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E. H., Ladd, H. F., Linn, R. L., . . .Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers.* Washington, DC: Economic Policy Institute.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice.* Maidenhead, Berkshire, England: Open University Press.

Black, P. & Wiliam, D. (1998a). *Assessment and classroom learning.* Assessment in Education, 5, 7-74.

Black, P., & Wiliam, D. (2004). *The formative purpose: Assessment must first promote learning.* In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability* (pp. 20-50): 103rd yearbook of the National Society for the Study of Education, Part II. Chicago, IL: University of Chicago Press.

Brookhart, S. M. (2017). *Formative assessment in teacher education.* In D. J. Clandinin & J. Husu (Eds.) *International handbook of research on teacher education* (pp. 927-943). London: Sage.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L F., Stevens, M. T., & Welsh, M E. (2016). *A century of grading research: Meaning and value in the most common educational measure.* Review of Educational Research, 86, 803-848.

Darling-Hammond, L. (2015). *Can value-added add value to teacher evaluation?* Educational Researcher, 44, 132-137.

DuFour, R. (2004). *What is a professional learning community?* Educational Leadership, 61(8), 6-11.

Ferrara, S., Maxey-Moore, K., & Brookhart, S. M. (in press). *Guidance in the Standards for classroom assessment: Useful or Irrelevant?* In S. M. Brookhart & J. H. McMillan (Eds.), Classroom assessment and measurement. New York: Routledge.

Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P. R., Shavelson, R. J., & Yin, Y. (2008). *On the fidelity of implementing embedded formative assessments and its relation to student learning: A collaboration between curriculum and assessment developers.* Applied Measurement in Education, 21, 360-389.

Gallimore, R., Ermeling, B.A., Saunders, W.M., & Goldenberg, C. (2009). *Moving the learning of teaching closer to practice: Teacher Education Implications of School-based Inquiry Teams.*

Elementary School Journal, 109(3), 537-553.

Goertz, M., Oláh, L., Nabors, & Riggan, M. (2009). *From testing to teaching: The use of interim assessments in classroom instruction* (Vol. RR-65). Philadelphia, PA: Consortium for Policy Research in Education.

Goldhaber, D. D., Goldschmidt, P., & Tseng, F. (2013). *Teacher value-added at the high-school level: Different models, different answers?* Educational Evaluation and Policy Analysis, 35(2), 220–236.

Graham, S., Hebert, M., & Harris, K. R. (2015). *Formative assessment and writing: A meta-analysis.* The Elementary School Journal, 115, 523-547.

Jacob, B. A., & Lefgren, L. (2005). *Principals as agents: Subjective performance measurement in education.* Cambridge, MA: National Bureau of Economic Research.

Konstantopoulos, S., Miller, S., van der Ploeg, A., Li, C.-H., & Traynor, A. (2011, September). *The impact of Indiana's system of diagnostic assessments on mathematics achievement.* Paper presented at the Fall 2011 Conference of the Society for Research on Educational Effectiveness, Washington, DC.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). *The intertemporal variability of teacher effect estimates.* Education Finance and Policy, 4, 572-606.

McElligott, J., & Brookhart, S. M. (2009). *Legal issues of grading in the era of high stakes accountability.* In T. Guskey, (Ed.), Practical solutions for serious problems in standards-based grading (pp. 57-74). Thousand Oaks, CA: Corwin Press.

McTighe, J. (2018). *Three key questions on measuring learning.* Educational Leadership, 75(5), 14-20.

Michigan Assessment Consortium. (2017). *Assessment literacy standards.* Retrieved 9/3/18 from http://michiganassessmentconsortium.org/sites/default/files/mac_AssessLitStds_2017_screen-9.19.17.pdf

Noyce, P. E., & Hickey, D. T. (Eds.) *New frontiers in formative assessment.* Cambridge, MA: Harvard University Press.

