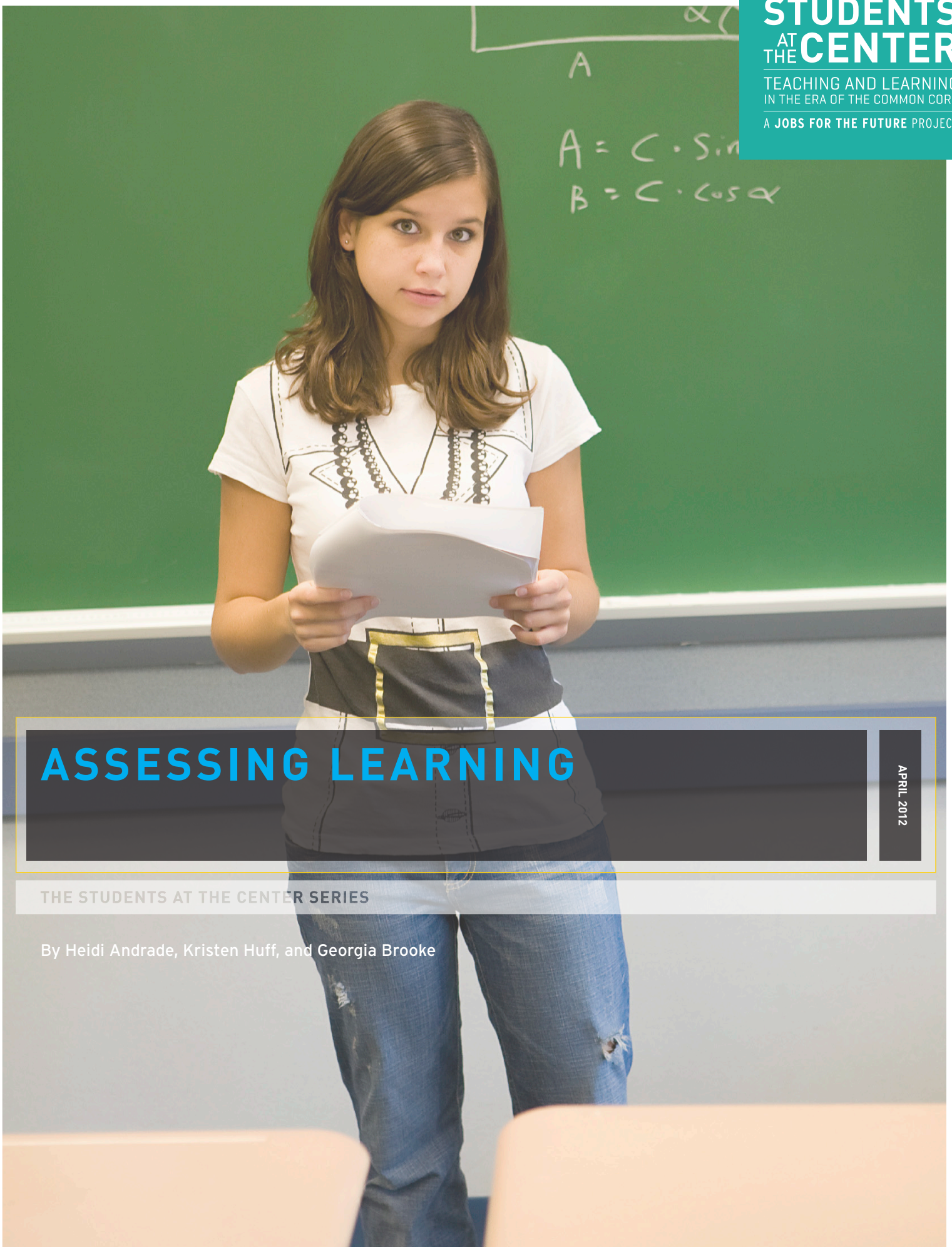


STUDENTS AT THE CENTER

TEACHING AND LEARNING IN THE ERA OF THE COMMON CORE

A **JOBS FOR THE FUTURE** PROJECT

ASSESSING LEARNING

APRIL 2012

THE STUDENTS AT THE CENTER SERIES

By Heidi Andrade, Kristen Huff, and Georgia Brooke

EDITORS' INTRODUCTION TO THE STUDENTS AT THE CENTER SERIES

Students at the Center explores the role that student-centered approaches can play to deepen learning and prepare young people to meet the demands and engage the opportunities of the 21st century. *Students at the Center* synthesizes existing research on key components of student-centered approaches to learning. The papers that launch this project renew attention to the importance of engaging each student in acquiring the skills, knowledge, and expertise needed for success in college and a career. Student-centered approaches to learning, while recognizing that learning is a social activity, pay particular attention to the importance of customizing education to respond to each student's needs and interests, making use of new tools for doing so.

The broad application of student-centered approaches to learning has much in common with other education reform movements including closing the achievement gaps and providing equitable access to a high-quality education, especially for underserved youth. Student-centered approaches also align with emerging work to attain the promise and meet the demands of the Common Core State Standards.

However, critical and distinct elements of student-centered approaches to learning challenge the current schooling and education paradigm:

- Embracing the student's experience and learning theory as the starting point of education;
- Harnessing the full range of learning experiences at all times of the day, week, and year;
- Expanding and reshaping the role of the educator; and
- Determining progression based upon mastery.

Despite growing interest in student-centered approaches to learning, educators have few places to which they can turn for a comprehensive accounting of the key components of this emerging field. With funding from the Nellie Mae Education Foundation, Jobs for the Future asked nine noted research teams to synthesize existing research in order to build the knowledge base for student-centered approaches to learning and make the findings more widely available.

The topic of this paper, as with each in the series, was selected to foster a deeper, more cohesive, research-based understanding of one or more core elements of student-centered approaches to learning. The authors in this series: synthesize and analyze existing research in their areas; identify what is known and where gaps remain related to student-centered approaches to learning; and discuss implications, opportunities, and challenges for education stakeholders who put students at the center. The authors were asked to consider the above definition of student-centered approaches, but were also encouraged to add, subtract, or critique it as they wished.

The authors were not asked explicitly to address the Common Core State Standards. Nevertheless, the research proceeded as discussions of the Common Core were unfolding, and several papers draw connections with that work. The thinking, learning, and teaching required for all students to reach the promised outcomes of the Common Core provide a backdrop for this project. The introductory essay looks across this paper and its companion pieces to lift up the key findings and implications for a new phase in the country's quest to raise achievement levels for all young people.

The nine research papers are loosely organized around three major areas of inquiry—learning theory; applying student-centered approaches; and scaling student-centered learning—although many of the papers necessarily cross more than one area:

1. **LEARNING THEORY:** What does foundational and emerging research, particularly in the cognitive and behavioral sciences, tell us about how students learn and about what motivates them to learn?
   Mind, Brain, and Education
   Christina Hinton, Kurt W. Fischer, Catherine Glennon
   Motivation, Engagement, and Student Voice
   Eric Toshalis, Michael J. Nakkula

2. **APPLYING STUDENT-CENTERED APPROACHES:** How are student-centered approaches to learning implemented? What is the nature of teaching in student-centered learning environments? How can students who are underrepresented in postsecondary education be engaged earlier and perform well in the math and reading activities that scaffold learning? How are advances in technology customizing curriculum and changing modes of learning to meet the needs of each student?

   Teachers at Work—Six Exemplars of Everyday Practice
   Barbara Cervone, Kathleen Cushman
   Literacy Practices for African-American Male Adolescents
   Alfred W. Tatum
   Latino/a and Black Students and Mathematics
   Rochelle Gutierrez, Sonya E. Irving
   Curricular Opportunities in the Digital Age
   David H. Rose, Jenna W. Gravel

3. **SCALING UP STUDENT-CENTERED APPROACHES TO LEARNING:** How have schools sought to increase personalization and with what outcomes for learning? What is the relationship between assessment and student-centered approaches? What can districts do to support student-centered approaches to learning?

   Personalization in Schools
   Susan Yonezawa, Larry McClure, Makeba Jones
   Assessing Learning
   Heidi Andrade, Kristen Huff, Georgia Brooke
   Changing School District Practices
   Ben Levin, Amanda Datnow, Nathalie Carrier

A number of distinguished researchers and practitioners serve as advisors to *Students at the Center* including Scott Evenbeck, founding president of the New Community College, City University of New York; Charles Fadel, Visiting Scholar, Harvard Graduate School of Education, MIT ESG/IAP, and Wharton/Penn CLO; Ronald Ferguson, Senior Lecturer in Education and Public Policy, Harvard Graduate School of Education and the Harvard Kennedy School; Louis Gomez, Professor and the John D. and Catherine T. MacArthur Foundation Chair in Digital Media and Learning, Graduate School of Education and Information Studies, UCLA; Susan Moore Johnson, Professor and the Jerome T. Murphy Professor of Education, Harvard Graduate School of Education; Jim Liebman, Simon H. Rifkind Professor of Law, Columbia University School of Law; Miren Uriarte, Professor, College of Public and Community Service, University of Massachusetts, Boston; and Arthur VanderVeen, Vice President, Business Strategy and Development at Compass Learning.

To download the papers, introductory essay, executive summaries, and additional resources, please visit the project website: www.studentsatthecenter.org.

Over the coming months, Jobs for the Future and the Nellie Mae Education Foundation will craft opportunities to engage a broad audience in the conversation sparked by these papers. We look forward to building a shared understanding and language with you for this important undertaking.

Nancy Hoffman, Adria Steinberg, Rebecca Wolfe

Jobs for the Future

JOBS FOR THE FUTURE

Jobs for the Future identifies, develops, and promotes education and workforce strategies that expand opportunity for youth and adults who are struggling to advance in America today. In more than 200 communities across 43 states, JFF improves the pathways leading from high school to college to family-sustaining careers.

