

Principles for a Comprehensive Assessment System

Assessment has long been at the center of education policy debates, and for good reason. The goal of schooling is to maximize student learning, and assessments provide a picture of what students know and are able to do. Assessments also have a strong influence on what goes on in classrooms.

The United States is now poised to make the most dramatic advance in assessment in decades. A state-led effort to develop common core standards in literacy and mathematics is defining what it means to be ready for colleges and careers, and this effort will invariably heighten the demand for assessments that measure a broader range of knowledge and skills and open the door for common assessment components across states. At the same time, the U.S. Department of Education is providing \$350 million for consortia of states to develop new assessments that measure the common core standards. And a reauthorization of the Elementary and Secondary Education Act (ESEA) will set guidelines for assessments and their use for years to come.

This opportunity could not be more timely. There is widespread concern that the most prominent assessments currently being used in the United States are inadequate and may have a significantly negative impact on student learning. These tests measure a limited range of knowledge and skills and thus provide an incomplete—and often inaccurate—picture of what students know and can do. They are considered a distraction from instruction and fuel the perception that students are being over-tested. They also provide an inaccurate picture of school performance and make it difficult to direct resources to the schools that need them the most. And in many cases, when high stakes are attached to the results, they create incentives for schools to limit learning opportunities for children—because of an undue focus on the tests.

In large part, this situation results from the fact that the primary means of assessment being used in the nation's schools—state tests administered at the end of the school year—are not up to the job. For many reasons, these tests cannot measure the full range of competencies students are expected to demonstrate and provide at best a blurry snapshot of student and school performance. Yet they are expected to inform students and parents about academic progress, teachers about what to do to improve instruction, community members about school success, and states and the nation about whether students and schools are meeting goals. These expectations, all important, are more than any assessment—even the most sophisticated—can bear.

This brief suggests the principles upon which the federal government and states should base their work in fashioning new assessments. Recognizing that no single test can fulfill all the needs for information by all stakeholders, it suggests the need for a comprehensive system of assessments. Most importantly, the brief argues that this system needs to be coherent and cohesive, aligned to standards for college and career readiness.

Of course, an assessment system alone cannot ensure that all students learn what they need to know to succeed after graduating high school. Teachers need curriculum and instructional tools to teach effectively, as well as the ability to use assessment information skillfully. Yet without strong assessments, any effort to raise standards for students will likely fail. Students, parents, teachers, community members, and public officials all need valid and reliable information to strengthen teaching and learning. To ensure that students know what they need to know, we have to know what they know now.

The Purposes of Assessment

Since Horace Mann administered tests of student performance in Massachusetts in 1840, assessments have played multiple roles in American education. As Mann did, states and districts continue to administer tests to inform the public about the state of student achievement and to hold schools accountable for meeting achievement goals. Tests are also used to inform parents about individual students' progress, and to certify whether students have met standards in order to advance to the next grade or graduate from high school. Teachers regularly use tests to determine if students understand the concepts they are trying to teach and if additional instruction is needed. Increasingly, school systems are looking to use tests to gauge teachers' abilities to improve student performance.

These purposes are important, but a single test cannot serve all of them. For example, to determine whether schools are making progress would require a broad survey of knowledge and skills taught in a particular year. But such tests are ill-suited for teachers to monitor individual students' learning. The broad survey tests are intended to be “summative”—administered at the end of the year, to see whether students have learned what they were expected to have been taught. But the results come back too late for teachers to use them during the course of the year. Moreover, the broad survey tests typically include only a few items on each topic, too few for teachers to get a good understanding of where students need help and what might be tripping them up, much less what strategies they might employ to ensure that all students understand the concepts they need to know.

The different tests include different types of test items as well. End-of-year tests are generally administered in a single setting; that is, students sit for an hour or two and answer questions posed to them in a booklet handed to them. But this type of test excludes long-term projects and papers that might take days or weeks to complete. Such projects are important for knowing whether students can apply their knowledge to solve complex problems, communicate their understanding, and reflect on their performance, among other skills, yet such competencies are seldom measured by on-demand tests.

