



# USING ASSESSMENTS

## TO IMPROVE LEARNING AND STUDENT PROGRESS

FOR THE LAST YEAR, DURING THE COURSE OF CONVERSATIONS WITH KEY STAKEHOLDERS ACROSS THE U.S., PEARSON'S LEADERSHIP HAS RECOGNIZED FOUR PREVAILING TOPICS OF INTEREST OR CONCERN. THE LEVEL OF THIS INTEREST COMPELLED US TO TAKE A DEEPER LOOK INTO THESE TOPICS; AS A RESULT, WE DEVELOPED FOUR ISSUES PAPERS THAT EXAMINE THE UNDERLYING RESEARCH AND DEFINE EFFECTIVE SOLUTIONS. THESE THEMES OR ISSUES—COLLEGE READINESS, TEACHING QUALITY, ASSESSMENTS FOR LEARNING, AND TECHNOLOGY AND EDUCATIONAL ACHIEVEMENT—HAVE SIGNIFICANT IMPLICATIONS FOR THE FUTURE PROSPERITY OF THE UNITED STATES.

### WE ARE ON THE EDGE OF A SIGNIFICANT OPPORTUNITY

to grow student learning through the use of formative assessment systems, but this will require a disciplined approach to some key fundamentals before we ask teachers to embrace new strategies or programs.



Our issues papers are not intended to be exhaustive overviews of the current state of American education. Rather, they outline the issues and associated challenges, and point to potential solutions that exhibit demonstrable results. The papers offer the reader, whether a legislator, administrator, school board member, teacher, or parent, a scan of the existing literature and a perspective on approaches that have demonstrated progress. For example, the discussion about teaching quality, perhaps the single most significant variable that influences student achievement, will consider the return on an investment in effective teaching and professional development. The technology paper broadens the dialogue on the potential efficacy and efficiency that can be harnessed with the latest advances in learning science and enterprise resources. The college readiness paper focuses on factors that contribute to a student's ability to succeed in higher education.

This particular paper examines the role of assessment in the teaching and learning process, and explores what we can do to help teachers and students utilize both "assessments for learning" and "assessments of learning" to gauge and improve student progress.





## INTRODUCTION

Assessments provide objective information to support evidence-based or data-driven decision making. We often categorize assessments by purpose (admissions, placement, diagnostic, end-of-course, graduation, employment, professional certification, and so on) but it may be more revealing to understand who uses the information from an assessment to guide their actions and to inform their decisions. The role of assessment in students' learning is commonly misunderstood, perhaps because each stakeholder group—students, parents, teachers, administrators, policy makers, and the business community—has different information needs. The purpose of this paper is to create a better understanding of how the formative use of assessment information can accelerate the teaching and learning process.

Dr. Richard J. Stiggins, author of the article *Assessment in Crisis: The Absence of Assessment FOR Learning* (Stiggins, 2003), states the following:

**“The evolution of assessment in the United States over the past five decades has led to the strongly held view that school improvement requires:**

- The articulation of higher achievement standards
- The transformation of those expectations into rigorous assessments and
- The expectation of accountability on the part of educators for student achievement, as reflected in test scores

**Standards frame accepted or valued definitions of academic success. Accountability compels attention to these standards as educators plan and deliver instruction in the classroom. Assessment provides one aspect of the evidence of success on the part of students, teachers, and the system.”**

Note, however, that this evolution of assessment as cited does not mention instruction or learning as a key element in assessment. This is because the primary users of the information generated by accountability assessments are not students and teachers, but administrators and policy makers. We need to understand how to create a more balanced system of assessments that can meet the information needs of students and teachers in addition to supporting the need for accountability assessments that drive standards-based education reforms.

The assessment and accountability provisions of No Child Left Behind (NCLB) Public Law 107-110, 2001 set targets for student achievement. In large part due to these targets, and their associated consequences, educators and policy makers have focused on how to accelerate student achievement for all students. A balanced system that combines assessments *for* learning with assessments of learning can provide

a rich stream of information and timely feedback loops that support the work of the teacher and student. There is a growing body of evidence that the use of well-constructed, externally developed assessments by teachers can lead to larger gains in student performance than other education reform strategies.



## ASSESSMENT *OF* LEARNING DEFINED

Assessments of learning are designed to provide information and insight on how much students have learned during the school year, whether standards are being met, and whether educators are helping their students learn. They are not designed to create information to be used to inform specific instruction as learning takes place. A misunderstanding of the use and value of end-of-year NCLB assessments has been partially responsible for common criticisms of NCLB. Such end-of-year assessments are valuable measures of what has been learned. The diagnostic lens resulting from these assessments can only be used for improving instructional practice or systemic improvements, rather than improving the teaching and learning process itself, because the information from the assessments is not available to students and teachers in a timely manner, and the assessments are summative measures of what has been learned over a course of instruction; they are not designed for use during a course of instruction.

The national focus on summative assessment has frustrated some educators, suggesting that too much emphasis is being placed on this one kind of assessment information. As consideration is being given to a revision of the national assessment policy, education leaders are weighing in, suggesting that a more balanced focus on both assessments of learning (reporting on what has been learned) as well as assessments for learning (designed to provide useful and timely feedback to teachers and students as learning progresses) is necessary if we are going to raise standards and ensure that every student is on a path toward proficiency.