WWW.JFF.ORG

The **Nellie Mae Education Foundation** is the largest charitable organization in New England that focuses exclusively on education. The Foundation supports the promotion and integration of student-centered approaches to learning at the middle and high school levels across New England. To elevate student-centered approaches, the Foundation utilizes a strategy that focuses on: developing and enhancing models of practice; reshaping education policies; increasing the body of evidenced-based knowledge about student-centered approaches and increasing public understanding and demand for high-quality educational experiences. The Foundation's initiative and strategy areas are: District Level Systems Change; State Level Systems Change; Research and Development; and Public Understanding. Since 1998, the Foundation has distributed over $110 million in grants.

WWW.NMEFOUNDATION.ORG

ABOUT THE AUTHORS

Heidi Andrade, Ed.D., is an associate professor of educational psychology and the Associate Dean for Academic Affairs at the School of Education, University at Albany-State University of New York. Her research and teaching focus on the relationships between thinking, learning, and assessment, with emphases on classroom assessment, student self-assessment, and self-regulated learning.

Kristen Huff, Ed.D., is the Senior Fellow for Assessment at the Regents Research Fund. Her main research interest is integrating evidence-centered design theory and practice into large-scale assessments to ensure that what is valued in student learning is measured on exams and communicated through instructionally-relevant score reports.

Georgia Brooke is an advanced doctoral student at the University at Albany. Her interests include formative assessment, gifted education, and the interplay between psychological and biological forces, including the impact of environmental pollution on cognitive functioning.

TABLE OF CONTENTS

INTRODUCTION

> Assessment practices that provide opportunities for open-ended responses and that enable youngsters to play to their strengths fly in the face of assumptions about uniformity. . . . The good school . . . does not diminish individual differences; it increases them. It raises the mean and increases the variance."
>
> —Elliot W. Eisner in "The Uses and Limits of Performance Assessment," 1999

Student-centered assessment embodies sound assessment practices that can be incorporated into any educational setting but are especially critical in student-centered learning contexts, where active engagement in learning and responsibility for the management of learning are core assumptions (Lea, Stephenson, & Troy 2003). In this paper, we begin to paint a picture of student-centered assessment by discussing existing classroom-based, local, state, and national assessment practices in terms of their role in a comprehensive system and how well each represents our defining characteristics of student-centered assessment. The picture that emerges includes a blend of classroom-based assessments, such as: student self- and peer assessments, formative tests, and portfolios; local assessments, including exhibitions and interim assessments created by teams of teachers to inform instruction; and large-scale assessments, including the criterion-referenced tests that states use for K-12 accountability purposes and the norm-referenced tests that are administered nationwide. We also feature computer-based assessments, which hold special promise in a balanced system.

While all the assessments we discuss play a valuable role, some are more student-centered than others, according to the definition used for the *Students at the Center* project. We point out some of the challenges faced by each type of assessment and outline possibilities for advancements. We conclude by proposing future directions for balanced assessment systems in the context of student-centered learning.

A DEFINITION OF STUDENT-CENTERED ASSESSMENT

Like any good assessment, student-centered assessment articulates appropriately challenging and developmentally appropriate learning targets. It also provides feedback to students, teachers, districts, and states about how to deepen learning. It is valid and reliable for the given context, and it is practicable and efficient (McMillan 2011). Our vision of student-centered assessment has several additional defining qualities. It is:

- Individualized;
- Focused on learning and growth;
- Motivating;
- Amenable to actively engaging students in the regulation of their own learning; and
- Informative and useful to a variety of audiences.

The first and most obvious feature of student-centered assessment is that it is **individualized**. Indeed, how could it *not* center on individual students' strengths, needs, and interests and still be student centered? Individualizing assessment involves differentiating learning targets, assignments, and tasks, providing focused feedback on students' learning (whether they are working alone or in groups), and adjusting teaching and learning processes as needed.

Student-centered assessment also **focuses on learning and growth**. That means it does more than measure and report student learning or the lack thereof—although it does those things as well. Student-centered assessment promotes learning and growth by providing useful feedback to the students themselves, their teachers, and others about what the students need in order to progress toward the learning target. This quality of student-centered assessment echoes modern conceptions of formative assessment in that assessment is a moment of learning, not just grading, ranking, or sorting (Andrade & Cizek 2010; Shute 2008).

Student-centered assessment involves the **active engagement of students** in setting goals for their learning and growth, monitoring their progress toward those goals, and determining how to address any gaps. Also called self-regulated learning, the ability to manage one's own learning and growth is a key type of expertise needed for 21st-century college and career success (Dembo & Seli 2008). Classroom assessment practices such as self-assessment, peer assessment, and portfolios have the potential to not only help students learn core content knowledge and skills, but also to develop important self-regulatory habits (Allal 2010; Andrade 2010).

Perhaps the most surprising aspect of our definition of student-centered assessment is that it is **motivating**. Many people associate being evaluated with mild to moderate anxiety, not motivation, and research has shown that grades can be associated with decreased motivation and lower achievement (Butler & Nisan 1986; Lipnevich & Smith 2008). However, recent studies have shown that formative assessment—particularly detailed, task-specific comments on student work—can activate interest in a task (Cimpian et al. 2007) and result in better performance (Lipnevich & Smith 2008).

» Recent studies have shown that formative assessment—particularly detailed, task-specific comments on student work—can activate interest in a task and result in better performance.

Related Paper in the *Students at the Center* Series[1]

For more on student-centeredness and motivation, see *Motivation, Engagement, and Student Voice*, by Eric Toshalis, Ed.D. and Michael J. Nakkula, Ed.D.

Finally, student-centered assessment is **informative and useful to a variety of audiences**. Daniel Resnick and Lauren Resnick (1985) have said that American students are the most tested and the least examined students in the world. We have test scores coming out of our ears, but we do not yet do a very good job of using assessment information to adapt curricula and instruction. Student-centered assessment provides useful information that stakeholders at all levels—including students, teachers, administrators, parents, districts, and states—can use to support learning. For an example of an assessment that is informative at the local level, consider public exhibitions of student work, which engage an audience from the community in discussions of the quality of student work and learning, and of the education students are getting (Davidson & Feldman 2010). For an example of an assessment that is informative at a broader level, consider a state test that guides policymakers at the district or state level in determining where to allocate limited resources or what programs appear to be working best, where, and for whom.

BALANCED ASSESSMENT

We envision student-centered assessment as part of a balanced system of formative, interim, and summative assessments that, taken together, provide detailed information about student learning to inform learning, instruction, and policy.

Formative assessment is the ongoing, minute-by-minute, day-by-day classroom assessment that is administered in the course of a unit of instruction. The purposes of formative assessment are to: identify students' strengths and weaknesses; foster increased autonomy and responsibility for learning on the part of the student; assist educators in planning subsequent instruction; and aid students in guiding their own learning, revising their work, and gaining self-evaluation skills (Cizek 2010).

Interim assessment is a more formalized, periodic process of measuring student achievement throughout the school year to provide information to educators and policymakers who can adjust curricula and instruction as needed.

Finally, summative assessment is formal, often standardized, and typically administered at the end of a unit of instruction, semester, or year. The primary purpose is to categorize the performance of a student or system in order to, for example, assign grades, award or deny a diploma, make promotion/retention decisions, or classify test takers according to defined performance categories (e.g., basic, proficient, or advanced) (Cizek 2010).

It is necessary to contextualize student-centered assessment in a balanced system of formative, interim, and summative assessment because no one assessment process can inform students' approaches to learning, teachers' approaches to instruction, administrators' school- and district-level decisions, and policymakers' decisions about policy. For example, formative student self-assessment is highly individualized and actively engages students in regulating their own learning, but it is not particularly useful to any audience other than the student. In contrast, summative large-scale assessments provide useful information to district or state policymakers but cannot serve their intended purposes if they are individualized. Only a complete system of formative, interim, and summative assessments can be individualized, focused on learning and growth, motivating, amenable to actively engaging students in regulating their own learning, and capable of generating useful information for a variety of audiences.

Ultimately we argue that a balanced system of assessments, created both inside and outside the classroom, is needed to support student-centered approaches to learning. Such a system may include everything from informal observations of student work to formal, standardized tests.

Only a complete system of formative, interim, and summative assessments can be individualized, focused on learning and growth, motivating, amenable to actively engaging students in regulating their own learning, and capable of generating useful information for a variety of audiences.

STUDENT-CENTERED ASSESSMENT PRACTICES

In this section, we introduce formative, interim, and summative assessment processes, make brief references to relevant research, and discuss the ways in which each process does and does not reflect our definition of student-centered assessment. With the exception of summative exhibitions, formative assessment processes tend to be more student-centered than interim and summative assessments. Table 1 presents an overview of assessment processes discussed here, along with our judgments of the student-centeredness of each.

TABLE 1
STUDENT-CENTERED QUALITIES OF SELECT ASSESSMENT PROCESSES

	INDIVIDUALIZED	FOCUSED ON LEARNING AND GROWTH	MOTIVATING	STUDENT SELF-REGULATION	INFORMATIVE TO A VARIETY OF AUDIENCES
FORMATIVE					
Self-assessments	●	●	●	●	
Peer assessments	●	●	●	●	●
Portfolios	●	●	●	●	●
Tests		●	●	●	●
INTERIM					
Criterion-referenced tests		●			●
SUMMATIVE					
Exhibitions	●	●	●	●	●
Tests based on learning progression		●			●
Diagnostic items		●			●
Large-scale tests					●

FORMATIVE STUDENT-CENTERED ASSESSMENT AT THE CLASSROOM LEVEL

The purpose of classroom-based assessment in a balanced, student-centered system is to provide timely information to students and teachers about where students are in their learning, what gaps in knowledge and understanding exist, and how teachers and students can work to deepen learning (Perie, Marion, & Gong 2009). Particularly student-centered forms of classroom assessment include self- and peer assessment, process portfolios, and formative tests.

series

Related Paper in the *Students at the Center* Series[2]

For more information on teachers in student-centered contexts, see *Teachers at Work—Six Exemplars of Everyday Practice*, by Barbara Cervone and Kathleen Cushman.