Unfortunately, the end-of-year tests that are useful for accountability purposes have played an outsized role in American schools in recent years. But even if they were limited to their most appropriate purpose, the end-of-year tests that most states and districts employ are inadequate and outmoded. They too often focus on relatively low-level knowledge and skills and fail to gauge whether students are meeting all the standards states and districts expect them to meet. There is little evidence that they are “instructionally sensitive”—that is, if they detect whether improvements in teaching improve student performance—and the scores continue to reflect students' backgrounds more than their school experiences. They fail to take advantage of advances in technology that enable much more sophisticated examinations of student learning.

A new assessment system would redesign the summative tests used for accountability purposes and embed them in a comprehensive and coherent system in which curriculum, instruction, and assessment are intertwined. Think of the system as a wheel—at the hub are the common core standards, and the



spokes include summative assessments, formative assessments, curriculum tasks, instructional tools, and professional development. In such a system, assessments are not separate and apart from classroom instruction, they are integral to it. All forms of assessment provide an ongoing information loop to teachers, school leaders, parents, policymakers, and the public.

To envision how such a system might shift school practice, consider what has happened in the retail industry. In the past, retail stores would close their doors for a day each year to take inventory. Now, thanks to the accurate and instantaneous information bar codes allow, retailers can keep track of their inventory in real time, 365 days a year. Likewise, a comprehensive assessment system could provide continuous, coherent, and high-quality information on student performance that teachers, school leaders, and district and state administrators could use to improve teaching and learning.

Organizing Principles

What would such a system look like? While it could take a number of forms, an ideal system would be organized around the following principles:

Coherence

The system aligns curriculum, instruction, and assessment around the key learning goals spelled out in the standards for college and career readiness.

The notion of “alignment” is critical to standards-based reform. The Elementary and Secondary Education Act has since 1994 required states to ensure that assessments are aligned to challenging standards. In practice, though, that has meant that states and test makers line up test items against standards documents and show that they match. More comprehensive studies have shown that this level of reported alignment is misleading, and that tests measure only a narrow range of the expectations for student learning described in the standards. A study of standards and assessments in nine states, for example, found that only 30 percent of mathematics assessment items in those states matched the content and cognitive demand of the mathematics standards’ expectations in fourth grade; only 26 percent matched the standards at eighth grade.¹ And, as previous research has shown, the study found that the test items tended to focus on relatively low-level content and skills.

This misalignment has serious consequences for instruction and learning. Researchers have consistently found that teachers place more emphasis on the tests than on the standards to guide instruction.² As a result, classroom instruction tends to focus on relatively low-level content and skills, rather than the important learning goals expressed by the standards. And the assessment results provide misleading information about the extent to which students have attained the learning goals the standards expect them to reach.

A coherent system would be organized around a limited number of “big ideas” in the standards, rather than attempt to align all pieces to every standard. The system would ensure that all components, at all levels, are truly aligned to those big ideas. This practice is common in other countries, such as Australia, Canada, England, and Singapore. In those countries, education officials use standards to develop curriculum guides, or syllabi, and professional development, along with external and classroom-based assessments. The set of assessments measure the full range of the core standards, and reinforce instruction that aims at higher levels of cognitive complexity as well as basic skills and knowledge.



Queensland, Australia

In Queensland, Australia, there have been no external examinations since 1971. Instead, the Queensland Studies Authority (QSA) has implemented a unique system of externally moderated school-based assessment for secondary schools. The QSA supports the assessment program throughout the state from Prep to Year 12. Its comprehensive system, Learning P–12, is an integrated approach to teaching, learning, assessment, and reporting. All assessment practices in the state are standards-based. The QSA begins with the national curriculum guidelines and then develops Essential Learnings that identify what students should know, understand, and be able to do in each content area. These key concepts are arranged along a developmental continuum. With guidance from the national curriculum and the state syllabi, districts design their own systems of curriculum, instruction, and assessment for each two-year upper-secondary course. These systems meet the needs of individual students, are responsive to the local context, and enable students to master the Essential Learnings for each content area. Using the QSA syllabus as a foundation, the district develops a comprehensive system that includes a range of assessment techniques to measure student achievement, including group work, oral presentations, projects, portfolios, and examinations. This process allows districts to exercise judgment about what is best for their students. In addition, the assessment process is beneficial for students because they are judged on their performance over the course of two years rather than on the basis of one high-stakes examination.