## ASSESSMENTS *FOR* LEARNING

Assessments for learning help teachers diagnose students' learning needs, inform students about their own learning progress, and allow modification of the instructional approach to improve learning.

Dr. Richard J. Stiggins stated the following in *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning* (Reeves, 2007):

**“Examples of assessments for learning are those that we use to diagnose student needs, support students’ practice, or help students watch themselves improving over time. In all cases, we seek to provide teachers and students with the kinds of information they need to make decisions that promote continued learning.**

**Assessments for learning occur while the learning is still happening and throughout the learning process. So early in the learning, students’ scores will not be high. This is not failure—it simply represents where students are now in their ongoing journey to ultimate success.”**

Assessments for learning provide the continuous feedback in the teach-and-learn cycle, which is not the intended mission of summative assessment systems. Teachers teach and often worry if they connected with their students. Students learn, but often misunderstand subtle points in the text or in the material presented. Without ongoing feedback, teachers lack qualitative insight to personalize learning for both advanced and struggling students, in some cases with students left to ponder whether they have or haven’t mastered the assigned content.

Assessments for learning are part of formative systems, where they not only provide information on gaps in learning, but inform actions that can be taken to personalize learning or differentiate instruction to help close those gaps. The feedback loop continues by assessing student progress after an instructional unit or intervention to verify that learning has taken place, and to guide next steps. As described by Nichols, Meyers, and Burling:

**“Assessments labeled as formative have been offered as a means to customize instruction to narrow the gap between students’ current state of achievement and the targeted state of achievement.... The label formative is applied incorrectly when used as a label for an assessment instrument...reference to an assessment as formative is shorthand for the particular use of assessment information, whether coming from a formal assessment or teachers’ observations, to improve student achievement. As William and Black (1996) note: ‘To sum up, in order to serve a formative function, an assessment must yield evidence that...indicates the existence of a gap between actual and desired levels of performance, and suggests actions that are in fact successful in closing the gap.’ ”**

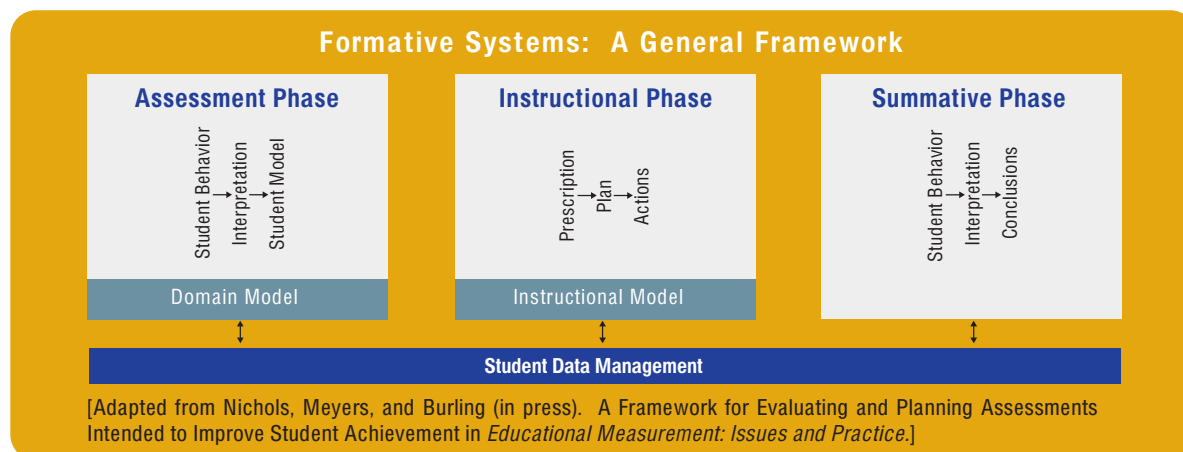
**STUDENTS LEARN, BUT OFTEN MISUNDERSTAND** subtle points in the text or in the material presented. Without ongoing feedback, teachers lack qualitative insight to personalize learning for both advanced and struggling students, in some cases with students left to ponder whether they have or haven’t mastered the assigned content.



## **A FRAMEWORK** for Designing and Evaluating Formative Systems

Paul D. Nichols, Jason Meyers, and Kelly Burling at Pearson are conducting research on a general framework for designing and evaluating formative systems. Assessments labeled as formative require empirical evidence and documentation to support the claim inherent in the label “formative” that improvements in learning are linked to the use of assessment information. They have developed a framework that can be used to design formative systems and evaluate evidence-based claims that the assessment

information can be used to improve student achievement. This work has many important implications for understanding when the use of assessment information is likely to improve student learning outcomes and for advising test developers on how to develop formative assessment systems. A general framework for formative systems is depicted in the following diagram:



## ASSESSMENT FOR LEARNING— Early Evidence

In 1998, The Nuffield Foundation commissioned Professors Paul Black and Dylan Wiliam to evaluate the evidence from more than 250 studies linking assessment and learning (Assessment for Learning, 1999). Their analysis was definitive—initiatives designed to enhance effectiveness of the way assessment is used to promote learning in the classroom can raise pupil achievement. A new term, “assessment for learning,” was established.

