

MEANINGFUL MEASUREMENT

The Role of Assessments in Improving High School Education in the Twenty-First Century

June 2009



© 2009 Alliance for Excellent Education. All rights reserved.

Suggested citation:

L. M. Pinkus, ed., *Meaningful Measurement: The Role of Assessments in Improving High School Education in the Twenty-First Century* (Washington, DC: Alliance for Excellent Education, 2009).

Ordering information:

Copies of *Meaningful Measurement: The Role of Assessments in Improving High School Education in the Twenty-First Century* can be downloaded from the Alliance's website at www.all4ed.org. To request print copies of the report, please visit http://www.all4ed.org/publication_material/order_form. The first copy of the report is complimentary. Additional copies are available at a charge of \$1 per copy to cover shipping and handling costs.

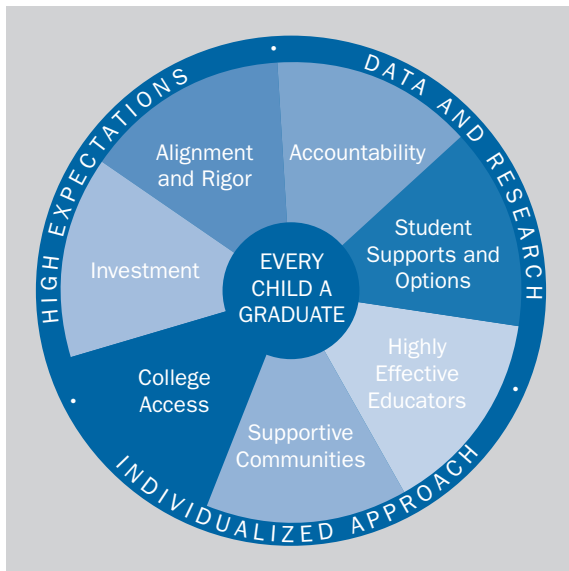
About the Alliance for Excellent Education

The mission of the Alliance for Excellent Education is to promote high school transformation to make it possible for every child to graduate prepared for postsecondary learning and success in life.

The Alliance for Excellent Education is a national policy and advocacy organization, based in Washington, DC, working to improve national and federal policy so that all students can achieve at high academic levels and graduate high school ready for college, careers, and citizenship in the twenty-first century.

The Alliance has developed a “Framework for Action to Improve Secondary Schools” that informs a set of federal policy recommendations based on the growing consensus of researchers, practitioners, and advocates about the challenges and solutions for improving secondary student learning.

The framework, shown graphically here, encompasses seven policy areas that represent key leverage points in ensuring a comprehensive, systematic approach to improving secondary education. The framework also captures



three guiding principles that apply to all of the policy areas. Although the appropriate federal role varies from one issue area to another, they are all critically important to reducing dropouts and increasing college and career readiness.

About the Editor

Lyndsay M. Pinkus is director of strategic initiatives at the Alliance for Excellent Education. Since joining the Alliance in January 2002, she has served in a variety of research, coordination, and advocacy roles, where her work has included managing policy and grant work on a range of issues including graduation rates, data, secondary school accountability, and secondary school improvement, and authoring a number of publications for the Alliance. Prior to rejoining the staff in January 2006, Ms. Pinkus served as a legislative associate at Washington Partners, LLC, providing government relations and policy research and analysis for a variety of clients, including the Alliance. She is a graduate of the School of Public Affairs at American University as a presidential scholar; the Public Affairs and Advocacy Institute at the Center for Congressional and Presidential Studies; and the Institute for Educational Leadership's Education Policy Fellowship program.

Acknowledgments

The Alliance for Excellent Education is greatly appreciative of the authors for sharing their time and expertise in writing the following chapters, as well as of the multiple Alliance staff members and advisors whose dedication contributed significantly to this volume.

The Alliance for Excellent Education is also grateful to Carnegie Corporation of New York for the financial support that made this publication possible.

The views expressed in this volume are those of the authors and do not necessarily represent those of the Alliance for Excellent Education or the funder.