SELF-ASSESSMENT

The purpose of self-assessment is to identify areas of strength and weakness in one's work in order to make improvements and promote learning, achievement, and self-regulation (Andrade & Valtcheva 2009). As defined by Paul Pintrich (2000), self-regulation is the tendency to monitor and manage one's own learning. Research suggests that self-regulation and student achievement are closely related: Students who set goals, make flexible plans to meet them, and monitor their progress tend to learn more and do better in school than students who do not (Zimmerman & Schunk 2011). Self-assessment is a key element of self-regulation because it involves awareness of the goals of a task and checking one's progress toward them. As a result of self-assessment, Dale Schunk (2003) found that both self-regulation and achievement can increase.

It is critical to recognize the nature of self-assessment as *formative*. Self-assessment is done on work in progress in order to inform revision and improvement; it is not a matter of having students determine their own grades. Given what we know about human nature, as well as research regarding students' tendencies to inflate self-evaluations that count toward final grades, we subscribe to a purely formative type of student self-assessment—that is, as feedback *for* oneself *from* oneself.

Done correctly, self-assessment can play a fundamental role in a balanced system of student-centered assessment. By encouraging students to critique their own work and explicitly identify both strengths and areas that need improvement, self-assessment is individualized. It involves active student engagement by putting ownership of the assessment process in the students' hands: They are in charge of monitoring progress toward goals by comparing their work to explicit criteria, identifying gaps, and making plans to close those gaps. Student involvement is even greater if their teacher involves students in generating the criteria for a task, perhaps by co-creating a rubric.

The focus of self-assessment is learning and growth: Students generate feedback through the self-assessment process and then have opportunities to use that feedback to improve their work. This process of identifying weaknesses and making improvements can be repeated until mastery is achieved. In this way, self-assessment provides useful information to the students themselves about the quality of their work. However, student self-assessments are of limited usefulness to audiences outside the classroom; hence, the need for other forms of assessment.

Effective self-assessment involves at least three steps:

1. Articulate performance targets. The teacher, the students, or, preferably, both clearly articulate the expectations for the task or performance. Students become better acquainted with the task at hand when they are involved in thinking about what counts and how quality is defined. Co-creating a rubric is an effective way to make expectations clear and readily available to students. A rubric is usually a one- or two-page document that lists criteria and describes varying levels of quality, from excellent to poor, for a specific assignment.

2. Checking progress toward the targets. Students take a first attempt at their assignment, be it an essay, lab report, choral performance, or speech. They monitor their progress on their assignments by comparing their performances-in-progress to the expectations, noting areas of strength and weakness and making plans for improvement.

3. Revision. Students use feedback from their self-assessments to guide revision. This step is crucial. Students, being savvy, will not assess their own work thoughtfully unless they know their efforts can lead to opportunities to make improvements and possibly increase their grades.

A SAMPLE SELF-ASSESSMENT

An example from self-assessment in writing might look like this:

After writing a first draft of a persuasive essay, students underline key phrases in the rubric using a variety of colored pencils.

Next, they use corresponding colors to underline or circle their evidence of having met each criterion in their drafts. For example, students underline "clearly states an opinion" in blue on their rubric, then underline their opinions in blue in their essay drafts. If they cannot find a clearly articulated opinion to underline, they write themselves a reminder to do so in their revision.

To assess one aspect of sentence fluency, they underline "sentences begin in different ways" in yellow on their rubric, use the same yellow pencil to circle the first word in every sentence in their essays, and then say the circled words out loud with an ear for repetition.

And so on, for each criterion and sub-criterion on their rubric.

Heidi Goodrich (1996) has generated a list of conditions that are necessary for effective self-assessment. Students need:

- Awareness of the value of self-assessment;
- Access to clear criteria on which to base the assessment;
- A specific task or performance to assess;
- Models of self-assessment;
- Direct instruction in and assistance with self-assessment;
- Practice;
- Cues regarding when it is appropriate to self-assess; and
- Opportunities to revise and improve the task or performance.

Research has examined the effects of self-assessment in a wide range of content areas, including writing (Evans 2001), mathematics (Ross, Hogaboam-Gray, & Rolheiser 2002), social studies (Lewbel & Hibbard 2001), science (Duffrin et al. 1998), and external examinations (MacDonald & Boud 2003). Findings suggest that student self-assessment can promote achievement and learner autonomy. Student reactions to self-assessment are generally positive, but they report needing support and practice to reap the full benefits of the process (Andrade & Du 2007).

PEER ASSESSMENT

The purpose of peer assessment is for learners to provide feedback to one another on the quality of a product or performance (Topping 2010). Students engaged in peer assessment help one another identify strengths, weaknesses, and target areas for improvement. According to Keith Topping, peer assessment happens both inside and outside of school and across different times and contexts, leading to the development of valuable metacognitive, personal, and professional skills. Similar to self-generated feedback, peer feedback is available in much greater volume and with greater immediacy than is teacher feedback.

Peer feedback can play an important role in a balanced system of student-centered assessment. It is individualized, and it actively engages students in the assessment process as peers familiarize themselves with the assessment criteria, examine a particular piece of work, and identify its unique strengths, weaknesses, and need for improvement—all of which they then discuss with the creator of the work. The focus is on growth through feedback, followed by opportunities to revise, improve, and promote mastery. Peer feedback is informative and useful for all in a peer feedback group as they take turns giving and receiving feedback on works in progress. Like self-assessment, however, peer assessment information has limited value for parents, administrators, and policymakers.

Topping (2010) argues that effective peer assessment involves the following steps:

1. Students and teachers co-create assessment criteria.

2. Peers are placed into pairs or small groups based on similar ability levels.

3. The teacher provides training by modeling how to assess a piece of work using explicit criteria.

4. Students get a checklist with peer assessment guidelines.

5. The activity to be assessed and timeline are specified.

6. The teacher monitors the progress of the peer assessment groups.

7. The quality of the feedback is examined.

8. Reliability is checked by comparing teacher- and peer-generated feedback.

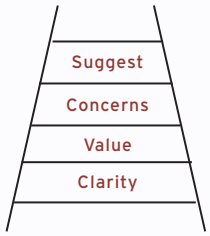
9. The teacher provides feedback to the students about the effectiveness of their assessments.

Research suggests that peer assessment can improve the quality and effectiveness of learning across grade levels, particularly in writing (Yang, Ko, & Chung 2005). Furthermore, both the assessee and the assessor benefit from peer assessment (Topping 2010). As Topping notes, "[L]istening, explaining, questioning, summarizing, speculating, and hypothesizing are all valuable skills of effective peer assessment." While an initial investment is necessary to establish effective peer feedback groups, it is likely to be worthwhile in terms of student learning.

THE LADDER OF FEEDBACK

When delivering peer feedback, it can be very helpful for students to follow a constructive process or protocol. For example, the "Ladder of Feedback" (Perkins 2003) guides users through four steps: clarification; value; concerns; and suggestions.

Ladder of Feedback



Because it is difficult, if not impossible, to give useful feedback on a piece of work the student reviewer does not understand, the first step on the ladder gives the assessor a chance to ask *clarifying* questions about the assessee's work. For example, a peer assessor may ask, "What is your research question?" when giving feedback on a proposal for an experiment to be conducted for a biology class. As the example indicates, questions of clarification are intended to seek information, not praise or criticize.

In the next step, the assessors identify something they *value* in their peer's work to help him or her build on strengths. As at every rung on the ladder, the feedback is phrased in terms of what the assessor values, not in absolute terms that deny the assessor's perspective. For example, "I really like how you propose a study of the organisms in Grafton Lake because it focuses on our own ecology and community" is preferable to "It's good that you will study the lake."

The third rung on the ladder is when the reviewer raises *concerns*. Again, the assessor should own the feedback by using language carefully. For example, "I'm concerned that you have several uncontrolled variables" is more constructive than "You have too many variables that you're not controlling for."

Finally, during the fourth step, the reviewer makes *suggestions* for improvement—for example, "You might consider revising your research questions to focus on only those variables you can control, such as. . . ." These suggestions should be stated as feedback, not mandates.

» Peer assessment can improve the quality and effectiveness of learning across grade levels, particularly in writing. Furthermore, both the assessee and the assessor benefit from peer assessment.

PORTFOLIOS

An academic portfolio is a purposeful collection of student work that includes student involvement in its construction and student reflection on its contents (Belgrad, Burke, & Fogarty 2008). The purpose is to scaffold student reflection and self-regulated learning, as well as to provide nuanced information—about a student's knowledge, dispositions, motivations, and needs—that can help teachers, students, and parents make decisions.

There are two general categories of portfolios: those that showcase a student's best work; and those that demonstrate growth and learning over time (Brookhart 2008). The latter are sometimes called process portfolios or process-folios (Seidel et al. 1997). The key feature of a process portfolio is evidence of students' learning processes and products. For example, a writing process portfolio typically includes several drafts, along with the student's comments on each draft. Together, the writing and the student reflections show improvement over time, with each subsequent piece showing more developed skill than earlier works.

All portfolios are individualized collections of student work that trace progress and highlight strengths via physical artifacts. For portfolios to be successful and student-centered, students must be actively engaged in their creation, especially by setting goals for the learning and achievement, selecting the pieces to include in the portfolio, and reflecting on what those pieces communicate about their own progress toward their goals. In this way, portfolios scaffold self-regulation. Process portfolios are designed to show progression from novice to mastery. Both process and showcase portfolios can be useful and informative to students, parents, teachers, and, sometimes, administrators. However, research on the effectiveness of portfolios suggests they are best used formatively for classroom assessment purposes, rather than summatively as large-scale evaluations, thus limiting their usefulness to audiences outside the school (Brookhart 2008; Herman & Winters 1994).