Learning outcomes demonstrated that the linking of assessment with instruction gained between one-half and three-quarters of a standard deviation unit—very large on almost any scale. This suggests that assessments for learning can deliver improved student achievement at a greater pace than many other interventions. Research including Black and Wiliam (Black and Wiliam, 1998) indicates that improving learning through assessment depends on five surprisingly simple key factors:

- The provision of effective (and continuous) feedback to students
- The active involvement of pupils in their own learning
- Adjusting teaching to take account of the results of assessment
- A recognition of the profound influence assessment has on the motivation and self-esteem of pupils, both of which are crucial influences on learning
- The need for pupils to be able to assess themselves and understand how to improve

By combining large-scale summative assessments of student learning with smaller in-school formative assessments for learning, educators can create a more comprehensive representation of student progress.

What is missing in assessment practice in this country is the recognition that, to be valuable for instructional planning, assessment needs to be a moving picture—a video stream rather than a periodic snapshot.

The table below shows the effect, in the number of additional months of progress per year, of three different educational interventions. The estimate of the effectiveness of class size is based on the data generated by Jepsen and Rivkin (2002) and the estimate for teacher content knowledge is derived from Hill, Rowan, and Ball (2005). The estimate for formative assessment (assessment for learning) is derived from Wiliam, Harrison, and Black (2004) and other small-scale studies.

Intervention	Extra Months of Learning Gained Per Year
Class-size reduction by 30%	3
Increase teacher content knowledge from weak to strong (2 standard deviations)	1.5
Formative Assessment (Assessment for Learning)	6 to 9

The data in this table provides additional evidence that when a teacher properly uses assessment, learning gains will follow. It is important to note that these studies suggest that assessment for learning outperforms both class-size reduction and improving teacher content knowledge as an effective intervention to accelerate learning gains in the classroom.



## COMPARING THE *ROLE OF STUDENT LEARNING* As a Part of Assessment of Learning and Assessment for Learning

Formative and summative assessments play important roles in driving and understanding student progress, but it is important to understand how each different type of assessment is used by both educators and students.

The following table provides a comparison of assessments of learning and assessments for learning\*:

A COMPARISON	
Assessment of Learning	Assessment for Learning
Strives to document achievement	Strives to increase achievement
Informs others about students	Informs students about themselves
Informs teachers about improvements for next year	Informs teachers about needed improvements continuously

## A COMPARISON (CONT'D)

### Assessment *of* Learning

### Assessment *for* Learning

Reflects aggregated performance toward the content standards themselves

Reflects specific student and teacher iterations to aspects of learning that underpin content standards

Produces comparable results

Can produce results that are unique to individual students

Produces results at one point in time

Produces results continuously

Teacher's role is to gauge success

Teacher's role is to promote success

Student's role is to score well

Student's role is to learn

Provides administrators program evaluation information

Not designed to provide information for use outside of the classroom

Motivates with the promise of rewards and punishments

Motivates with the promise of success

*\* This table is a customized version of a table first published by the NEA in Balanced Assessment: The Key to Accountability and Improved Student Learning, 2003.*

The next section will expand on the role of both types of assessments and their relationship to student success.



## BALANCED ASSESSMENT: Why We Need Both to Raise Student Performance

As early as 2003, educators and leaders in the educational measurement community started calling for the use of both summative and formative assessments as part of a more balanced assessment system. In 2008, the Council of Chief State School Officers (CCSSO) devoted their Student Assessment Conference toward the need of implementing such balanced assessments.

**“By combining large-scale summative assessments of student learning with smaller in-school formative assessments for learning, educators can create a more comprehensive representation of student progress. This is not to minimize the role of external assessments in favor of internal assessments only. Both assessments of and for learning are important” (Stiggins, 2002), and “while they are not interchangeable, they must be compatible” (Balanced Assessments, 2003). The key to maximizing the usefulness of both types is to intentionally align assessments of and for learning so that they are measuring the same student progress. (Reeves, 2007)**

## STUDENTS ARE LIKELY TO LEARN MORE

if teachers use assessment for learning to improve instruction during specific teacher/student interactions.

Administrators, teachers, and students could all benefit from a balanced assessment approach in their schools:

- Administrators (local, state, and federal) need the results of the assessment for accountability purposes. Such uses imply that administrators are spending their funding allocations appropriately and that, by so doing, students benefit from their school experience.
- Communication between administrators and teachers is essential to share goals and accountability data and to meet the mission of the school. Teachers need to know how to utilize assessment results to improve instruction, thereby increasing the learning of their students. Such simple things like the timing of the assessment (ongoing vs. one snapshot at the end of the year), availability of the results (ongoing vs. after the course of instruction has ended), as well as the alignment of the curriculum standards must all be taken into account by the teacher when using the assessments.
- Students share in this mixed-purpose testing as well. Students will improve their learning and their performance on assessments via the feedback they receive from the assessment. Similarly, students need to document their attainment of the educational goals of learning as part of the assessment as well.
- Parents will also benefit from the balanced assessment approach when they receive ongoing information about specific subject areas and standards that will need additional focus to help their child achieve individual learning goals.

**“What is missing in assessment practice in this country is the recognition that, to be valuable for instructional planning, assessment needs to be a moving picture—a video stream rather than a periodic snapshot.” (Heritage, 2007)**

Clearly, all constituencies need a balanced approach to assessment, one that fills the need for accountability and improved learning via enhanced and targeted instruction. Of additional importance is communication between administrators, teachers, students, and parents.