Oláh, L. N., Lawrence, N. R., & Riggan, M. (2010). *Learning to learn from benchmark assessment data: How teachers analyze results.* Peabody Journal of Education, 85, 226-245.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Perie, M., Marion, S., & Gong, B. (2009). *Moving toward a comprehensive assessment system: A framework for considering interim assessments.* Educational Measurement: Issues and Practice, 28(3), 5-13.

Ruiz-Primo, M. A., & Brookhart, S. M. (2018). *Using feedback to improve learning.* New York: Routledge.

Saunders, W. M., Goldenberg, C. N., & Gallimore, R. (2009). *Increasing achievement by focusing grade level teams on improving classroom learning: A prospective, quasi-experimental study of title 1 schools.* American Educational Research Journal, 46, 1006-1033.

Schneider, M. C., & Randel, B. (2010). *Research on characteristics of effective professional development programs for enhancing educators' skills in formative assessment.* In H. L. Andrade & G. J. Cizek (Eds.), Handbook of formative assessment (pp. 251-276). New York: Routledge.

Shepard, L. A. (2001). *The role of classroom assessment in teaching and learning.* In V. Richardson (Ed.), Handbook of research on teaching, 4th ed. (pp. 1066-1101). Washington, DC: American Educational Research Association.

Shepard, L. A., & Penuel, W. R. (2018). *Classroom assessment principles to support learning and avoid the harms of testing.* Educational Measurement: Issues and Practice, 37(1), 52-57.

Stiggins, R. J. (1991). *Assessment literacy.* Phi Delta Kappan, 72(7), 534-539.

Stiggins, R. (2014a). *Revolutionize assessment.* Thousand Oaks, CA: Corwin.

Stiggins, R.J. (2014b) *Defensible teacher evaluation: Student growth through classroom assessment.* Thousand Oaks CA: Corwin.

Stiggins, R. (2017). *The perfect assessment system.* Alexandria, VA: ASCD.
Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* Albany, NY: SUNY Press.

Supovitz, J. (2009). *Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform.* Journal of Educational Change, 10, 211-227.

Wiliam, D. (2010). *An integrative summary of the research literature and implications for a new theory of formative assessment.* In H. L. Andrade and G. J. Cizek (Eds.), Handbook of formative assessment (pp. 18-40). New York: Routledge.

Wiliam, D. (2016). *Leadership for teacher learning.* West Palm Beach, FL: Learning Sciences International.

Wilson, M. (2004). *A perspective on current trends in assessment and accountability: Degrees of coherence.* In M. Wilson (Ed.), Towards coherence between classroom assessment and accountability (pp. 272-283): 103rd yearbook of the National Society for the Study of Education, Part II. Chicago, IL: University of Chicago Press.

Xu, Y., & Brown, G. T. L. (2016). *Teacher assessment literacy in practice: A reconceptualization.* Teaching and Teacher Education, 58, 149-162.

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., & Furtak, E. M. (2008). *On the impact of formative assessment on student motivation, achievement, and conceptual change.* Applied Measurement in Education, 21, 335-359.

Zimmerman, B. J., & Schunk, D. H. (2011). *Self-regulated learning and performance.* In B. J. Zimmerman & D. H. Schunk (Eds.), Handbook of self-regulated learning and performance (pp. 1-12). New York: Routledge.

Join us for the 3rd annual

Formative Assessment National Conference

in Washington DC Area on July 29-31, 2019

SESSION HIGHLIGHT

National Panel on Grading for Learning: How Can Grading Support Learning?

The nation's foremost assessment experts will discuss the best practices and policies on grading and how grading can support learning.

PANELISTS

Sue Brookhart, Thomas Guskey, Jay McTighe, Dylan Wiliam, Michael Toth (Moderator)

https://www.dylanwiliamcenter.com/conference/

175 Cornell Rd. Suite 118 | Blairsville, PA 15717
1400 Centrepark Blvd. | West Palm Beach, FL 33401
724.459.2100 | 877.979.3316

www.learningsciences.com