CHAPTER

4

The Role of Interim Assessments in a Comprehensive Assessment System

Judy Wurtzel*

Aspen Institute

Scott Marion, Marianne Perie, and Brian Gong

National Center for the Improvement of Educational Assessment

The standards-based reform movement has resulted in the widespread use of summative assessments designed to measure students' performance at specific points in time. Recognizing that these end-of-year tests are not intended to and do not provide useful information to regularly inform and track student learning during the year, educators are looking for additional assessments to fill that need. Many vendors are now selling what they call “benchmark,” “diagnostic,” “formative,” and/or “predictive” assessments with promises of improving student performance. These assessment systems often lay claim to the research documenting the powerful effect of formative assessment on student learning. However, the research in this area, including the seminal Black and Wiliam (1998) meta-analysis, evaluated formative assessments of a very different character than essentially all current commercially available interim assessment programs.

This chapter differentiates between true classroom formative assessment and the interim assessments currently in the marketplace and provides a framework for considering the appropriate role of interim assessments. It looks at six issues: (1) distinguishing among assessment types; (2) key questions for educational leaders; (3) determining the purpose for the interim assessment; (4) characteristics of an effective interim assessment system; (5) current commercially available interim assessment systems; and (6) implications for district, state, and federal decisionmakers. Our goals are to help district leaders make better decisions about the purchase and use of interim assessment systems and to help state and federal policymakers consider what role they might play in supporting effective interim assessment practices.

Issue 1: Distinguishing Among Assessment Types

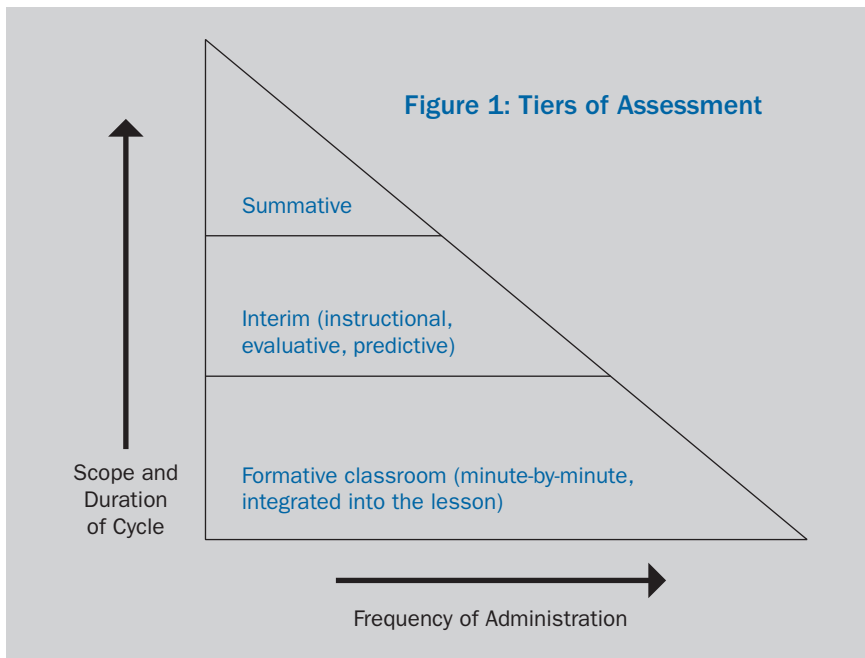
Our schema recognizes three assessment types—summative, interim, and formative—and distinguishes among them based on their intended purposes, audience, and use of the information, rather than simply on when the assessment is given.

Summative assessments are generally given once, at the end of some unit of time (such as the semester or school year), to evaluate students' performance against a defined set of learning targets (e.g., content standards). Because summative assessments are given at the end of a period of instruction, they are not particularly useful for educators to use in adjusting instruction or interventions to address individual student needs. These assessments typically are administered statewide (but can be national or district) and are usually used as part of an accountability program or to otherwise inform policy. State tests mandated under the No Child Left Behind Act (NCLB) are of this type.

Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes.¹ It is done by the teacher in the classroom for the explicit purpose of diagnosing where students are in their learning, where gaps in knowledge and understanding exist, and how to help teachers and students improve student learning. Formative assessment is embedded within the learning activity

and linked directly to the current unit of instruction. The assessments/activities generally are small-scale (a few seconds or a few minutes, and certainly less than a class period) and short-cycle (they are often called “minute-by-minute” assessments, or formative instruction). Tasks or prompts may vary among students depending on the teacher’s judgment about the need for specific information about a student. Providing corrective feedback and modifying instruction are essential aspects of a classroom formative assessment.

Interim assessment is the suggested term for the assessment that falls between formative and summative assessments, including the medium-scale, medium-cycle assessments currently in wide use. Interim assessments evaluate students’ knowledge and skills relative to a specific set of academic goals, typically within a limited time frame, and are designed to inform decisions both in the classroom and beyond the classroom (such as at the school or district level). They may be given at the classroom level to provide information for the teacher, but unlike formative assessments, the results of interim assessments can be meaningfully aggregated and reported at a broader level. As such, the timing of the assessment administration is likely



The triangle in Figure 1 illustrates the distinctions between the three types of assessment.

to be controlled by the school or district rather than by the teacher. Many of the assessments labeled “benchmark,” “formative,” “diagnostic,” or “predictive” fall within our definition of interim assessments.

It is important to begin with these definitions, because even assessment experts have been hobbled by the lack of clear definitions of assessment types. There is concern among some experts that there has been some co-opting of the formative assessment label and research by those purveying nothing of the sort. This imprecision has led to a blurring of the perceived differences between “formative assessment” and “interim assessment.” Districts putting interim assessments in place may be getting important and actionable data, but they are rarely getting the power of true formative assessment practices.

Issue 2: Key Questions for Educational Leaders

When deciding whether to include interim assessments in a district’s overall assessment system, it is important to be clear about the intended purpose and use of interim assessments and how the particular assessments will work in the teaching-learning cycle. As a start, it will be helpful to address the following questions:

1. What do we want to learn from this assessment?
2. Who will use the information gathered from this assessment?
3. What action steps will be taken as a result of this assessment?
4. What professional development or support structures should be in place to ensure that the action steps are taken and are successful?
5. How will student learning improve as a result of using this interim assessment system, and will it improve more than if the assessment system were not used?

The answers to these questions will reveal a theory of action about how assessments will lead to improved student learning and drive many of the design decisions. Importantly, these questions and the associated answers serve as the beginning of a validity argument to support (or refute) the particular assessment system. While this chapter focuses primarily on the first question, all five, especially the last, are essential to consider.

In addition, it is important to reflect on how an intended assessment works in the context of other assessments in use. Large school districts often have a plethora of assessments put in place for a variety of reasons over the course of many years. It is suggested that districts conduct an assessment audit that examines what assessments exist, their intended purposes, the results produced, and the utility of the data. Based on the audit results, districts may be able to eliminate less useful assessments, reduce assessment burden, avoid distracting educators and the public with non-useful results, and reclaim instructional time.

Issue 3: Determining the Purpose for the Interim Assessment

A critical task for policymakers is to answer the first question posed above—“What do I want to learn from this assessment, and why do I want to learn it?”—and then find or develop a set of assessments that best fits that purpose. Despite claims to the contrary, single assessments rarely serve multiple purposes well. They tend to work best when the limited number of purposes have been prioritized explicitly. Interim assessments can be thought of in terms of three general classes of purposes—instructional, evaluative, and predictive—with many specific purposes within each.

A. Instructional purposes: Interim assessments designed to serve instructional purposes should provide results that enable educators to adapt instruction and curriculum to better meet student needs. Within this general category of instructional purposes, policymakers and assessment leaders must go further and prioritize specific instructional purposes to better guide assessment design and/or selection. For example, interim assessments might be used to enrich the curriculum, determine students’ strength and weakness in a particular domain, or provide feedback to students for motivational and metacognitive reasons.