## IS IT POSSIBLE TO USE ASSESSMENTS to Raise Standards and Promote Learning?

It is possible to use assessments to both raise standards and to promote learning, although few teachers have been formally prepared to construct their own reliable formative assessments or to use the information resulting from assessments as an integral tool in teaching. Assessment literacy, or the lack of such literacy, is the theme of many education conferences and seems to be lightly covered by teacher preparatory programs.

**“Despite the importance of assessments in education today, few teachers receive much formal training in assessment design or analysis. A survey by Stiggins (1999) showed, for example, that less than half the states require competence in assessment for teacher licensure. Lacking specific training, teachers often do what they recall their own teachers doing...they treat assessments strictly as evaluation devices,**

**administering them when instructional activities are completed and using them primarily to gather information for assigning students' grades."**

**Dr. Thomas R. Guskey, *Using Assessment to Improve Teaching and Learning* (Reeves, 2007)**

Clearly, the relationship of both teachers and students with educational assessment needs to be transformed if they are to benefit from assessment information. Training on how to harness the power of assessment is an ever-growing need today.

**"To move classroom assessment into its appropriate role beside standard testing in a balanced assessment system, policymakers must invest in key areas of pre-service education, licensure requirement changes, and professional development in assessment literacy. The paper calls for investments in high quality large-scale assessment and more funding for classroom assessment. These additional resources should be used, in part, to ensure that teachers and administrators have the competencies necessary to use classroom assessment as an essential component of a balanced assessment program." (Balanced Assessments, 2003)**



## **A NEW APPROACH TO ASSESSMENT**

The need to embrace a balanced assessment system in order to improve student learning will require policy and implementation changes for both states and districts. Such changes will allow assessments for learning and assessments of learning to be integrated into a balanced system of assessments. Some states have already taken steps forward in this regard—asking for interim assessments and curriculum and learning support materials. Yet, more is needed.

**"Standards and tests alone cannot improve student achievement. As states raise standards for all students, they need to ensure that teachers and students have the supports essential for success. It will also require a substantial state effort in partnership with local districts to provide current teachers with high quality and engaging curriculum, instructional strategies and tools, formative assessments, and ongoing professional development needed for continuous improvement." (Achieve, 2008)**

What can we do now to better support teacher and student success?

We can work to establish a closer connection between instruction, formative assessment, and summative assessment in every district in the United States. This connection will be strengthened by leaders from all levels working together to establish:

- The development and growth of formal teacher education on assessment for learning
- Funding to support assessment for learning professional development practices and strategies
- Development of tools and reports that allow teachers to manage and use "just in time" results to improve individual student instruction

**WE BELIEVE THAT STANDARDS WILL RISE** and student learning will show gains when teachers and students are given the training and tools to support balanced assessment systems of both formative and summative assessments.



- Student-friendly learning systems, reports, and tools that provide clear information to students for managing their own learning
- Longitudinal data-management systems so student progress can be measured over time, and so educators and parents can project whether a student is on a path to proficiency and important benchmarks, such as college readiness.

Considerable research is being focused on the real impact of creating a closer connection between formative and summative assessments. In a paper published in the February 2003 issue of *Education Policy Analysis Archives*, Meisels et al. examined the change in scores on the Iowa Tests of Basic Skills for 3<sup>rd</sup> and 4<sup>th</sup> graders whose teachers used curriculum-embedded assessment for at least three years. Effect gain sizes were between .7 – 1.5, far exceeding the contrast group’s average changes.

**“Perhaps the most important lesson that can be garnered from this study is that accountability should not be viewed as a test, but as a *system*. When well-constructed, normative assessments of accountability are linked to well-designed, curriculum-embedded instructional assessments, children perform better on accountability exams, but they do this not because instruction has been narrowed to the specific content of the test. They do better on the high stakes tests because instruction can be targeted to the skills and needs of the learner using standards-based information the teacher gains from ongoing assessment and shares with the learner. ‘Will this be on the test?’ ceases to be the question that drives learning. Instead, ‘What should I learn next?’ becomes the focus.” (Meisels, 2003)**



## **REAL IMPROVEMENTS** for Teachers and Students

**“... if we are serious about improving student achievement, we have to invest in the right professional development for teachers. This is an important point, because too often, professional development is presented as a fringe benefit—part of a compensation package to make teachers feel better about their jobs. This certainly seems to be the way teacher professional development is viewed by many outside the world of education, and also, sometimes, by policymakers.”**

Dr. Dylan Wiliam, *Content Then Process: Teacher Learning Communities in the Service of Formative Assessment* (Reeves, 2007)

We can provide assessment literacy training for all teachers, opening the floodgates of knowledge on how to best use reliable, externally developed assessments for learning. This, coupled with targeted professional development, will empower teachers to bring new relevance to the teach-and-learn cycle and will result in improved learning.

Such professional development activities could include:

- Setting of clear learning goals
- Framing appropriate learning tasks

- **Deployment of instruction with appropriate pedagogy to evoke feedback**
- **Use of feedback to guide student learning and teacher instruction**
- **“Assessment for learning” teams in school consortiums or school buildings to support ongoing professional development**

Black and Wiliam (2005) argue that talking about improving learning in classrooms is of high interest to teachers because it is central to their professional identities. Teachers want to be effective and to have a positive impact on student learning.