Teachers are central to the assessment system in Queensland. They are trained in the Essential Learnings, the formative assessment process, and the use of evidence from assessments to make informed and comparable judgments. They further develop their competence in assessment by engaging in professional conversations about student work. Teachers work collaboratively to analyze samples of student work, compare them to the Essential Learnings, and make judgments about student learning. Through this process, teachers build their knowledge and expertise about assessment.

In order to ensure that the results of assessments are consistent, defensible, and comparable from school to school across the state, Queensland uses a system of moderation for subjects that count toward university entrance. There are several phases of the moderation system. First, the syllabus for the two-year course is approved by the state. Then, the school district develops a work program to describe how the district will implement the syllabus, given its local context, and support students in achieving mastery of the Essential Learnings. The moderation process then provides a mechanism for the state to ensure that the syllabus criteria and standards are consistently implemented statewide. After the first year of the course, schools select a sample of portfolios of student work from Year 11 students and present them to district review panels of expert practitioners. These panels provide schools with feedback about implementation of the course, the effectiveness of assessment instruments, evidence of student achievement, and whether teachers are making comparable judgments about the quality of student performance. After the second year of the course, schools select a sample of portfolios from students in Year 12. The district review panels then verify that the assessments of the quality of student learning are comparable across the district to ensure consistency in the <http://unabridged.merriam-webster.com/interpretation> and application of the performance standards. State panels also review a random sample of student portfolios to ensure that judgments about student progress are comparable across the state of Queensland. Panel members follow up with districts in cases in which they identify discrepancies. Research evidence indicates that Queensland's assessment system produces reliable and comparable assessment of student achievement across the state.

Comprehensiveness

The system consists of a toolbox of assessments that meet a variety of different purposes and that provide various users with information they need to make decisions.

While the end-of-year tests that have loomed large in American education provide important information for accountability purposes, they are less useful for teachers, who need fine-grained information about student strengths and weaknesses, or for school leaders, who need accurate information about teacher effectiveness. A better-stocked toolbox would help serve all of these functions and more.

For example, such a system would include formative assessments that show teachers whether students truly understand the content or where they are struggling, along with tools to suggest steps they could take to help students overcome their difficulties. The system would also include measures that provide data to inform school leaders about teachers' effectiveness at improving student learning over the course of a year and that suggest professional development strategies. And it would include classroom



assessments that provide students and parents with an ongoing record of student progress, along with indicators to show areas where improvement is needed.

The system would also include a mix of measures to show whether students are attaining all of the important knowledge and skills they are expected to learn. In Rhode Island, for example, students must complete a project or submit a portfolio, in addition to completing their coursework and passing tests, in order to graduate from high school. These projects offer opportunities for students to investigate a topic with the help of a community mentor, write a lengthy research paper, and make an oral presentation to teachers and community members.

Importantly, a comprehensive system would also recognize the unique needs of high schools. In the upper grades, high school students often branch off into various pathways that reflect their interests. While such pathways should all lead to preparation for postsecondary success, they require different assessments, and a comprehensive system would accommodate those differences.

Accuracy and Credibility

The information from assessments supports valid inferences about student progress toward college and career readiness, as well as actionable information for multiple users.

To serve as credible measures of standards for college and career readiness, assessment systems must not only indicate whether high school graduates are prepared for postsecondary success; they must also show whether younger students are on a path that will lead to future success. In order to accomplish this goal, the assessment system should be grounded in a clear, evidence-based idea of learning and development that leads to college and career readiness.

In recent years, cognitive scientists have developed such paths, known as “learning progressions,” in several subject areas, and some assessment programs have used them as the basis for measures to determine student progress.³ For example, the Developmental Assessment program, developed by the Australian Council for Educational Research and used in several states in that country, is based on a set of “progress maps” that describe knowledge and skills in the sequence in which they develop. Such maps provide a picture, from beginning learning to expertise, and enable students, parents, teachers, and the public to see student progress over time. In that way, the assessments based on the progress maps can also be used to measure growth in student performance.

Accurate and credible assessments also measure—and support—good instruction. Currently, many of the tests used for accountability purposes fall short on this score. Students who come into class with high levels of background knowledge can perform well, regardless of what their teacher did.⁴ And as a result, they do little to create incentives for teachers to improve their practice.