When the purpose is to enrich the curriculum, assessments should be designed to have students explore concepts in greater depth or provide tasks that stretch students and teachers to do things at deeper cognitive levels than they might otherwise. The assessment itself contributes to enriching the instruction. When the purpose is to illuminate the strengths and weaknesses of individuals or groups of students, an

assessment system often contains a bank of items aligned with the state curriculum that teachers can use to create a test to evaluate student learning on the concepts taught to date. Results are reported immediately, and data are disaggregated by content standard (or some other unit of learning), allowing teachers to identify strengths and weaknesses in the students' learning. Ideally, to provide actionable information the assessment is fully aligned with the specific classroom or at least school curriculum and provides more in-depth analyses of student misconceptions along with instructional tools and strategies for improving instruction. When the specific instructional purposes are motivating and providing feedback to students, tasks should engage students and encourage them to wrestle with challenging subject matter knowledge. Quick feedback afforded by computer-based testing programs and rich tasks that make student thinking and productions explicit, such as exhibitions and projects, can both achieve these aims. Unfortunately, many purveyors of computer-based interim assessment systems only provide selected response formats (e.g., multiple choice), thereby delivering on only the fast turnaround part of the promise.

B. Evaluative purposes: The primary goal of interim assessments designed for evaluative purposes is to provide information to help the teacher, school administrator, curriculum supervisor, or district policymaker learn about curricular or instructional choices and take specific action to improve the program. These actions are intended to influence subsequent teaching and thereby, presumably, improving the learning. This can be thought of as a program evaluation designed to change curriculum and instruction over the years. An example would be assessments given at various points throughout the year to provide more details about student performance on instructionally relevant subdomains (e.g., adding simple fractions)—not with the intention of intervening but for evaluating the effectiveness of a program or strategy. Another set of important evaluative purposes is to enforce some minimal quality through standardization of curriculum and pacing guides, centralizing coordination for highly mobile urban student populations and high teacher turnover, or as a lever to overcome differences in learning expectations and grading standards.

C. Predictive purposes: Predictive assessments are designed to determine each student's and groups of students' likelihood of meeting some criterion score on the end-of-year tests or other future outcome. While such predictions are of great interest, they obviously must be coupled with further analysis and action. For example, districts might use predictive assessments as a screener to identify students who are not on track to score proficient on the end-of-year test so that they can be given further probes to determine areas of weakness and be provided with remedial instruction, extra support, and/or tutoring. This scenario highlights the value of having formative, interim, and summative assessment types aligned in a comprehensive system.

D. Multiple purposes: Given constrained resources, it is no wonder that educational leaders are tempted to use a single assessment for as many purposes as possible. Unfortunately, one of the truisms in educational measurement is that when an assessment is designed to fulfill too many purposes it rarely fulfills any purpose well.

This does not mean that certain interim assessment systems cannot fulfill more than one purpose. If the system is intended to provide rich information about individual students' strengths and weaknesses tied to a particular set of curricular goals, then these results can likely be aggregated to the subgroup, school, and/or district level to provide evaluative and predictive information. On the other hand, if the primary goal is to gather predictive or early-warning information, it is unlikely that the assessment will contain rich enough information for full instructional or even evaluative purposes. Therefore, if users want to fulfill multiple purposes, they must design a system to fulfill the finest-grain purposes first and then aggregate the results to more general levels in the educational system. However, users still need to be sure that multiple purposes are not contradictory, such as might be the case when an assessment is used for both instructional and accountability purposes.

Issue 4: Characteristics of an Effective Interim Assessment System to Be Used for Instructional Purposes

Once educational leaders are clear about purposes for the interim assessment system, they still face many additional considerations and decisions about the system's design and implementation. While there is little research evidence about the characteristics of effective interim assessments, our work with states and districts suggests some commonsense guidance. This chapter focuses on characteristics of interim assessments for instructional purposes because most districts appear to want to use assessments for this purpose, most vendors say their assessments can meet that purpose, and there are more concerns about claims for instructional purposes than for evaluative and predictive purposes.