Teachers need tools to help them manage the collection of data, the analysis of assessment data, and the reporting of the results—to students, parents, and administrators. The use of common tools across a school building, district, or state will allow educators to establish the same language of classroom assessment, sharing their results and experiences through common recording and reporting systems.

**“Teachers who are supported to collect and analyze data in order to reflect on their practice are more likely to make improvements as they learn new skills and practice them in the classroom. Through the evaluation process, teachers learn to examine their teaching, reflect on practice, try new practices, and evaluate their results based on student achievement.” (Speck and Knipe, 2001)**

Students are likely to learn more if teachers use assessment for learning to improve instruction during specific teacher/student interactions.

“Research shows that when students are involved in the assessment process—by co-constructing the criteria by which they are assessed, self-assessing in relation to the criteria, giving themselves information to guide, collecting and presenting evidence of their learning, and reflecting on their strengths and needs—they learn more, achieve at higher levels, and are more motivated. They are also better able to set informed, appropriate learning goals to further improve their learning.” (See Crooks, 1988; Black and Wiliam, 1998; Davies, 2004; Stiggins, 1996; and Reeves, 2007)

Actions to achieve greater student involvement in their own assessment could include:

- **Assessment literacy training for students**
- **Training and implementation of student self-evaluation**
- **More student peer review and collaboration**
- **Involvement of students in their own parent/teacher meetings, allowing them a key role in presenting their own goals and progress to their parents**
- **Informative reporting systems that are designed to provide feedback to individual students concerning their own progress and achievement**

“When done well, assessment can help:

- **Build a collaborative culture**

## TEACHERS WHO ARE SUPPORTED

to collect and analyze data in order to reflect on their practice are more likely to make improvements as they learn new skills and practice them in the classroom.

- Monitor the learning of each student on a timely basis
- Provide information essential to an effective system of academic intervention
- Inform the practice of individual teachers and teams
- Provide feedback to students on their progress in meeting standards
- Motivate students by demonstrating next steps in their learning
- Fuel continuous improvement process
- Serve as the driving engine for transforming a school.”

Dr. Richard DuFour, *Once Upon a Time: A Tale of Excellence in Assessment* (Reeves, 2007)



## WHAT'S NEXT?

A valid question from readers of this paper might be: “What’s next?” After all, many states are facing huge budget deficits and some are actually scoping back their assessment systems, often eliminating some of the more innovative or valuable components (performance assessments, writing assessments, tools, and resources for parents) that are not legislative mandates. President Barack Obama and Education Secretary Arne Duncan will have the opportunity to shape assessment practice in the U.S. and are already calling for higher standards and higher quality assessment systems. In the coming months and years, educational assessments will likely continue to be a focus of attention and debate at the national, state, and local level. Moreover, educational assessments will receive increasing attention at the international level as businesses and communities worldwide compete for talent in a global economy.

Moving forward, we are on the edge of a significant opportunity to grow student learning through the use of formative assessment systems, but this will require a disciplined approach to some key fundamentals before we ask teachers to embrace new strategies or programs. Those fundamentals include:

- Focus on assessment literacy for teachers, students, and parents
- Support for a balanced emphasis on assessment for learning as well as assessment of learning
- Inclusion of assessment for learning in funding initiatives
- Funding and support for securing well-constructed, reliable, externally developed formative assessments for use by teachers in their classrooms
- Ongoing professional development focused on improving consistency across faculty and districts, and on teachers’ use of evidence from assessment results to support learning
- Funding and support for building longitudinal data systems to support more timely and informed decision-making and to implement more sophisticated analyses, including projection models for determining if students are making adequate progress toward learning goals

We believe that standards will rise and student learning will show gains when teachers and students are given the training and tools to support balanced assessment systems of both formative and summative assessments.

Two new studies published in the October 2008 issue of *Applied Measurement in Education (AME)* support this proposition. In the Guest Editor's Introduction to the two studies, Stanford University's Richard J. Shavelson explained the findings:

**“After five years of work, our euphoria devolved into a reality that formative assessment, like so many other education reforms, has a long way to go before it can be wielded masterfully by a majority of teachers to positive ends. This is not to discourage the formative assessment practice and research agenda. We do provide evidence that when used as intended, formative assessment might very well be a productive instructional tool.”**

Teachers must learn how best to adapt formative assessment to their needs and the needs of their students. As pointed out by Black and William,

**“. . . if the substantial rewards promised by the evidence are to be secured, each teacher must find his or her own patterns of classroom work. Even with optimum training and support, such a process will take time.” (Black and William 1998)**

We believe that the education community can work together to support transformational change in how teachers and students use assessment. Teachers, students, administrators, parents, educational researchers, elected officials, and educational service providers can all contribute ongoing support for further research, educational offerings, and easy-to-use common tools, resulting in best practices in assessment that make sense to teachers and students in classrooms across the U.S.