Barrett (2007) describes the value of electronic portfolios, or e-portfolios, that harness technology to enable archiving, linking, storytelling, collaborating, and publishing. Electronic portfolios use computers and/or the Internet as a container, allowing students to collect and organize their portfolio materials in audio, video, graphics, and text. A study by Chi-Cheng Chang and Kuo-Hung Tseng (2009) suggests that the use of electronic portfolios is positively associated with improvements in student performance, goal setting, problem solving, and reflection.

E-PORTFOLIOS

MAIN SITE: Student portfolios from New Technology High School in Napa, California.[3]

EXAMPLE 1: This one contains more reflection on growth, but does not state explicit goals.[4]

EXAMPLE 2: This one states explicit goals, but contains less reflection.[5]

FORMATIVE USES OF SUMMATIVE TESTS

Traditionally, tests come at the end of a unit of study and are used summatively to determine grades. Formative uses of summative tests involve two testing events that bookend careful analyses of learning by students and teachers. The results of the first test are used formatively, while the results of the second test are used summatively. By way of example, consider a formative test of students' understanding of the Pythagorean theorem. After administering the test but before going over the correct answers, the teacher asks small groups of students to compare answers and resolve any disagreements using their notes, textbook, and other resources. Only after students have had ample opportunity to discover and correct their mistakes does the teacher go over the correct answers with the class. Students then note the kinds of mistakes they tended to make on the test and what they need to study before the next one. A day or two later, the teacher administers the second, summative test, which has different questions but covers the same content.

Formative tests differ in a very important way from practice tests, which usually involve students taking a test, passively listening as the teacher goes over the correct answers, then taking another test. It is not really hearing the correct answers to the test that makes formative use of testing work. Rather,

it is the hard thinking that happens in between the tests that matters (Bloom 1984). This approach to testing is based on Benjamin Bloom's approach to mastery learning, which emphasizes the value of formative assessment and corrective procedures that re-teach content to struggling learners in a new way (Guskey 2010). Research shows that mastery learning is related to learning gains, especially for struggling students, and that it has positive effects on students' attitudes toward course content (Kulik, Kulik, & Bangert-Drowns 1990). In fact, after reviewing meta-analyses from over 40 areas of educational research, Chen-Lin Kulik, James Kulik, and Robert Bangert-Drowns concluded that "few educational treatments of any sort were consistently associated with achievement effects as large as those produced by mastery learning."

Formative uses of summative testing are individualized: they provide information about what each student does and does not know, at least in terms of what was tested. This approach to testing is designed with learning and growth in mind. The explicit goal of the first test is to activate learning about the content of the second test. Although teachers can do all the work of analyzing the results of the first test to identify areas that need to be re-taught, testing at its best actively engages students in the regulation of their own learning when they themselves determine the gaps in their knowledge and make plans for filling in those gaps. Having a grasp of the targets for their learning (as articulated by the first test) and a chance to learn more and earn a higher grade is likely to be motivating, especially to students who need extra time or resources.

Of the four classroom assessment processes discussed in this section, formative uses of summative testing are informative and useful to the widest variety of audiences. The usefulness to administrators can be enhanced if the first test is also used as an interim test and analyzed in terms of the instructional and curricular needs of a class or entire grade level.

INTERIM STUDENT-CENTERED ASSESSMENT AT THE SCHOOL AND DISTRICT LEVELS

School- and district-level assessments tend to be more useful to a wider audience than classroom-level assessments but often at the expense of individualization, student self-regulation, and motivation. Criterion-referenced interim assessments appear to hold special promise for focusing on learning and growth, and thereby informing adjustments to curriculum and instruction.

CRITERION-REFERENCED INTERIM ASSESSMENTS

Schools and districts across the nation are reporting impressive gains in student achievement through the use of teacher-created, criterion-referenced assessments (Bambrick-Santoyo 2008). Such assessments are developed by teams of teachers from within and across schools in particular grades and subject areas; they work together to develop items that directly measure the curricula enacted in their classrooms. The teachers use the same assessments on an interim basis throughout the school year (about every six weeks), get together to discuss the results at length, and share pedagogical approaches to helping one another's students succeed. For example, if Ms. Garcia's third graders all aced the question on 100s place value, but Mr. Lawson's third graders bombed it, the teachers meet so that Ms. Garcia can share with Mr. Lawson how she worked with her students on 100s place value. The key to the success of these efforts is that teachers work together to develop the items, discuss the results, and then adjust their pedagogy accordingly when they return to their classrooms (Bambrick 2008).

Criterion-referenced interim assessments focus on learning and growth and are useful to teachers and administrators. The fact that teachers use the same

Testing at its best actively engages students in the regulation of their own learning when they themselves determine the gaps in their knowledge and make plans for filling in those gaps.

assessments with different classes and in different schools tends to limit their ability to individualize and to engage students in the regulation of their own learning and thereby motivate them—but not necessarily. The student-centeredness of interim assessments would increase sharply if schools used them the way teachers use formative tests: by involving students in analyzing the results of their performance and making plans to deepen their learning.

SUMMATIVE STUDENT-CENTERED ASSESSMENT AT THE SCHOOL LEVEL: EXHIBITIONS

Exhibitions are public demonstrations of mastery that occur at culminating moments, such as at the conclusion of a unit of study or at high school graduation (Davidson 2009). Their purpose is to support sustained, personalized learning while assuring commitment, engagement, and high-level intellectual achievement aligned with established standards. Exhibitions ensure continuity between formative classroom assessments and high-stakes summative assessments by employing teaching, learning, and assessment practices in classroom settings to rehearse, emphasize, and otherwise reinforce progress toward successful final exhibitions. According to Jill Davidson, exhibitions represent a paradigm shift from evaluating academic achievement strictly through "seat time" toward a system of authentic demonstrations of mastery designed to simulate the kinds of open-ended challenges faced by people working in a field of study.

series

Related Paper in the *Students at the Center* Series[6]

For more information on personalized learning, see *Personalization in Schools*, by Susan Yonezawa, Larry McClure, and Makeba Jones.

Exhibitions are a rare example of a summative assessment process that exemplifies each of our characteristics of student-centered assessment. They are individualized to student interests. They involve personalized, ongoing feedback from a variety of sources before the official, summative exhibition. They actively engage students in regulating learning by requiring them to set short-term and long-term goals and to seek out feedback in order to achieve a successful outcome. Because exhibitions are typically presented to an audience that includes practicing experts, they provide an authentic, real-world task that can increase student motivation. By definition, exhibitions are demonstrations of mastery that provide useful information about student learning and achievement to students, teachers, parents, administrators, and community members.

According to Davidson (2009), successful exhibitions:

- Provide multiple opportunities for revision based on frequent feedback;
- Are open to the public;
- Involve school-wide participation;
- Are high stakes (e.g., associated with graduation requirements);
- Occur at culminating moments; and
- Are demonstrations of mastery.

In addition to these key features, Davidson lists four supportive conditions that facilitate the effective implementation of exhibitions:

- School-wide, exhibitions-aligned instructional and assessment processes;
- Structures that support sustained collaboration and inquiry among students and teachers;
- Strong connections with the community outside the school; and
- Active participation in a system or network of other exhibitions-driven schools.

Because exhibitions are typically presented to an audience that includes practicing experts, they provide an authentic, real-world task that can increase student motivation.

While the literature on exhibitions is thin, it suggests that schools placing a central focus on exhibitions have lower dropout rates, higher college-going rates, and improved college performance and persistence (Coalition of Essential Schools 2006).

Video: A Successful Exhibition[7]

Consider the following example of a successful exhibition provided by Davidson (2009):

A tenth-grade science student is interested in the ways that stem cells might provide a cure for multiple sclerosis. The student reviews scientific literature and interviews scientists who are active in stem cell research. Throughout the school year, the student works on a research paper covering the symptoms and effects of multiple sclerosis and offers an analysis of the potential of stem cells as a cure. Before exhibition day, the student submits her paper to an exhibitions committee comprised of teachers and members from the scientific research community.

On the exhibition day, the student gives a public, oral, and multimedia presentation of her research to an audience of her peers, teachers, family members, and community members. She then fields comments and questions and receives a critique of her paper.

Throughout the exhibition process, the student can demonstrate successful mastery of scientific research and inquiry, the conventions of scientific writing, and the ability to think critically.

FORMATIVE, INTERIM, AND SUMMATIVE ASSESSMENT USING ASSESSMENT TECHNOLOGIES

Modern assessment technologies can serve formative, interim, and summative purposes. They give feedback to students and enable teachers to respond to the learning needs of each student with greater speed, frequency, focus, and flexibility. Key features of student-centered assessment technologies include: systematic monitoring of student progress to inform instructional decisions; the identification of misconceptions that may interfere with student learning; rapid feedback to students, teachers, and others; and information about student learning needs during instruction (Russell 2010).

Computer-based assessment programs integrate the management of learning (e.g., organizing student assignments, assessments, and performance), curricular resources, embedded assessments, and detailed student-level and aggregate reporting of strengths and weaknesses. Examples of computer-based programs that feature student-centered assessments include ALEKS, DreamBox Learning, Time To Know, CompassLearning Odyssey, Wowzers, Carnegie Learning, SuccessMaker, and WriteToLearn. Some programs, such as DreamBox Learning and Time To Know, integrate instruction and assessment into one platform. Others, such as WriteToLearn, have a more exclusive focus on assessment.

Perhaps the greatest advantage of these computerized systems in terms of student-centeredness is the degree to which they are individualized. The programs listed above harness the flexible, adaptive capabilities of artificial intelligence to respond to each student's work with detail and immediacy. The feedback generated by these assessment technologies has the potential to motivate students by providing specific information about not only their strengths but also how they can improve their work while they are still engaged in the task. In addition, student-centered computerized assessments also generate information that teachers can use to inform instruction and, in many cases, make summative judgments about student achievement.