There is no one-size-fits-all assessment, only a best design for a desired use and the existing constraints and resources. Given that, the general characteristics of any interim assessment to be used for instructional purposes should include

- provision for qualitative insights about understandings and misconceptions and not just a numeric score;
- immediate implications for what to do besides reteaching every missed item;
- rich representation of the content standards students are expected to master;
- high-quality test items, including rich open-ended tasks, that are directly linked to the content standards and specific teaching units;
- a good fit within the curriculum (preferably a curriculum aligned to a similar conception of student learning as the formative assessment strategies) so that the test is an extension of the learning rather than a timeout from learning;
- a good fit with curriculum pacing so that students are not tested on content not yet taught;
- clear reporting that provides actionable guidance on how to use the results;
- validation of the uses of the information provided by the assessment;

- administration features (speed, availability of normative information, customization, timing flexibility, adaptability) that match the assessment purposes; and
- professional development for teachers.

While each item on this checklist could be discussed in depth, this chapter focuses on reporting results, inclusion of data in accountability systems, and item type, because of their importance to policymakers.

One strategy for defining the desired characteristics is to focus on reporting. Score reports make test results actionable. Designing the reporting system at the beginning clarifies all the information desired from the assessment. Score reports should be vetted with those who need to use the information—teachers, in most cases, but also school leaders.

Another key issue is item type. Interim assessments can include a wider range of item types than what is typically the case. In particular, extended performance tasks can serve instructional purposes more readily than other interim assessment item types. They enrich the curriculum, increase student motivation by engaging them in meaningful interactions with rich subject matter, and provide opportunities for teachers to learn about student thinking. As long as the results can be aggregated and used at a level beyond the classroom (which can be done through systematic observations, rubrics, and other scoring methods), an assessment with these types of tasks falls under our definition of interim.

Issue 5: Current Commercially Available Interim Assessment Systems

Once decisionmakers have determined purposes and key design characteristics, they can then determine whether to choose among commercially available interim assessment systems or develop their own. Test companies offer interim assessment products, often labeled “formative” or “benchmark,” for a wide variety of purposes. The best current commercially available systems can

- provide an item bank reportedly linked to state content standards;

- assess students on a flexible time schedule wherever a computer and, if necessary, an Internet connection are available;
- provide immediate or very rapid results;
- highlight content standards in which more items were answered incorrectly; and
- link scores on these assessments to the scores on end-of-year assessments to predict results on end-of-year assessments.

Many of the better commercially available interim assessment products can address questions such as:

- Is this student on track to score “proficient” on the end-of-year NCLB tests?
- Is the student improving over time?
- What proportion of students are at risk of scoring below proficient on the end-of-year NCLB tests?
- On which content standards are the students performing relatively well (or poorly) (for a student, classroom, school, district, state)?
- How does this student’s performance compare to the performance of other students in the class?

Unfortunately, most commercially available interim assessment systems currently do not

- address multiple purposes (i.e., instructional, evaluative, or predictive) well;
- provide rich detail about the curriculum assessed;
- help teachers understand the nature of a student’s misconception(s);
- report detailed information on the student’s depth of knowledge on a particular topic;
- further a student’s understanding through the type of assessment task; and
- give teachers the information on how to implement an instructional remedy.

Furthermore, these systems typically do not answer the following questions:

- Why did a student answer an item incorrectly?
- What are some possible strategies for improving performance in this content area?
- What did the student learn from this assessment?
- What type of thinking process is this student using to complete this task?

Given this analysis, there is continued concern about the weakness of available interim assessment systems for instructional purposes. Nonetheless, interim assessments can play a productive role in this and other areas. In particular, in terms of predictive and evaluative purposes, interim assessments can help districts determine whether all schools have similarly high standards, whether highly mobile students have exposure to a common curriculum, and which students are “off track” so they can intervene.