**“As educators, school leaders, and policymakers, we exist in a world where too often assessment equals high-stakes tests. This is a very limited view of assessment... we call for a redirection of assessment to its fundamental purpose: the improvement of student achievement, teaching practice, and leadership decision-making. The stakes could not be higher. We have two alternatives before us: Either we heed the clarion call of Schmoker (2006) that there is an unprecedented opportunity for achieving results now, or we succumb to the complaints of those who claim that schools, educators, and leaders are impotent compared to the magnitude of the challenge before them.”**

Dr. Douglas Reeves, *From the Bell Curve to the Mountain: A New Vision for Achievement, Assessment, and Equity* (Reeves, 2007)



## GLOSSARY

### A

#### **Age-equivalent:**

The measure of a student's ability, skill, and/or knowledge as compared to the performance of a norm group. For example, if a score of 40 on a particular test is said to have an age-equivalent of 8-3, this means that the median score of students who are 8 years and 3 months old is 40. Thus, a student who scores above his chronological age is said to be performing above the norm and a student who scores below his chronological age is said to be performing below the norm.

#### **Alternative assessments:**

Non-traditional methods of assessment and evaluation of a student's learning skills. Usually students are required to complete multifaceted, meaningful tasks. For example, teachers can use student presentations, group tests, journal writing, books, portfolios, compare/contrast charts, lab reports, and research projects as alternative assessments. Alternative assessments often depend on rubrics to score student results.

#### **Authentic assessments:**

Any performance task in which the students apply their knowledge in a meaningful way in an educational setting. Students are required to use higher cognitive ability to solve problems that are part of their reality.

#### **AYP (Adequate Yearly Progress):**

An important part of No Child Left Behind, AYP is the minimum level of improvement a school or district must achieve each year. Schools are evaluated based on student performance on high-stakes standards. For a school to make AYP, every accountability group must make AYP.

### B

#### **Balanced assessment:**

A system that integrates assessments administered at different levels and for different purposes (e.g., state tests for demonstrating AYP, district tests, classroom assessments of various kinds) to improve the learning environment and student achievement. Some of the tests may be standardized (i.e., administered and scored in a standardized way), while others may be less formal in nature.

**Benchmarks:**

A reference point for determining student achievement. Benchmarks are general expectations about standards that must be achieved to have mastery at a given level. Benchmarks ensure that students are able to communicate and demonstrate learned tasks at a particular level. (Benchmark assessments may or may not be used by teachers . . . it depends on who administers them and why/how the results will be used.)

**Bias:**

Bias refers to an assessment procedure or tool that provides an unfair advantage or disadvantage for a particular group of students. Bias exists in many situations and impacts the reliability and validity of the assessment scores. Bias can be found in content, differential validity, mean differences, misinterpretation of scores, statistical models, and wrong criterion.

## C

**Classroom assessment:**

An assessment that is administered in a classroom, usually by the classroom teacher. Such assessments may be summative (used to assign grades) or formative (used to help the student learn certain content or skills before they are assessed in a summative manner). Common examples of classroom assessments include direct observation, checklists, teacher-created tests and quizzes, projects, portfolios, and reports.

**Comparability:**

The process of making a comparison of similar collected data. In assessment it refers to comparing data collected from like or similar groups to make effective decisions. For example, School A can compare data from their results on a norm-referenced test with other schools that have similar demographics to make informed decisions. They can also compare data from similar groups within the school.

**Concurrent validity:**

Measures how scores on one assessment are related to scores on a similar assessment given around the same time. Educational publishers might administer parts of a new form of an assessment to different schools as a screening for the new form and then compare the results to the prior form.

**Construct:**

A hypothetical phenomenon or trait such as intelligence, personality, creativity, or achievement.

**Construct validity:**

The extent to which an assessment accurately measures the construct that it purports to measure and the extent to which the assessment results yield accurate conclusions.

**Constructed response item:**

A type of test item or question that requires students to develop an answer, rather than select from two or more options. Common examples of constructed response items include essays, short answer, and fill-in-the-blank.

**Content validity:**

A measure of how well a specific test assesses the specific content being tested. For example, if a teacher created a final exam that only included items from the last three units of study, it would have a low content validity.

**Criteria:**

A set of values or rules that teachers and students use to evaluate the quality of student performance on a specific task. Criteria are useful in determining how well a student product or performance has matched identified tasks, skills, or learning objectives. Teachers identify and define specific criteria when developing rubrics.

**Criterion-referenced assessment:**

As opposed to a norm-reference assessment that compares student performance to a norm group of students, a criterion-referenced assessment compares student performance to a criterion or set of criteria, such as a set of standards, learning targets, or a rubric.

## F

**Form:**

In assessment, a form is a particular version of a particular test. Publishers test multiple forms at each grade level each year to ensure that the assessment is valid and reliable.

**Formative assessment:**

A process or series of processes used to determine how students are learning and to guide the instructional process. It also helps teachers develop future lessons, or ascertain specific student needs. Because of their diagnostic nature, it is generally inappropriate to use formative assessments in assigning grades.

## H

**High-stakes test:**

Tests or assessments are high stakes whenever consequences are attached to the results. For example, data from state accountability assessments are published publicly and can often have significant implications for schools and districts; and, schools and districts that do not achieve AYP are subject to sanctions. At an individual level, students who perform poorly may be at risk of failing a class or not being promoted.

## I

### **Item:**

In assessment, an item is an individual question or prompt.

## M

### **Metacognition:**

The knowledge of an individual's cognitive processes that includes monitoring his or her learning. Metacognition strategies often help learners plan how to approach a specified task, monitor comprehension, and evaluate results.