By collecting evidence on whether assessments are “instructionally sensitive”—that they detect the effects of good instruction—states and state consortia can ensure that assessment systems accurately and credibly measure performance and create incentives for good instruction. At the same time, such measures would be seen as fair indicators to inform judgments about teacher effectiveness.

Fairness

The assessments enable *all* students to demonstrate what they know and are able to do.



States have made great strides in including students with disabilities and English language learners in assessment programs over the past eight years. One study of eight states, for example, found that the percentage of students with disabilities who took high school reading tests increased from 87 percent to 94 percent between 2001 and 2005. The performance of students with disabilities also rose substantially over that period.⁵

But states and state consortia can do more to ensure that assessments are fair by designing them from the outset to allow participation by all students. This means, among other things, making sure that the language and images used are clear to all and are relevant to the concepts that are tested, enabling the use of responses other than writing, and providing accommodations that facilitate participation.

Similarly, assessment systems should also ensure fairness by allowing students at all ranges on the achievement continuum to demonstrate what they know and can do. For accountability purposes, however, the goal should be to determine whether students are on track toward college and career readiness.

Fairness also implies transparency. Students should know what the expectations are, and assessments should measure what they are expected to learn.

Design Principles

These organizing principles for a comprehensive, coherent assessment system imply a set of design criteria. While states and state consortia designing new assessment systems might start from different places, depending on the type of assessment system they currently have in place, the assessments they build can all meet these criteria.

Matching Learning Goals

The types of assessment tasks should match the learning goals students are expected to demonstrate. Some of these goals might lend themselves to multiple-choice items, but others will likely require extended tasks and performances. The assessment system should include curriculum-based components to accommodate learning goals that cannot be measured by on-demand assessments.

Assessment systems in other countries include tasks students complete during the year, as well as end-of-year tests, to measure student performance. For example, in Alberta, Canada, only half of a student's score on the twelfth-grade examination, which is used as a requirement for graduation, is based on an end-of-year examination. The other half is derived from work conducted throughout the year, including projects, portfolios, oral presentations, and exhibitions—all scored by teachers. Teachers also score the open-ended components of the examination, which count for 50 percent of the exam score.

The experience in Alberta shows that teachers can accurately and reliably evaluate student work. Teachers are trained extensively in the standards for student performance, and their scores are moderated—that is, they are audited by independent reviewers to ensure consistency. And teacher involvement in scoring has improved their evaluation of student performance throughout the year, according to Jim Dueck, assistant deputy minister for accountability and reporting for the Alberta Ministry of Education. In fact, he states, Alberta teachers are the least likely of any teachers in Canada to use nonacademic factors in awarding marks to students.⁶



Alberta, Canada

In the province of Alberta in Canada, students are tested at four levels of the education system: in language arts and mathematics in grade three; in language arts, mathematics, science, and social studies in grades six and nine; and in core academic courses in grade twelve through the diploma examination system. These diploma examinations provide students with a credential indicating that they demonstrated mastery of the content in that subject area.

Students' final grades in diploma examination system courses are based on a fifty-fifty blend of the diploma examination score and school-based assessments. The diploma examination score itself is also a fifty-fifty blend of multiple-choice and open-ended items. The Ministry of Education provides teachers with detailed blueprints for the diploma examinations that clearly outline the alignment of the assessment with the curriculum, describe the items and tasks that will appear on the examination, and provide rubrics on which student responses will be judged.

The 50 percent of the student's grade for school-based assessments might include unit tests, quizzes, projects, portfolios, presentations, or exhibitions. Because some learning goals are better measured with performance tasks than with standardized tests, the format of the assessment is aligned appropriately with the curriculum standard that is being measured. In this way, students provide evidence, through a variety of assessments, that they have demonstrated mastery of all the curriculum standards associated with the course.

Teachers in Alberta are involved with all aspects of assessment. In the development stage, teachers work collaboratively to design test blueprints and scoring criteria, generate items, and review examinations to ensure fairness and alignment with the curriculum. Before scoring begins, teachers provide selections of student work that can be used as examples during training. After practicing with these examples and engaging in professional development to reach consensus on appropriate scores for different levels of quality work, teachers from across the province then score the examinations. To ensure consistency in examination scoring, multiple readers score them, and those scores are calibrated. Teachers also serve on advisory panels to provide state officials with recommendations for improvements of the assessment program for the following year. This involvement throughout the process helps teachers to build their knowledge about assessment and ensure that the end-of-course examinations are aligned with classroom instruction.