» **Perhaps the greatest advantage of computerized systems is the degree to which they are individualized. The programs harness the flexible, adaptive capabilities of artificial intelligence to respond to each student's work with detail and immediacy. The feedback generated by these assessment technologies has the potential to motivate students by providing specific information about not only their strengths but also how they can improve their work while they are still engaged in the task.**

AN EXAMPLE OF A STUDENT-CENTERED ASSESSMENT TECHNOLOGY: WRITETOLEARN

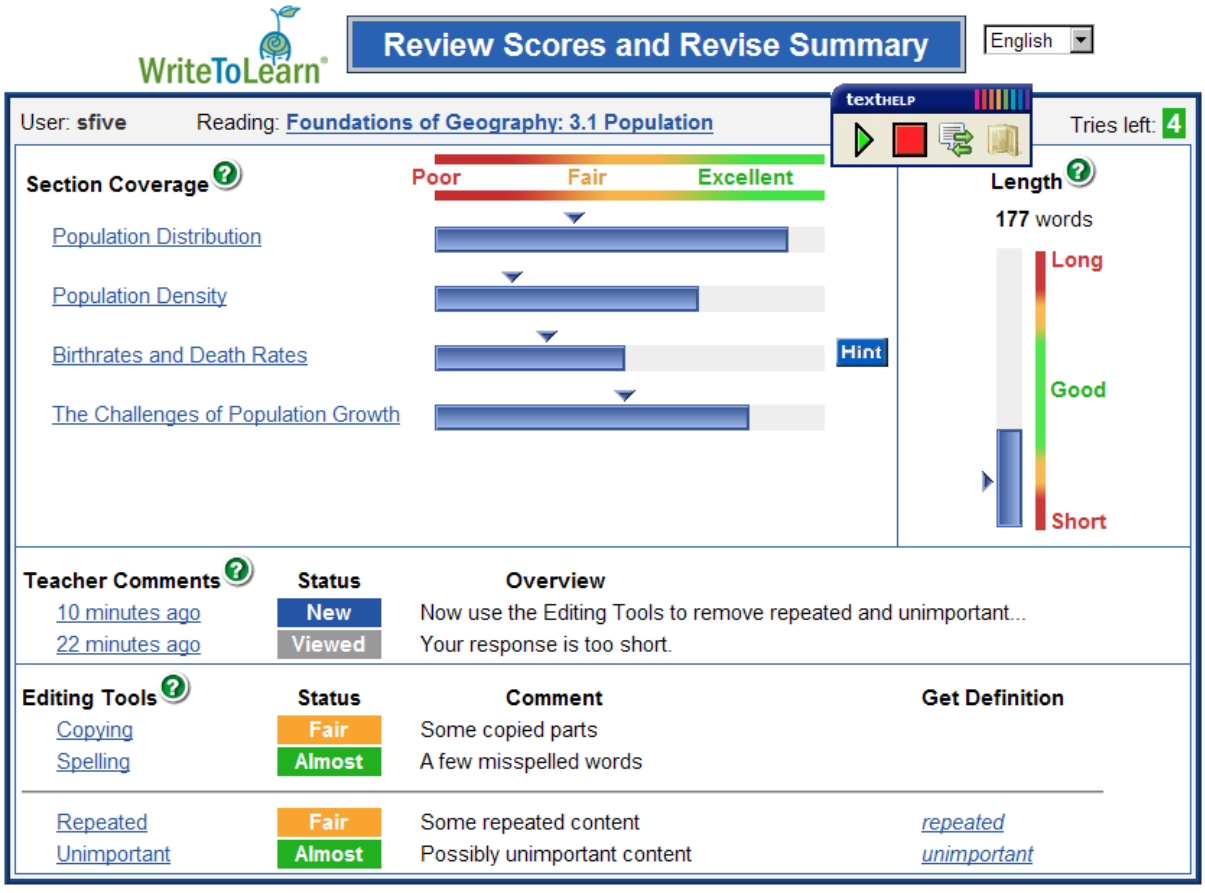
WriteToLearn is a particularly good example of a student-centered assessment technology with strong research support. WriteToLearn promotes reading comprehension and writing skills by providing students with immediate, individualized feedback (Landauer, Lochbaum, & Dooley 2009). The program, designed for students in grades 4 through 12, has two components: Summary Street, where students read and summarize articles or book excerpts; and the Intelligent Essay Assessor, where students write topic-prompted essays.

Summary Street identifies the semantic similarities between students' summaries and the piece they summarized. It then generates a detailed feedback report on content, length, copying, spelling, redundancy, and irrelevancy. Both the teacher and the students get feedback reports. For example, the student feedback screen indicates that the strengths of a student's summary of a passage about geography include its attention to population distribution and density, and that an area in need of improvement is the coverage of birth and death rates (*see Figure 1*). It also contains a suggestion from the teacher about how to use the program to make improvements to the summary.

The Intelligent Essay Assessor mimics the way human scorers make judgments about quality by comparing student essays with hundreds of others that have been expertly scored. The Intelligent Essay Assessor can measure writing traits (e.g., focus, grammar, word choice), and assess writing quality in terms of characteristics (e.g., voice, semantic coherence).

By logging into WriteToLearn, teachers can set up individualized assignments for the students in their classes and then monitor activity and progress. When students log in, they can begin drafting assignments and then submit them for instant, detailed, computer-

FIGURE 1
SCREENSHOT OF WRITETOLEARN'S SUMMARY STREET



Provided by and reproduced with permission of Pearson Education, Inc., publisher of Write-to-Learn

generated feedback. Students can complete several cycles of drafting and revision, while teachers monitor individual and class activity in real time through graphical displays and visualizations. When students are ready, they can submit the assignment to the teacher for summative assessment.

The research on WriteToLearn is promising. One study used a counterbalanced design to find a positive relationship between the use of Summary Street and student summary scores after just two weeks of using the program (Wade-Stein & Kintsch 2004). That study also found that students spent significantly more time on generating summaries than students not using the program, which suggests the program may promote motivation and engagement. Another study, using an experimental design, found that eighth-grade students who used Summary Street scored significantly higher on a test of comprehension than students who did not use the program (Franzke et al. 2005). Student writing in the treatment group was also judged to be better than the writing of students in the control group.

Many assessment technology platforms, especially those listed here, have the potential to play a role in a balanced system of formative, interim, and summative student-centered assessment. They are individualized in terms of the feedback they provide to a variety of audiences, including students, teachers, and others. They are explicitly designed to promote learning and growth through the use of feedback. They can be controlled by students, who can decide when to get feedback and what to do with it. And there is evidence to suggest they can motivate students. Continued research on their effectiveness in student-centered learning environments would be valuable.

series

Related Paper in the *Students at the Center* Series[8]

For more on student motivation, see *Motivation, Engagement, and Student Voice*, by Eric Toshalis and Michael J. Nakkula.

SUMMATIVE STUDENT-CENTERED ASSESSMENT AT THE STATE AND NATIONAL LEVELS

Large-scale assessment tends to be the least student-centered of the processes discussed here. We include them because they are ubiquitous in U.S. schools and unlikely to go away any time soon. Also, and on a more hopeful note, recent advances suggest that large-scale tests can do more than measure and report on a narrow band of student knowledge and skills.

LEARNING PROGRESSION-BASED ASSESSMENTS AND DIAGNOSTIC ITEMS

Learning progressions articulate in detail how knowledge, skills, and abilities change as one moves from less to more sophisticated understanding in a given content domain. Sometimes these are referred to as the building blocks or steps that students need to go through to reach a learning target (Popham 2008). For example, if the learning target is for students to understand that it gets colder at night because part of the Earth is facing away from the Sun's heat, they must first understand that the Earth both orbits around the Sun and rotates on its own axis (*see Table 2 on page 15*).

Learning progressions have the potential to support student-centered assessment in the classroom by revealing for teachers typical preconceptions that students have as they move from less to more sophisticated understanding of a particular learning objective. A recent commission on learning progressions in science went so far as to say (Corcoran, Mosher, & Rogat 2009):

> We are convinced that it is not possible for the reform goals with respect to "all students" to be met unless instruction in our schools becomes much more adaptive. That is, the norms of practice should shift in the direction in which teachers and other educators take responsibility for continually seeking evidence on whether their students are on track to learning what they need to if they are going to reach the goals, along

TABLE 2
EXCERPT OF LEVELS 2 AND 3 FROM A GRADE 4 LEARNING PROGRESSION

LEVEL 2	LEVEL 3
The student recognizes that: > The Sun appears to move across the sky every day > The observable shape of the Moon changes every 28 days The student may believe that the Sun moves around the Earth. COMMON ERROR: All motion in the sky is due to the Earth spinning on its axis. COMMON ERROR: The Sun travels around the Earth. COMMON ERROR: It gets dark at night because the Sun goes around the Earth once a day. COMMON ERROR: The Earth is the center of the universe.	The student knows that: > The Earth orbits the Sun > The Moon orbits the Earth > The Earth rotates on its axis However, the student has not put this knowledge together with an understanding of apparent motion to form explanations and may not recognize that the Earth is both rotating and orbiting simultaneously. COMMON ERROR: It gets dark at night because the Earth goes around the Sun once a day.

Source: Briggs et al. (2006)

> with tracking indicators of what problems they may be having, and then for making pedagogical responses to that evidence designed to keep their students on track. . . . [T]eachers will not be able to engage in such processes until they have in their minds some idea about how students' learning in the subjects . . . develops over time in their school.

Unfortunately, research indicates that teachers may not have the tools that would enable them to use learning progressions as the basis for their assessments. Margaret Heritage and her colleagues (2009) conducted a study of the type of evidence about learning that teachers gathered from reviewing student work, and concluded that most standards and curricula do not provide explicit detail on "how learning progresses in a domain[,] what the precursor skills and understandings are for a specific instructional goal, what a good performance of the desired goal looks like, and how the skill or understanding increases in sophistication from the current level students have reached." Christina Schneider and Bruce Randel (2010) concluded that teachers often do not transform standards into learning objectives at the intended and appropriate levels of complexity. Expecting teachers to use learning progressions as the basis of their student-centered assessments may not be feasible until the necessary standards, curricula, and other instructional support systems are in place.