Issue 6: Implications for District, State, and Federal Decisionmakers

Districts clearly set the policy and practice context for effective use of interim assessments. While interim assessments are typically not used for policy purposes at the state level, and not at all at the federal level, there is also a considerable role for states—and a limited role for the federal government—in support of effective interim assessment use. Productive roles for leaders at the district, state, and federal levels are suggested below.

Leadership: Establishing and maintaining the vision

Leaders at all levels must be clear and coherent as they articulate their vision for learning, instruction, assessment, and school quality and then work to ensure consistency among the various initiatives and policies. This broad vision should support the meaningfulness of formative and interim assessments. For example, leaders who dwell on the large-scale summative results as the only important measure of school quality can undermine the use of other assessment approaches. Leaders should conduct assessment audits and, when designing assessment systems, be thoughtful about their intended purposes, the results produced, and the usefulness of the data.

These steps will help eliminate less useful assessments, reduce the assessment burden, and reclaim instructional time.

Standards, assessment, and curriculum design decisions

In a standards-based environment, interim assessment practices are driven by decisions about standards, curriculum, and state assessments. Leaders can insist on high-quality college- and career-readiness standards that are developed according to the most up-to-date learning theories, which support the assessment of specific learning targets in ways that best facilitate learning and instruction. They can ensure that state assessments focus on a limited number of meaningful outcomes and use rich item formats; this signals to the field the types of learning outcomes and tasks that are valued, and thus supports better interim and formative assessment practices. They can support the development and use of challenging curriculum that includes embedded formative and interim assessments to equip teachers with the tools to translate the ambitious learning goals articulated in good standards into meaningful, rich learning experiences for students. While most of these decisions are made by state leaders, there is an important federal role in supporting state efforts through funding support, accountability policies (see below), and the R&D investments needed to develop these next-generation standards, assessments, and curriculum.

Accountability policies

While there is no hard evidence on the best approach, our sense is that the results of interim assessments should be made public within a district (among teachers, administrators, and parents) but should not be used for school or district accountability purposes. This is particularly true if assessments are to be used for instructional purposes and the goal is for teachers to use assessment results as the basis for conversations among themselves and with their students about the nature of students' work and the changes in their own practice that are needed to improve this work. For such conversations and analyses to take place, teachers must believe in—and not fear—the assessment results. However, state policies, such as those found in Wyoming and Rhode Island, where results of local interim assessments (including portfolios of student work) are used for graduation certification can direct attention and efforts toward improving local

assessment practices. (It should be noted that in both of these states there has been a significant amount of state support to help local districts build and use sound assessment systems.)

Funding and other resources

Money matters! It is cheaper to score multiple-choice items than constructed-response items or performance tasks, and it often costs less to buy a computer-based testing system than to invest in professional development for all teachers. Even within the reality of constrained budgets, saving a few dollars on an assessment system might actually “cost” more in terms of opportunities for learning. State and federal governments can help underwrite the costs of effective assessment systems and associated professional development. Beyond new money, leaders can reallocate existing resources to support formative and other local assessment initiatives in ways that make clear their importance.

Professional development policies

A consistent research finding is that the effectiveness of any test used for instructional purposes is dependent on how the teacher uses the information to give feedback to the students. Teachers need to learn how to administer the assessment, learn from the results, and adjust instruction accordingly. While districts must take the lead in this area, states can play a valuable role in funding, designing, and/or providing such professional development. The federal government can support this as well. Moreover, the state should work with teacher credentialing and teacher education institutions to make this a more salient part of teacher pre-service training.

Quality control

The state can be invaluable in the area of quality control, and the federal government can aid those efforts through funding and research. First, it can vet potential vendors and provide information on the characteristics of the assessments available, quality of the items, and the degree of alignment with state curriculum. The state department of education may simply provide the information to districts or choose to allow state funds to be spent only on interim assessment systems that meet specific qualifications. Secondly,

the state can provide access to research various interim assessments and best practices in their use. Thirdly, the state department of education can network districts using similar assessment systems.