Clarity in Reporting

The design of the reports should be considered at the outset and should provide different audiences with actionable information that informs next steps to further advance student learning.

Performance reports are in many ways the most important part of assessment systems—they are the ways that students, parents, teachers, and policymakers learn what students know and can do. Yet too often they are an afterthought, developed after the assessment is designed. As a result, they are often too arcane to serve a useful function, and in some cases they do not accurately convey student learning. For example, efforts to overlay descriptions of the knowledge and skills students at particular performance levels can demonstrate are not always accurate.⁷

By mapping assessment tasks to the learning progressions, assessment systems can report student performance in ways that are both accurate and informative. One example of such a system is Australia's National Assessment Program: Literacy and Numeracy (NAPLAN). Under that program, parents receive reports on individual student performance that indicate the "band," or level of achievement, from 1 to 10, the student reached, along with a description of the knowledge and skills a student in that band can demonstrate (see Figure 1 on the next page).



Figure 1: Skills for Year 5 (Fifth Grade), 2009 National Assessment Program: Literacy and Numeracy (NAPLAN)

Band	Reading	Writing	Language Conventions	Numeracy
8	Analyses and interprets the content of complex texts. Identifies the writer's main message by connecting information across a persuasive text.	Writes a cohesive story that attempts to engage the reader. Makes deliberate and appropriate word choices to create an emotional response. Attempts to use features of recognisable text types such as fantasy or adventure stories, and devices such as humour or drama. Constructs most complex sentences correctly.	Identifies an error, then correctly spells words with difficult letter patterns (<u>ex</u> hausted, <u>exc</u> itement). Recognises and applies a variety of grammatical conventions. Demonstrates the correct use of punctuation in written English such as commas, speech marks and apostrophes.	Solves a word problem involving subtraction of decimals to two decimal places. Recognises different visual representations of a common fraction. Solves a complex problem involving multiplication, division and conversion between grams and kilograms. Calculates the area of rectangles given their side lengths. Visualises the number of painted faces of a composite 3D object.
7	Infers the motivation of a character in a more difficult story. Identifies how values are presented in a persuasive text. Understands how factual texts are structured and recognises the effect of slogans and the purpose of layout features such as subheadings.	Writes a story with a beginning and a detailed complication. Develops characters and setting through description or dialogue. Joins and orders ideas using connecting words and maintains clear meaning through the story. Correctly spells most common words and some difficult words, including words with less regular spelling patterns and silent letters.	Correctly spells some difficult words ending with 'tion' (illustr <u>ation</u>) and with long vowel sounds (a <u>u</u> thor). Identifies an error, then correctly spells some difficult words (precious) and words that contain silent letters (kn <u>i</u> tting). Recognises common grammatical conventions such as the correct tense of a verb phrase. Recognises the correct use of punctuation for direct and indirect speech.	Applies strategies such as multiplication or grouping to solve a word problem. Converts centimetres to metres. Calculates elapsed time in hours and minutes using digital am and pm times. Determines the total number of specified features of a common 3D object.
6	Makes connections between ideas that are implied and not clearly stated. Interprets figurative language to understand a character's feelings in a story. Makes inferences about a character's actions and feelings. Understands how texts are structured and recognises the purpose of parts of a text such as the title.	Organises a story into paragraphs that focus on one idea or a group of related ideas. Uses accurate words or groups of words when describing events and ideas. Punctuates nearly all sentences correctly with capitals, full stops, exclamation marks and question marks. Correctly uses more complex punctuation marks some of the time.	Correctly spells words with long vowel sounds (c <u>u</u> rved) and words ending in 'ous' (danger <u>ous</u>). Identifies an error, then correctly spells two-syllable words (often). Recognises common grammatical conventions such as the correct use of adjectives, verbs and adverbs. Recognises the correct use of punctuation such as commas in lists and possessive apostrophes.	Solves a problem involving addition and subtraction. Solves a problem involving division with a remainder. Divides whole objects into fractions. Continues a spatial pattern beyond the next term. Uses a balance to determine equivalence. Finds the width of a rectangle given the perimeter and length. Locates information in a timetable. Identifies a reflex angle. Interprets a map and follows directions to locate a position.
5	Connects and interprets ideas in a story. Identifies the meaning of a word in context and the main idea of a paragraph. Connects ideas and pictures in a longer illustrated factual text. Identifies to whom a pronoun refers in a sentence.	Structures a story to include a beginning, a complication and events that are related to each other. Includes enough supporting detail for the story to be easily understood by the reader, although the conclusion may be weak or simple. Correctly structures most simple and compound sentences and some complex sentences.	Correctly spells words containing unusual vowel patterns (th <u>ou</u> ght) and two- and three-syllable words ending with 'le' or 'er' (tit <u>le</u> , publis <u>h</u> er). Identifies an error, then correctly spells compound words (sometimes). Recognises common grammatical conventions such as the correct form of verbs. Recognises the correct use of punctuation such as question marks within direct speech.	Solves a money problem involving multiplication and rounding. Interprets data in a table to solve a subtraction problem. Selects a column graph to match information in a table. Compares volume by reading different scales. Identifies a square pyramid from its net and visualises the opposite faces on the net of a cube.
4	Makes an inference from clearly stated information and identifies the common message in two posters. Makes inferences about a character's actions in a story. Locates information in a longer illustrated factual text.	Writes a story in which characters or setting are briefly described. Correctly punctuates some sentences with both capital letters and full stops. May demonstrate correct use of capitals for names and some other punctuation. Correctly spells most common words.	Correctly spells common two-syllable words (bec <u>au</u> se). Identifies an error, then correctly spells common compound words (overn <u>igh</u> t). Recognises common grammatical conventions such as the correct use of adjectives and prepositions. Identifies the structure of a question. Recognises the correct use of punctuation such as capital letters for proper nouns and the placement of speech marks for direct speech.	Adds 2 two-digit numbers. Uses knowledge of number facts up to 10 x 10 to determine equal values. Continues a number pattern based on subtraction. Identifies a common 2D shape made when another shape is divided and identifies a prism displayed in an everyday context.
3	Connects clearly stated information from different parts of a text. Makes connections between the words and pictures in a story. Finds clearly stated information in a story and a poster.	Attempts to write a story containing a few related events, although these are usually not elaborated. Correctly orders the words in most simple sentences. May experiment with using complex sentences but with little success. Orders and joins ideas using a few connecting words but the links are not always clear or correct.	Spells common two-syllable words with regular spelling patterns. Recognises common grammatical conventions such as the correct use of adverbs and pronouns. Recognises some correct use of punctuation in written English.	Reads time and recognises the position of the hands on a clock at the half-hour. Identifies a specific date on a calendar. Identifies the most likely outcome of a simple chance event.