Fortunately, there is great momentum to use learning progressions as the basis for large-scale assessment (Alonzo, Neidorf, & Anderson forthcoming). The potential contribution of learning progressions to improved assessment design is significant. Learning progressions can inform the development of achievement-level descriptions (ALDs), which are

» Expecting teachers to use learning progressions as the basis of their student-centered assessments may not be feasible until the necessary standards, curricula, and other instructional support systems are in place.

TABLE 3
DIAGNOSTIC ITEM BASED ON A LEARNING PROGRESSION

IT IS MOST LIKELY COLDER AT NIGHT BECAUSE:	
A. The Earth is at the furthest point in its orbit around the Sun.	Level 3
B. The Sun has traveled to the other side of the Earth.	Level 2
C. The Sun is below the Earth and the Moon does not emit as much heat as the Sun.	Level 1
D. The place where it is night on Earth is rotated away from the Sun.	Level 4

Source: Briggs et al. (2006)

the basis of score interpretation and decision making for large-scale, criterion-referenced assessments. Developing ALDs from empirically based learning progressions and using them as the basis for designing large-scale assessments would go a long way toward bridging the gap between large-scale assessment and the reality of how students learn and build knowledge (Huff & Plake 2010).

When learning progressions are used as the basis for designing multiple-choice items, the items can then be used diagnostically. When the information from a learning progression, including the steps toward understanding and typical preconceptions or errors, are used as answer choices, student responses to the item yield information about their understanding that far surpasses what can be learned from simply classifying the student response as right or wrong (Wylie & Wiliam 2007). Table 3 provides an example of an item that can be used diagnostically.

Although these types of items help assessments meet the student-centered assessment criteria for focusing on learning and growth and for being informative to a variety of audiences, there are challenges to using them on large-scale assessments that are designed for summative purposes. Diagnostic items are typically designed to tease out nuanced differences in student misunderstandings; often, they do not meet the technical criteria for summative items whose primary purpose is to determine whether or not a student has met a specific learning target. Researchers are investigating ways to incorporate learning-progression-based diagnostic items into large-scale summative assessments (Alonzo, Neidorf, & Anderson forthcoming).

LARGE-SCALE ASSESSMENTS

Large-scale assessments for K-12 education typically come in two flavors: criterion-referenced and norm-referenced. The most ubiquitous examples of criterion-referenced tests are those that states use for K-12 accountability. These tests are designed to measure a particular set of state-based learning outcomes (or standards). Results for students are reported in terms of the degree to which the student has or has not met the expected performance standard (e.g., basic, proficient, or advanced) and increasingly the degree to which the student has improved since the previous year (i.e., growth measures). In contrast, norm-referenced large-scale assessments are designed to measure less curriculum-based knowledge and skills. They are administered nationwide so that student performance can be reported in terms of national norms. Students from New Mexico to Massachusetts can interpret their nationally normed percentile rank in the same way. For example, the 89th percentile represents the same performance, regardless of whether the student is from Santa Fe or Amherst.

Both statewide criterion-referenced and nationally norm-referenced assessments provide useful information to school and district administrators, policymakers, parents, and other stakeholders. The results enable policymakers to compare performance across a state as well as nationally. Policymakers at the district or state level can use this type of information in determining, for example, where to allocate limited resources and what kinds of educational programs are more successful than others in particular contexts (e.g., schools with

innovative instructional approaches) and/or with particular student groups (e.g., English language learners; students with special needs).

Recent research suggests that criterion-referenced assessments may also have utility within a student-centered learning context. An evaluation of the Adaptive Content with Evidence-based Diagnosis learning system found that it could enhance student learning by providing test takers with elaborated, task-level feedback without compromising the assessment's technical quality (Shute, Hansen, & Almond 2008). The authors conclude that state-mandated tests might be augmented to "support student learning with instructional feedback without jeopardizing the primary purpose of the assessment." Such an augmentation to large-scale tests would go a long way toward making them more effective in promoting learning and growth.

CHALLENGES AND POSSIBILITIES FOR ADVANCEMENT IN LARGE-SCALE ASSESSMENT

To be more in line with the defining criteria of student-centered assessment, large-scale assessment could be improved in three areas:

- Evidence-centered design, which bases assessments on theories of knowing and learning;
- Instructionally relevant score reporting, or richer, more meaningful, more useful feedback on assessment results; and
- Addressing the educational context of a wide array of students.

EVIDENCE-CENTERED ASSESSMENT DESIGN

Conventional approaches to the design of large-scale assessments rely heavily on a psychometric framework that prioritizes the need to rank-order examinees for the purposes of selection, classification, or summative evaluation. Although the primary purpose of standards-based assessments used in statewide testing are not to rank-order students but rather to classify them into performance categories or achievement levels, these tests are still typically designed from the same psychometric framework. Paul Nichols (1994) and others (NRC 2001; Pellegrino et al. 1999; Snow & Lohman 1989) have argued compellingly that educational assessments designed from psychometric models are not optimal for informing instruction because the tasks are not based on an explicit model of how students learn. In addition, scoring models that are primarily used to rank-order students cannot reflect the complexity of a learner's cognitive strengths and weaknesses. Consequently, the test results are not necessarily connected to classroom learning and instruction and have limited utility for educators and students.

To maximize the educational benefits of assessment, exams should be situated within an aligned and integrated system of curriculum, instruction, and assessment (Nichols 1993, 1994; NRC 2001; Pellegrino et al. 1999). In this system, curricula should sequence learning objectives that reflect our understanding of how students build knowledge and expertise in the domain, much like learning progressions do. Instruction should employ strategies that facilitate knowledge building and active learning. Assessment design should be informed by the same cognitive framework that shapes the curriculum and provide feedback to teachers that informs instruction. That is, learning and instruction are optimized when a cognitive model of learning not only provides the framework for assessment design but also provides the framework for the educational system in which the assessment is used.[9] James Pellegrino (2002; NRC 2001) elaborates on this integrated system of curriculum, instruction, and assessment by suggesting a framework that consists of three interrelated elements:

- A model of student learning in the academic domain;
- A set of beliefs (or hypotheses) about the kinds of observations that will provide evidence of student competencies in the domain, where such competencies are defined by the model; and
- A framework for interpreting the results of the assessment.

This general cognitive assessment framework can be put into operation in large-scale assessment through evidence-centered design (Huff, Steinberg, & Matts 2010; Mislevy, Almond, & Steinberg 2002; Steinberg et al. 2003). ECD is an innovative approach to measuring complex, higher-order thinking, and detailing the specific assessment features that are

required to elicit student understandings. Its primary goals are to ensure that: the evidence obtained from an assessment is optimal given the purpose, use, and audience for the results; and the inferences from the assessment are supported by a comprehensive and coherent argument that relates each aspect of the assessment (the questions/tasks, the student performance, and the inferences from the results) back to the model of learning that is at the heart of the assessment enterprise. Evidence-centered design begins with an articulation of:

- The claims about students that are the target of learning and measurement (e.g., the student can synthesize in writing the themes from multiple, sufficiently complex informational texts);
- The observable evidence required in student work to warrant the claims (e.g., whether the characteristics of a masterful essay that synthesizes themes are from multiple, sufficiently complex informational texts); and
- Tasks that are optimally designed to support student work in providing the required evidence (e.g., the requirements for the texts, prompt, and other scaffolds that would best support a student in providing a masterful essay that displays the desired characteristics).

Although many who have not used ECD comment that it is equivalent to what has always been implicit in conventional assessment design, explicitness is one of its primary strengths. For example, skills such as "explain," "analyze," and "synthesize" must be defined in terms of observable features of student work. In turn, the need for observable features of student work shapes the tasks students engage in to provide evidence of their learning. The explicit articulation of claims and evidence enables us to remove the ambiguity inherent in many standards, learning objectives, and test specifications, and it helps ensure that instruction and assessment are aligned. Once the claims, observable evidence, and task features are articulated, they can be used to design any type of assessment, from a classroom assessment to be used formatively to a large-scale assessment to be used nationwide.

Evidence-centered design reflects several of the defining characteristics of student-centered assessment. Imagine a situation where the claims, observable evidence, and task features are explicitly articulated and shared among students, teachers, parents, and policymakers. It would be tremendously empowering, motivating, and informative for all stakeholders to have a shared understanding of the targets of learning and measurement (the claims), the observable evidence of student work that is required, and the types of tasks that can best support students in providing the required evidence of student learning.