Evaluation

Given the lack of research supporting the use of interim assessment and the many questions about the power and validity of different types of interim assessments, it is suggested that decisionmakers at all levels deliberately and continuously evaluate the effectiveness of interim assessment strategies within and across districts and adjust accordingly. Evaluations should include teacher surveys or focus groups to determine how the data were used and if there is evidence that the information gained from interim assessments improved student learning. Leaders should critically consider whether investments in improving daily classroom assessment practices are a more effective strategy for improving student learning than purchasing interim assessments.

Conclusion

Our hope is that policymakers will take at least six points from their reading of this chapter.

First, interim assessments, as defined in this chapter, are distinct from formative assessments. While a definition may seem trivial, it is clear that the many terms currently used to describe interim assessments (benchmark, periodic, predictive, formative) have impeded clear discussions and decisions about whether and how interim assessments should be used.

Second, the research supporting the efficacy of assessment to improve teaching and learning is based on formative assessment—minute-by-minute classroom assessment. While interim assessment has considerable intuitive appeal, there simply is no research base to support the claim that interim assessments improve student learning.

Third, there are useful and valid purposes for interim assessments within a comprehensive assessment system. However, in deciding whether implementing an interim assessment system is an appropriate strategy

and, more specifically, what interim assessment design is appropriate, policymakers must go through the analysis laid out above of the purpose and expected use of interim assessment.

Fourth, policymakers should evaluate commercially available assessments cautiously. In particular, if policymakers desire interim assessments to serve instructional purposes, they should ask whether they meet the suggested criteria of effective assessments.

Fifth, policymakers should seek to eliminate the “zone of wishful thinking” in the design and implementation of interim assessment systems. Policymakers often hope that data will automatically lead to improved practice. However, experience shows that data must be accompanied by the reporting systems, professional development, support structures, and management practices that will impact teacher and student beliefs and behaviors. Each of these elements should be considered at the initial phases of designing or selecting and implementing an interim assessment system.

Finally, as with any assessments, policymakers should regularly ask whether the benefits of interim assessments outweigh the costs in terms of instructional time, teacher time, and fiscal resources. Further, they should be considered in light of the possibility of providing professional development to implement true formative assessment practices. One reason school districts invest in interim assessment systems rather than promoting formative classroom assessment may be that they lack the capacity to implement formative assessment well at scale. As Black and Wiliam have noted, “The improvement of formative assessment cannot be a simple matter. There is no quick fix that can alter existing practice by promising rapid rewards. On the contrary, if the substantial rewards promised by the research evidence are to be secured, each teacher must find his or her own ways of incorporating the lessons and ideas set out above into his or her own patterns of classroom work and into the cultural norms and expectations of a particular school community. This process is a relatively slow one and takes place through sustained programs of professional development and support.”²

Leaders interested in improving formative assessment practices should support interim assessment systems designed with explicit attention to

increasing teachers' ability to do formative classroom assessment. The choice of item types, the format of reports and data analysis, and the structure and content of professional development can be carried out in ways that help teachers learn how to embed assessment within a learning activity, provide immediate corrective feedback, and modify instruction to meet students' needs. Over the long term, the focus of assessment efforts can move from interim assessment to the formative assessment practices that research suggests have the most payoff for student learning.

(This chapter was adapted from *The Role of Interim Assessments in a Comprehensive Assessment System: A Policy Brief*, by Marianne Perie, Scott Marion, Brian Gong of the National Center for the Improvement of Education Assessment, and Judy Wurtzel of the Aspen Institute.)³

The views expressed in this chapter are those of the authors and do not necessarily represent those of the Alliance for Excellent Education.