Source: NAPLAN, http://www.naplan.edu.au/home_page.html



Use of Technology

The use of technology can expand the types of assessment tasks that can be used, enhance the involvement of teachers, and make assessments more efficient.

As in many other sectors, technology has the potential to transform assessments and their use. Online testing, for example, can provide much better information about student abilities. Computer-adaptive testing, in which the items students respond to depend on their performance, can provide more accurate estimates of student performance, because it can pinpoint a student's level of achievement more efficiently. Oregon has pioneered the use of computer-adaptive tests for accountability purposes. In addition, computerized testing enables the use of simulations, animations, and other techniques that offer opportunities for students to engage in complex tasks that are unlikely in a paper-and-pencil setting.⁸

Computerized assessment also can save districts and states the considerable expense and logistical challenge of handling and distributing test materials. And online scoring makes it possible to receive instantaneous results.

Technology also facilitates teacher involvement in assessment development and scoring. Through online communities, large numbers of teachers can contribute to the development and review of assessment items and tasks. And states can engage teachers in scoring assessments without incurring the expense of bringing them together in a central location. Online scoring software can also serve as a check on the accuracy of human scoring.

Technology platforms can also make assessment results more accessible and useful. Results can be downloaded directly into spreadsheets that can allow school faculties to analyze data more easily and effectively. And results can be linked to instructional tools that provide guidance for teachers and principals about what to do when students are struggling on a particular topic.