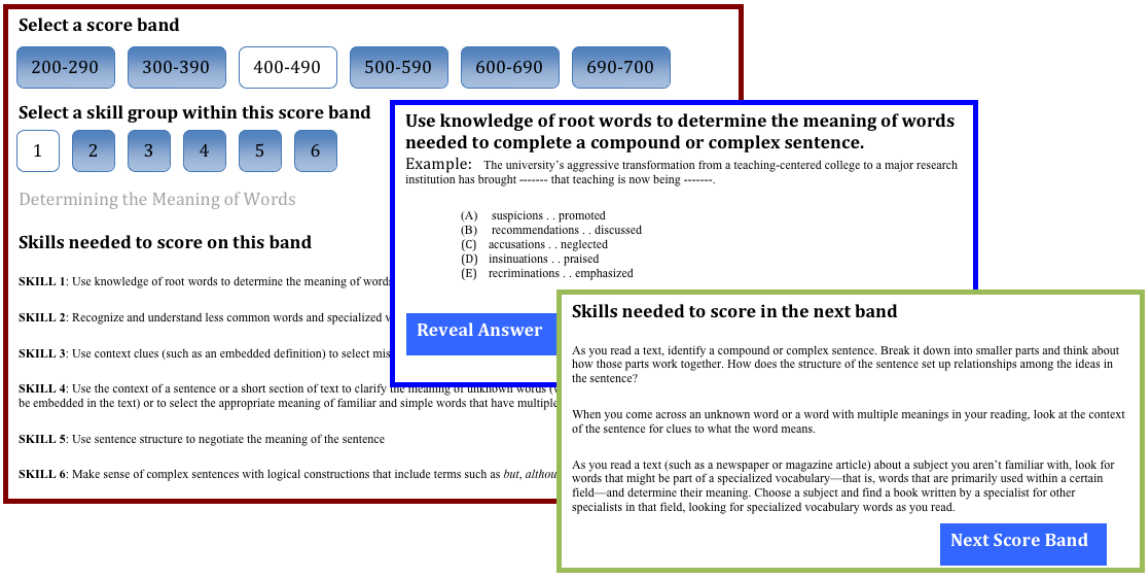
INSTRUCTIONALLY RELEVANT SCORE REPORTING

Another area of improvement for large-scale assessment relates to the types of information provided in score reports. The demand for more instructionally relevant information has increased (Huff & Goodman 2007; VanderVeen et al. 2007). The problem is that criterion-referenced and norm-referenced large-scale assessments are not designed to provide detailed information about the strengths and weaknesses of individual students. If they were, it might undermine their use for summative assessment.

That said, there are ways to improve the types of information that can be provided. First, large-scale assessments could use an item-mapping or scale-anchoring approach (Beaton & Allen 1992) like those used by the National Assessment for Educational Progress or the SAT (*see Figure 2 on page 20*). The benefit of this approach is that it can provide detailed feedback to students without compromising the summative design requirements. The drawback is that two students who both score, say, 500 on the reading portion will still receive the same feedback even though they may have answered the same set of questions differently.

A second approach to providing more feedback for instructional use is a suite of psychometric models referred to as diagnostic classification models (Rupp, Templin, & Henson 2010). In general, such models provide feedback based on the response patterns of individual students, and they provide an analysis of student strengths and weaknesses on specific, fine-grained skills. Although the feedback is rich, it is not always reliable given that most large-scale assessments are not designed to support this type of

FIGURE 2
EXAMPLE OF SCALE ANCHORING: THE SAT



Source: Adapted from the College Board website; accessed September 3, 2011

detailed feedback. However, some would argue that this type of information is still useful for instructional purposes. Needless to say, both approaches to providing more instructionally relevant score reports would be greatly improved if they were based on an assessment that was designed with student cognition in mind, such as evidence-centered design.

Improving the type of instructionally useful feedback provided by large-scale assessments is in keeping with the goals of student-centered assessment: improved feedback would help various stakeholders, including the student, connect the results of an assessment to his or her learning goals. One can imagine how this connection would help motivate and empower students.

THE ROLE OF CONTEXT

One challenge for large-scale assessment that needs further attention is the role of context. Allowing students to learn and build deep conceptual knowledge in real-life contexts is a key to developing the 21st-century skills required for college and career success. However, assessing the knowledge and skills that are important regardless of context is challenging. For example, suppose two high school biology teachers are conducting a unit on the various ways in which DNA is transferred to the next generation. One teacher is in Chicago, so she uses quagga mussels as the context for her students' project. They collect samples from Lake Michigan, conduct experiments, and record their observations. The other teacher is in Glenwood Springs, Colorado, so he uses a native plant that grows along the Colorado River as the context for the unit. How can both sets of students be assessed on the foundational knowledge and skills without putting either set at an advantage or disadvantage?

The developers of large-scale assessments routinely face this challenge. The role of context is not yet resolved, but progress can be made with more advances in assessment technology (e.g., simulation-based computerized assessments) and more sophisticated psychometric and assessment design innovations (e.g., the use of task models that can produce "families" of items that measure the same target but have different contexts) (Luecht 2002). Solving the problem of context in large-scale assessment would be a first step toward helping large-scale assessments be more individualized, which is the first defining characteristic of student-centered assessment systems and the one most challenging for large-scale assessment to attain.

CONCLUSION

We believe that a balanced system of formative, interim, and summative assessments can support student-centered assessment and learning. In general, the evidence suggests that classroom-based, formative assessments tend to be more easily individualized, focused on learning and growth, motivating, and amenable to actively engaging students in the regulation of their own learning, while interim and summative tests tend to be more useful to a wide audience of stakeholders; hence the need for each form of assessment.

Yet even an exquisitely balanced assessment system would present challenges. For one, the breadth and depth of the data about learning provided by modern measurement systems is awe-inspiring. Never before have students, teachers, parents, school districts, states, and the nation had access to such large quantities of high-quality information. But the sheer quantity of assessment data threatens to overwhelm us all. Even as we design new assessment processes, we must work to ensure they are useful to and used by the appropriate audiences. Students must learn how to take advantage of feedback to improve their work, deepen their understandings, and regulate their own learning. Teachers must learn how to individualize instruction and assessments and to make adjustments to instruction based on assessment results. Schools and districts must learn how to combine formative, interim, and summative results and interpret them in meaningful ways. And policymakers must learn to create and use balanced assessment systems that inform but do not overburden or overwhelm those they are designed to assist.

If that were not enough, we must also continually assess the assessments. From the perspective of evidence-centered design, we should be articulating our claims and goals for our assessment system, such as "students are revising their work based on individualized feedback," describing the observable evidence for those claims, and designing or identifying sources of that evidence. That is the job of researchers who, in collaboration with educators, can help ensure that recent advances in assessment are as student-centered as possible.

ENDNOTES

[1] See series paper: http://www.studentsatthecenter.org/papers/motivation-engagement-and-student-voice

[2] See series paper: http://www.studentsatthecenter.org/papers/teachers-work

[3] See: http://newtechhigh.org/?page_id=969

[4] See: http://www.wix.com/lindseygiaccio/portfolio

[5] See: http://newtechhigh.org/sean/

[6] See series paper: http://www.studentsatthecenter.org/papers/personalization-schools

[7] For a video of an exhibition, see: http://www.edutopia.org/urban-academy

[8] See series paper: http://www.studentsatthecenter.org/papers/motivation-engagement-and-student-voice

[9] See Huff & Goodman 2007 for an extended discussion of cognitively diagnostic assessment.

REFERENCES

Allal, L. 2010. "Assessment and the Regulation of Learning." In E.B.P. Peterson, ed. *International Encyclopedia of Education, Vol. 3*. Oxford, UK: Elsevier.

Alonzo, A.C., Neidorf, T., & Anderson, C.W. Forthcoming. "Using Learning Progressions to Inform Large-Scale Assessment." In A.C. Alonzo & A.W. Gotwals, eds. *Learning Progressions in Science*. The Netherlands: Sense Publishers.

Andrade, H.L. 2010. "Students as the Definitive Source of Formative Assessment: Academic Self-assessment and the Self-regulation of Learning." In H.L. Andrade & G.J. Cizek, eds. *Handbook of Formative Assessment*. New York, NY: Routledge.

Andrade, H.L. & Cizek, G.J., eds. 2010. *Handbook of Formative Assessment*. New York, NY: Routledge.

Andrade, H.L. & Du, Y. 2007. "Student Responses to Criteria-referenced Self-assessment." *Assessment and Evaluation in Higher Education*. Vol. 32, No. 2.

Andrade, H.L. & Valtcheva, A. 2009. "Promoting Learning and Achievement through Self-assessment." *Theory Into Practice*. Vol. 48, No. 1.

Bambrick-Santoyo, P. 2008. "Data in the Driver's Seat." *Educational Leadership*, Vol. 65, No. 4.

Barrett, H.L. 2007. "Researching Electronic Portfolios and Learner Engagement: The Reflect Initiative." *Journal of Adolescent & Adult Literacy*. Vol. 50, No. 6.

Beaton, A. & Allen, N. 1992. "Chapter 6: Interpreting Scales Through Scale Anchoring." *Journal of Educational and Behavioral Statistics*. Vol. 17, No. 2.

Belgrad, S., Burke, K., & Fogarty, R. 2008. *The Portfolio Connection: Student Work Linked to Standards, 3rd ed*. Thousand Oaks, CA: Corwin Press.

Bloom, B. 1984. "The Search for Methods of Group Instruction as Effective as One-to-one Tutoring." *Educational Leadership*. Vol. 41, No. 8.

Briggs, D.C., Alonzo, A.C., Schwab, C., & Wilson, M. 2006. "Diagnostic Assessment with Ordered Multiple Choice Items." *Educational Assessment*. Vol. 11, No. 1.

Brookhart, S. 2008. "Portfolio Assessment." In T.L. Good, ed. *21st Century Education: A Reference Handbook, Vol. 1*. Thousand Oaks, CA: SAGE.

Butler, R. & Nisan, M. 1986. "Effects of No Feedback, Task-related Comments, and Grades on Intrinsic Motivation and Performance. *Journal of Educational Psychology*. Vol. 78, No. 3.

Chang, C. & Tseng, K. 2009. "Use and Performances of Web-based Portfolio Assessment." *British Journal of Educational Technology*. Vol. 40, No. 2.

Cimpian, A., Arce, H.M., Markman, E.M., & Dweck, C.S. 2007. "Subtle Linguistic Cues Affect Children's Motivation. *Psychological Science*. Vol. 18, No. 4.

Cizek, G.J. 2010. "An Introduction to Formative Assessment: History, Characteristics, and Challenges." In H.L. Andrade & G.J. Cizek, eds. *Handbook of Formative Assessment*. New York, NY: Routledge.

Coalition of Essential Schools. 2006. *Measuring Up: Demonstrating the Effectiveness of the Coalition of Essential Schools*. Oakland, CA: Author.

Corcoran, T., Mosher, F.A., & Rogat, A. 2009. *Learning Progressions in Science: An Evidence-based Approach to Reform*. New York, NY: Center on Continuous Instructional Improvement, Teachers College, Columbia University.