### **Multiple measures:**

Using more than one assessment process or tool to increase the likelihood of accurate decisions about student performance. In some cases, this might mean multiple snapshots in time, such as classroom assessments, a checklist, and standardized assessments.

## N

### **NCLB or No Child Left Behind:**

The No Child Left Behind Act of 2001 was signed into law by President George W. Bush on January 8, 2002 in order to “close the achievement gap with accountability, flexibility, and choice so that no child is left behind.” This act greatly expanded the federal government's role in developing educational policy. While the intent of this act has many goals, such as improving academic achievement of the disadvantaged, preparing, training, and recruiting high quality teachers, and promoting parental choice, the emphasis on accountability and high-stakes testing has caused much controversy.

### **Norm group:**

A reference group consisting of students with similar characteristics to which an individual student or groups of students may be compared. For the comparative data to be valid, the norm group must match in terms of urbanicity, geographic region, and socioeconomic status.

### **Norm-referenced assessment:**

As opposed to criterion-based assessments that compare student performance to a pre-determined set of criteria, standards, or a rubric, norm-referenced assessments are used to compare a student's performance to other students with similar characteristics and who took the same test under similar circumstances.

## O

### **Objective scoring:**

Scoring that does not involve human judgment at the point of scoring.

### **Outcomes:**

Outcomes are results. They may be expressed in terms of a numerical score, an achievement level, or descriptively.

## P

### **Percentile:**

A type of score associated with norm-referenced tests. A percentile score indicates the percentage of students within the same group scored below a particular score. For example, a student receives a raw score of 35 and a percentile rank of 78. This means the student scored higher than 78% of his peers.

### **Performance-based assessment:**

An assessment that requires students to carry out a complex, extended learning process in which a product is required. The assessment requires that content knowledge and/or skills be applied to a new situation or task. Often performance-based assessments mimic real-world situations.

### **Portfolios:**

A meaningful collection of student works for assessment purposes. The collection may include student reflections on the work as well. Portfolios are most often intended to highlight the individual's activities, accomplishments, and achievements in one or more school subjects. A portfolio usually contains the student's best works and can demonstrate a student's educational growth over a given period of time.

### **Proficiency:**

A description of levels of competency in a particular subject or related to a particular skill. In scoring rubrics, proficiency refers to the continuum of levels that usually span from the lowest level (below the standard) through the highest level (exceeding the standard) with specific performance indicators at each level.

### **Prompt:**

A direction, request, or lead-in to a specific assessment task. Teachers often use writing prompts, for example, to direct students to write about a specific subject. (As in, "Write a three-paragraph essay to persuade teachers to start recycling programs in their classrooms.") Prompts can be used to direct students as to what content should be included, how the task should be approached, and/or what the product should be.

## Q

### **Qualitative:**

Analysis or evaluation of non-statistical data such as anecdotal records, observations, interviews, photographs, journals, media, or other kinds of information in order to provide insights and understandings. A survey that asks open-ended questions would be considered qualitative.

### **Quantitative:**

Analysis or evaluation designed to provide statistical data. Results are limited to items that can be measured or expressed in numerical terms. Although the results of quantitative analysis are less rich or detailed than qualitative analysis, the findings can be generalized and direct comparisons can be made.

## R

### **Raw score:**

The number of items a student answered correctly. Because every test is different, raw scores are not as useful as derived scores that describe a student's performance in reference to a criterion or the performance of other students.

### **Reliability:**

The reliability of a specific assessment is a numerical expression of the degree to which the test provides consistent scores. Reliability in testing tells whether the assessment will achieve similar results for similar students having similar knowledge and/or skills relative to the assessment tasks. It will also tell if a student will score relatively the same score on a test or assessment if they take that assessment again. Reliability is considered a necessary ingredient for validity.

### **Rubric:**

A rubric is a guideline that provides a set of criteria against which work is judged. A good rubric ensures that the assessment of these types of tasks is both reliable and fair. A rubric can be as simple as a checklist, as long as it specifies the criteria for evaluating a performance task or constructed response. A rubric details the expectations for the performance of an objective and the scale against which each objective is evaluated.

## S

### **Scale:**

A scale is the full range of possible scores for a specific task or item. For example, a true-false question has a scale of two, either correct or incorrect. Rubrics for performance tasks often use a four- to six-point scale.

**Selected response:**

An item type that requires a student to select the correct response from a set of provided responses. There may be as few as two options or as many as needed. Examples of selected response items include multiple choice, matching, and true/false questions.

**Self-assessment:**

A process used by students to evaluate their own work to determine the quality of the work against specified learning objectives. Students also determine what actions need to be taken to improve performance.

**Standard deviation:**

The average amount that a set of scores varies from the mean of that set of scores. Scores that fall between -1.00 and +1.00 of the mean are said to be in the “normal range” and include approximately 68% of all scores in any given set of scores.

**Standardized testing:**

This term refers to how a test is administered and scored. A standardized test must maintain the same testing and scoring conditions for all participants in order to measure performance that is relative to specific criteria or similar groups. Standardized tests have detailed rules, specifications, and conditions that must be followed.

**Standards:**

Statements about educational expectations of what students are required to learn and be able to do. Content standards describe WHAT students are expected to learn in terms of content and skills while performance standards describe HOW WELL students must perform relative to the content standards in order to demonstrate various levels of achievement (e.g., basic, proficient, advanced).