About the Authors

Judy Wurtzel* is codirector of the Aspen Institute Program on Education and Society. The program helps local, state, and national education leaders share knowledge about how school systems can improve the education and life chances of poor and minority students, and works with them to create programs and policies to accomplish these goals. Current initiatives include the Aspen Senior Congressional Staff Network, the Aspen Urban Superintendents' Network and complementary Urban Literacy and Mathematics Leadership Networks, and a project on rethinking human capital in urban school districts. From 1999 to 2005, Ms. Wurtzel was executive director of the Learning First Alliance, a permanent partnership of twelve leading educational associations dedicated to improving public education. During the Clinton administration, she was a senior adviser to the deputy secretary of the U.S. Department of Education, working on a range of elementary and secondary education issues.

Scott Marion is the vice president of the National Center for the Improvement in Educational Assessment, where his current projects include evaluating the technical quality of state alternate assessment systems, exploring the instructional usefulness of interim assessment approaches, and helping states design valid accountability systems. Dr. Marion serves on the U.S. Department of Education's National Technical Advisory Committee and on a national research committee investigating the issues

and challenges associated with incorporating value-added measures in educational accountability systems. Dr. Marion received his PhD in measurement and evaluation from the University of Colorado at Boulder, an MS in science education from the University of Maine, and a BA in biology from the State University of New York. Prior to joining the Center for Assessment, Dr. Marion was most recently the director of assessment and accountability for the Wyoming Department of Education and was responsible for overseeing the Wyoming Comprehensive Assessment System and designing the technical and policy structures to implement a multiple-measures, locally created graduation assessment system. Dr. Marion regularly publishes and presents the results of his work in peer-reviewed journals and at several national conferences. For more information about the Center for Assessment and Dr. Marion, please see www.nciea.org.

Marianne Perie is a senior associate with the National Center for the Improvement of Educational Assessment. She received her PhD in educational research, measurement, and evaluation from the University of Virginia. Prior to joining the center, she worked on district, state, and international assessments as well as the National Assessment of Educational Progress (NAEP) as an employee of first the American Institutes for Research and then the Educational Testing Service. Her primary interests are standard setting, reporting, accountability, and validity studies. She has conducted standard-setting studies in more than sixteen states, districts, and foreign countries. She taught a course in standard setting as part of the federally funded Graduate Certificate Program and coauthored a revision of the 1982 publication *Passing Scores*, published in 2008 as *Cutscores*. Dr. Perie is currently working with several states on exploring a validity argument on alternate assessments for students with significant cognitive disabilities and enhancing their technical documentation. For other publications, see www.nciea.org and click on “publications.”

Brian Gong is the executive director and cofounder, with Rich Hill, of the National Center for the Improvement of Educational Assessment. Dr. Gong’s experience as the associate commissioner for curriculum, assessment, and accountability in the Kentucky Department of Education, in addition to his previous work in research and development at Educational Testing Service, allows him to bring a multifaceted approach to solving the complex educational problems. Dr. Gong was the author of an important Council of Chief State School Officers publication on the design of educational accountability systems, and continues to advise several of the council’s projects. Dr. Gong received his PhD in curriculum and instruction from Stanford University. Additionally, Dr. Gong currently serves as an advisor to the U.S. Department of Education as it considers how to incorporate student growth-based models into state accountability systems for NCLB.

*On May 19, 2009, Judy Wurtzel was named deputy assistant secretary for planning, policy, and evaluation at the U.S. Department of Education.

¹ L. Shepard, "Benchmark Assessments Be Formative?: Distinguishing Formative Assessment from Formative Program Evaluation," paper presented at the CCSSO Large Scale Assessment Conference, June 2006, San Francisco, CA.

² P. Black and D. Wiliam, "Assessment and Classroom Learning," *Educational Assessment: Principles, Policy and Practice* 5, no. 1 (1998): 7–74; P. Black and D. Wiliam, "Inside the Black Box: Raising Standards Through Classroom Assessment," *Phi Delta Kappan* 80 (October 1998).

³ M. Perie, S. F. Marion, B. Gong, and J. Wurtzel, *The Role of Interim Assessments in a Comprehensive Assessment System: A Policy Brief* (Washington, DC: Achieve, Inc., the Aspen Institute, and the National Center for the Improvement of Educational Assessment Inc., 2007).