Teacher Engagement

Involving teachers in the development, scoring, and use of assessments can strengthen assessment and improve instruction.

One reason many teachers see testing as something apart from the daily work of teaching and learning is because in many cases they have had little to do with the development and scoring of assessments. They receive materials from a testing company, hand them out to students, collect them and send them back, and then get a report some months later that might seem incomprehensible to them.

Involving teachers in the development and review of assessment tasks can help them see that the tasks students are expected to perform are worthy exercises. And they can connect these tasks to the learning goals students are expected to meet.

Involving teachers in scoring assessments, meanwhile, helps familiarize them with the expectations for performance. Teachers can understand what good work looks like, and why work that does not meet expectations falls short. If the scoring takes place in a face-to-face setting, teachers can compare notes on instructional strategies.



Alberta's assessment system has involved teachers extensively. Some 35,000 Alberta teachers have participated in the system in some capacity, according to Dueck, and a 1993 report on the system by the General Accounting Office (now the Government Accountability Office) concluded that the teacher involvement has been extremely valuable. The report states, "Widespread teacher involvement ... helps increase teachers' knowledge of curricula and instruction and aids in the development of tests that are compatible with good classroom instruction."⁹

To be sure, involving teachers in assessment development, review, and scoring requires extensive professional development. Teachers need to be able to understand the learning goals, how classroom and assessment tasks can enable students to demonstrate these goals, and examples of good performance. Teachers also need to know how to interpret assessment reports and strategies they might employ to improve performance. Few teacher education programs, however, offer instruction in assessment literacy, and professional development on the topic for practicing teachers is scarce.

Federal Policy Implications

Federal education policy over the past two decades has played a dominant role in shaping state assessment systems. The Improving America's Schools Act of 1994 required states to put in place tests at three grade levels that were aligned to state content standards, and in response many states revamped testing programs or established them for the first time. The No Child Left Behind Act of 2002 added additional requirements, mandating tests in reading and mathematics each year from grade three through eight, and once in high school; science tests were added in 2006. NCLB also required tests for students with disabilities and tests of English language development, and the law provided funds for the development of new assessment systems.

The pending reauthorization of the ESEA offers an opportunity for the federal government to reshape testing practice yet again. The Obama administration has signaled that it intends to shift the ESEA to focus on college and career readiness, which will require new assessments to measure student progress toward that goal. To help ensure the development of comprehensive, coherent assessment systems that support student learning toward standards for college and career readiness, the federal government should take the following actions:

- **Require that tests provide credible and valid information on student progress toward college and career readiness.**

The development of standards for college and career readiness provides an opportunity to shift the education system from a focus on an arbitrarily defined notion of "proficiency" to an emphasis on students' preparation for success after high school. This shift can only be accomplished if all state assessments are truly aligned to the same college- and career-ready standards and indicate the extent to which students are on a path toward achieving those standards. Results from every state's assessments should show whether students are prepared for postsecondary success, and preparation should mean the same thing whether students attend school in St. Petersburg or Seattle.

Assessments can meet these criteria in several ways. One way would be to develop a single assessment for all states that is aligned to the standards for college and career readiness. The measures and the reports on the measures would be identical.



Another way would be to ensure, through the peer review process, that the assessments used by each state (or consortia of states) are aligned to the big ideas in the standards for college and career readiness and indicate student progress toward those standards. To accomplish this, states or state consortia would have to design their assessments to measure both the content and cognitive complexity the standards indicate, and would have to design reports from the outset so they show performance against standards.

The reauthorization of the ESEA should also use progress toward college and career readiness as its measure of accountability for schools and districts.

- **Require that tests for accountability purposes measure the breadth of standards for college and career readiness, and authorize the inclusion of curriculum-based assessments that measure competencies that cannot be assessed effectively by end-of-year tests.**

While on-demand tests provide important information about student performance, they cannot assess all that students would be expected to know and be able to do to be prepared for colleges and careers—for example, the ability to conduct research and write an extended essay—and thus provide limited and potentially misleading information about students’ preparation for postsecondary success. The heavy emphasis on such tests also encourages some schools to focus on the kinds of skills that can be assessed on paper-and-pencil tests and to downplay the deeper learning experiences that might better prepare students for the future.