Davidson, J. 2009. "Exhibitions: Connecting Classroom Assessment with Culminating Demonstrations of Mastery." *Theory Into Practice*. Vol. 48, No. 1.

Davidson, J. & Feldman, J. 2010. "Formative Assessment Applications of Culminating Demonstrations of Mastery. In H.L. Andrade & G.J. Cizek, eds. *Handbook of Formative Assessment*. New York, NY: Routledge.

Dembo, M. & Seli, H. 2008. *Motivation and Learning Strategies for College Success, 3rd ed*. New York, NY: Lawrence Erlbaum.

Duffrin, N., Dawes, W., Hanson, D., Miyazaki, J., & Wolfskill, T. 1998. "Transforming Large Introductory Classes into Active Learning Environment." *Journal of Educational Technology Systems*. Vol. 27.

Eisner, E. 1999. "The Uses and Limits of Performance Assessment." *Phi Delta Kappan*. Vol. 80, No. 9.

Evans, K.A. 2001. "Rethinking Self-assessment as a Tool for Response." *Teaching English in the Two-Year College*. Vol. 28.

Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., & Dooley, S. 2005. "Summary Street: Computer Support for Comprehension and Writing." *Journal of Educational Computing Research*. Vol. 33.

Goodrich, H. 1996. *Student Self-assessment: At the Intersection of Metacognition and Authentic Assessment*. Unpublished doctoral dissertation. Cambridge, MA: Harvard University.

Guskey, T.R. 2010. "Formative Assessment: The Contributions of Benjamin S. Bloom." In H.L. Andrade & G.J. Cizek, eds. *Handbook of Formative Assessment*. New York, NY: Routledge.

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. 2009. "From Evidence to Action: A Seamless Process in Formative Assessment?" *Educational Measurement Issues and Practice*. Vol. 28, No. 3.

Herman, J. & Winters, L. 1994. "Portfolio Research: A Slim Collection." *Educational Leadership*. Vol. 52, No. 2.

Huff, K. & Goodman, D. 2007. "The Demand for Cognitive Diagnostic Assessment." In J. Leighton & M. Gierl, eds. *Cognitive Diagnostic Assessment for Education*. Cambridge, UK: Cambridge University Press.

Huff, K. & Plake, B. 2010. "Innovations in Setting Performance Standards for K-12 Test-Based Accountability." *Measurement*. Vol. 8, No. 2-3.

Huff, K., Steinberg, L., & Matts, T. 2010. "The Promises and Challenges of Implementing Evidence-Centered Design in Large-Scale Assessment." *Applied Measurement in Education*, Vol. 23, No. 4.

Kulik, C.C., Kulik J.A., & Bangert-Drowns, R.L. 1990. "Effectiveness of Mastery Learning Programs: A Meta-analysis." *Review of Educational Research*. Vol. 60, No. 2.

Landauer, T., Lochbaum, K., & Dooley, S. 2009. "A New Formative Assesment Technology for Reading and Writing." *Theory Into Practice*. Vol. 48, No. 1.

Lea, S., Stephenson, D., & Troy, J. 2003. "Higher Education Students' Attitudes to Student-centred Learning: Beyond 'Educational Bulimia.'" *Studies in Higher Education*. Vol. 28, No. 3.

Lewbel, S.R. & Hibbard, K.M. 2001. "Are Standards and True Learning Compatible?" *Principle Leadership (High School Ed.)*. Vol. 1, No. 5.

Lipnevich, A. & Smith, J. 2008. *Response to Assessment Feedback: The Effects of Grades, Praise, and Source of Information*. Research report RR-08-30. Princeton, NJ: Educational Testing Service.

Luecht, R. 2002. "From Design to Delivery: Engineering the Mass Production of Complex Performance Assessments." Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

MacDonald, B. & Boud, D. 2003. "The Impact of Self-assessment on Achievement: The Effects of Self-assessment Training on Performance in External Examinations." *Assessment in Education*. Vol. 63.

McMillan, J.H. 2011. *Classroom Assessment: Principles and Practices for Effective Standards-Based Instruction*. New York, NY: Pearson.

Mislevy, R., Steinberg, L., & Almond, R. 2002. "Design and Analysis in Task-based Language Assessment." *Language Testing*. Vol. 19, No. 4.

National Research Council. 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Nichols, P. 1993. *A Framework for Developing Assessments that Aid Instructional Decisions*. ACT Research Report 93-1. Iowa City, IA: American College Testing.

Nichols, P. 1994. "A Framework for Developing Cognitively Diagnostic Assessments." *Review of Educational Research*. Vol. 64.

Pellegrino, J. 2002. "Understanding How Students Learn and Inferring What They Know: Implications for the Design of Curriculum, Instruction, and Assessment." In M.J. Smith, ed. *NSF K-12 Mathematics and Science Curriculum and Implementation Centers Conference Proceedings*. Washington, DC: National Science Foundation and American Geological Institute.

Pellegrino, J., Baxter, G., & Glaser, R. 1999. "Addressing the 'Two Disciplines' Problem: Linking Theories of Cognition and Learning With Instructional and Assessment Practice." *Review of Research in Education*. Vol. 24.

Perie, M., Marion, S., & Gong, B. 2009. "Moving Toward a Comprehensive Assessment System: A Framework for Considering Interim Assessments." *Educational Measurement*. Vol. 28, No. 3.

Perkins, D. 2003. *King Arthur's Round Table: How Collaborative Conversations Create Smart Organizations*. Hoboken, NJ: John Wiley & Sons.

Pintrich, P. 2000. "The Role of Goal Orientation in Self-regulated Learning." In M. Boekaerts, P. Pintrich, & M. Zeidner, eds. *Handbook of Self-Regulation*. San Diego, CA: Academic Press.

Popham, J. 2008. *Transformative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Resnick, D. & Resnick, L. 1985. "Standards, Curriculum, and Performance: A Historical and Comparative Perspective." *Educational Researcher*. Vol. 14, No. 4.

Ross, J.A., Hogaboam-Gray, A., & Rolheiser, C. 2002. "Student Self-evaluation in Grade 5-6 Mathematics Effects on Problem-solving Achievement." *Educational Assessment*. Vol. 8.

Rupp, A., Templin, J., & Henson, R. 2010. *Diagnostic Measurement: Theory, Methods, and Applications*. New York, NY: Guilford Press.

Russell, M.K. 2010. "Technology-aided Formative Assessment of Learning: New Developments and Applications." In H.L. Andrade & G.J. Cizek, eds. *Handbook of Formative Assessment*. New York, NY: Routledge.

Schneider, M.C. & Randel, B. 2010. "Research on Characteristics of Effective Professional Development Programs for Enhancing Educators' Skills in Formative Assessment." In H.L. Andrade & G.J. Cizek, eds. *Handbook of Formative Assessment*. New York, NY: Routledge.

Schunk, D. 2003. "Self-efficacy for Reading and Writing: Influence of Modeling, Goal-setting, and Self-evaluation." *Reading & Writing Quarterly*. Vol. 19, No. 2.

Seidel, S., Walters, J., Kirby, E., Olff, N., Powell, K., & Veenema, S. 1997. *Portfolio Practices: Thinking through the Assessment of Children's Work*. Washington, DC: NEA Publishing Library.

Shute, V. 2008. "Focus on Formative Feedback." *Review of Educational Research*. Vol. 78, No. 1.

Shute, V., Hansen, E., & Almond, R. 2008. "You Can't Fatten a Hog by Weighing It – Or Can You? Evaluating an Assessment for Learning System Called ACED." *International Journal of Artificial Intelligence in Education*. Vol. 18.

Snow, R. & Lohman, D. 1989. "Implications of Cognitive Psychology for Educational Measurement." In R. Linn, ed. *Educational Measurement, 3rd ed*. New York, NY: Macmillan.

Steinberg, L., Mislevy, R., Almond, R., Baird, A., Cahallan, C., DiBello, L., et al. 2003. *Introduction to the Biomass Project: An Illustration of Evidence-Centered Assessment Design and Delivery Capability*. CRESST Technical Report 609. Los Angeles, CA: Center for the Study of Evaluation, CRESST, UCLA.

Topping, K.L. 2010. "Peers as a Source of Formative Assessment." In H.L. Andrade & G.J. Cizek, eds. *Handbook of Formative Assessment*. New York, NY: Routledge.

VanderVeen, A., Huff, K., Gierl, M., McNamara, D., Louwerse, M., & Graesser, A. 2007. "Developing and Validating Instructionally Relevant Reading Competency Profiles Measured by the Critical Reading Section of the SAT." In D.S. McNamara, ed. *Reading Comprehension Strategies: Theory, Interventions, and Technologies*. Mahwah, NJ: Erlbaum.

Wade-Stein, D. & Kintsch, E. 2004. "Summary Street: Interactive Computer Support for Writing." *Cognition and Instruction*. Vol. 22.

Wylie, C. & Wiliam, D. 2007. "Analyzing Diagnostic Items: What Makes a Student Response Interpretable?" Paper presented at the annual meeting of the American Educational Research Association and the National Council on Measurement in Education.

Yang, J.C., Ko, H.W., & Chung, I.L. 2005. "Web-based Interactive Writing Environment: Development and Evaluation." *Educational Technology & Society*. Vol. 8, No. 2.

Zimmerman, B. & Schunk, D., eds. 2011. *Handbook of Self-regulation of Learning and Performances*. New York, NY: Routledge.

JOBS FOR THE FUTURE

TEL 617.728.4446 FAX 617.728.4857 info@jff.org

88 Broad Street, 8th Floor, Boston, MA 02110
122 C Street, NW, Suite 650A, Washington, DC 20001

WWW.JFF.ORG

nellie mae EDUCATION FOUNDATION

TEL 781.348.4200 FAX 781.348.4299

1250 Hancock Street, Suite 205N, Quincy, MA 02169
WWW.NMEFOUNDATION.ORG