**Standards-based assessments:**

Assessments designed to align with a particular set of standards such as a state’s content and performance standards. They may include selected response, constructed response, and/or performance-based items to allow students adequate and valid opportunities to perform relative to the standards.

**Subjective scoring:**

Scoring conducted by a human scorer. Assessments that are most often scored subjectively include essays, open-ended response items, portfolios, projects, and performances. When assessments will be used for decision-making it is important for every scorer to be reliable with the other scorers. This is usually accomplished through the use of clear rubrics. Despite having a clear scoring key or rubric, two scorers might disagree on the score of a particular response. It is preferred to have two or more scorers look at items that require subjective scoring, especially on high-stakes tests.

**Summative assessment:**

Assessments that occur after instruction and provide information on mastery of content, knowledge, or skills. Teachers use summative assessments to collect evidence of learning and assign grades.

**T****Teach-and-learn cycle:**

A research-based instructional approach through which educators teach, assess, report, diagnose, and prescribe solutions to improve student achievement.

**V****Validity:**

The validity of a specific assessment is a numerical score that indicates whether the content or skill being assessed is the same content or skill that the students have learned. Validity helps educators make accurate conclusions concerning student achievement. It enables educators to make accurate interpretations about student learning.

**W****Weighted scoring:**

A method of scoring that assigns different scoring values to different questions of a test. This requires conversion of all scores to a common scale in order to make a final grade.



## REFERENCES

- Achieve (2008). American Diploma Project Algebra II End-of-Course Exam: 2008 Annual Report. Washington, DC. URL <http://www.achieve.org>.
- Assessment for Learning: Beyond the Black Box (1999). Assessment Reform Group, Cambridge: School of Education, Cambridge University.
- Balanced Assessment: The Key to Accountability and Improved Student Learning. (2003). National Education Association, Student Assessment Series, (16).
- Black, P. J., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8–21.
- Black, P. J., and Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.
- Black, P. J., and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–75.
- Cech, Scott J., “Test Industry Split Over Formative Assessment,” *Education Week*, September 17, 2008.
- Crooks, T. (1988). The impact of classroom evaluation on students. *Review of Educational Research*, 58(4), 438–481.
- Davies. A. (2004). *Finding Proof of Learning in a One-to-One Computing Classroom*. Courtenay, BC: Connections Publishing.
- Furtak, Erin Marie, Ruiz-Primo, Maria Araceli, Shemwell, Jonathan T., Ayala 1, Carlos C., Brandon, Paul R., Shavelson, Richard J., and Yin, Yue (2008). “On the Fidelity of Implementing Embedded Formative Assessments and Its Relation to Student Learning,” *Applied Measurement in Education*, 21:4, 360–389.
- Heritage, Margaret H. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan* 89(2).
- Hill, H. C., Rowan, B., and Ball, D. L. (2005). Effects of teachers’ mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.
- Jepsen, C., and Rivkin, S. G. (2002). What is the tradeoff between smaller classes and teacher quality? (NBER Working Paper #9205). Cambridge, MA: National Bureau of Economic Research.

Meisels, S. J., Atkins-Burnett, S., Xue, Y., Nicholson, J., Bickel, D. D., and Son, S-H. (2003, February 28). Creating a system of accountability: The impact of instructional assessment on elementary children's achievement test scores, *Education Policy Analysis Archives*, 11(9). URL <http://epaa.asu.edu/epaa/v11n9/>.

Popham, James W. (2008). *Transformative Assessment*, Association for Supervision and Curriculum Development, Alexandria, Virginia. ISBN 978-1-4166-0667-3.

Poskitt, Jenny, and Taylor, Kerry (2008). National Education Findings of Assess to Learn (AtoL) Report, New Zealand Ministry of Education. URL <http://www.education-counts.govt.nz/publications/schooling/27968/27984/2>.

Reeves, Douglas, Editor, Ainsworth, Larry, Almeida, Lisa, Davies, Anne, DuFour, Richard, Gregg, Linda, Guskey, Thomas, Marzano, Robert, O'Connor, Ken, Stiggins, Rick, White, Stephen, and William, Dylan. *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning*, Solution Tree, Indiana, ISBN 978-1-934009-06-2.

Schmoker, M. (2006). *Results now: How we can achieve unprecedented improvements in teaching and learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

Shavelson, Richard J. (2008). "Guest Editor's Introduction," *Applied Measurement in Education*, 21:4, 293–294.

Speck, M., and Knipe, C. (2001). *Why Can't We Get It Right? Professional Development in Our Schools*. Corwin Press, Inc. Thousand Oaks, CA.

Stiggins, R. (1996). *Student-Centered Classroom Assessment*. Columbus, OH: Merrill.

Stiggins, Richard J. 1999. Assessment, student confidence, and school success. *Phi Delta Kappan* 81(3).

Stiggins, Richard J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan* 83(10), 758–765.

Yin, Yue, Shavelson, Richard J., Ayala, Carlos C., Ruiz-Primo, Maria Araceli, Brandon 1, Paul R., Furtak, Erin Marie, Tomita, Miki K., and Young, Donald B. (2008). "On the Impact of Formative Assessment on Student Motivation, Achievement, and Conceptual Change," *Applied Measurement in Education*, 21:4, 335–359.