Through legislation and regulatory guidance, the federal government could encourage the development and use of assessments that incorporate measures of classroom work conducted throughout the year as well as end-of-year, on-demand assessments. Such assessments should be scored by teachers through a moderated scoring system, a practice that other countries and regions, such as Queensland, Australia, have used effectively. The experience in Queensland and elsewhere demonstrates that teachers can score assessments accurately and reliably, and that the involvement of teachers improves classroom instruction by providing them with a clear sense of the expectations for student performance.

- **Target professional development funds toward the improvement of teachers’ assessment knowledge and skills.**

In a comprehensive assessment system, teachers play a greater role in assessing student performance than they do now. Teachers need to know how to develop and review assessment tasks, how to score them accurately and reliably, how to develop and employ effective formative assessments to track student knowledge and skills over time, how to interpret assessment results, and how to modify instruction based on assessment results. The involvement of teachers in assessment programs in other countries suggests that these practices are valuable in improving instruction, but teachers need support to be able to perform these practices effectively. The professional development needs are particularly acute for high school teachers, who typically teach many more students than elementary teachers do.

The federal government could direct states to include plans for developing teachers’ assessment knowledge and skills as part of their plans for professional development funds under Title II. The Higher Education Act could also create incentives for higher education institutions to make assessment literacy an integral part of teacher-preparation programs.



- **Ensure that assessments fairly include virtually all students, including English language learners and students with disabilities, and that they provide accurate information about student knowledge and skills.**

Over the past eight years, states have made great strides in including students with special needs in assessment programs, and achievement by students who had often been underserved has improved. However, new assessment techniques and technologies, such as universal design for learning and computer-based assessments, could enable more students to participate in regular assessments.

The federal government could encourage states and state consortia to incorporate new technologies that expand participation in assessment. However, in no way should such changes lower expectations for students; *all* students should be expected to graduate from high school prepared for college and careers.

- **Maintain the National Assessment of Educational Progress (NAEP) as an independent monitor of state and district performance.**

The NAEP has long been an invaluable source of information on student academic progress, and the requirement that all states participate in the assessment has only added to its value. Now, the NAEP serves an important monitoring function to show whether state-reported levels of performance or progress over time are corroborated by a well-respected, independent assessment. The addition of the Trial Urban District Assessment (TUDA) has extended this monitoring function to urban districts as well.

The federal government should continue state-level NAEP to maintain this monitoring function and should make the TUDA a permanent program. The content of the assessment, however, will need to change to reflect the standards for college and career readiness.

- **Provide states with financial support for assessment operating costs.**

The U.S. Department of Education has spurred bold thinking by launching a grant program to provide \$350 million for state consortia to develop new assessments tied to the standards for college and career readiness. Assessment development is a costly and labor-intensive process—new items and tasks must be written and field-tested, for example—and states and state consortia might be reluctant to take on the challenge, particularly in the current budget climate, without the federal incentive. Pooling their resources as consortia, an effort made possible by common standards, can also save money in the development process.

Once the initial development work is completed, though, states will continue to face ongoing costs to operate the new assessments, and if they include tasks that require human scoring, they will likely be more expensive than current tests. The federal government could encourage the implementation of new assessments by helping to offset some of the operating costs of the assessments. However, the federal government should not pay the entire costs; states should look for ways to reduce costs, such as the use of technology.



- **Provide support for the development and implementation of formative assessments and instructional tools.**

Formative assessments that provide real-time feedback to students and teachers about student understanding are integral components of comprehensive systems that support teaching and learning. They indicate not only how much students know and can do, but also where they might be struggling and why. These assessments are integral to instruction, and they are most effective when accompanied by instructional tools that offer guidance to teachers about strategies they might use to enhance student understanding.

The federal government's main interest is holding schools accountable for student performance, but the government can provide financial support to states and districts and create incentives for them to support professional development to enable teachers to improve their formative assessment skills. The federal funds could also support research and development in the design and use of instructional tools that are linked to the standards for college and career readiness. The government could also support efforts by teacher-preparation programs and professional development to enable teachers to assess students accurately, interpret results, and use these tools effectively.

- **Invest in research on learning progressions.**

Learning progressions indicate the pathways students take toward expertise in a subject area. They are critical to the development of an assessment system that measures student progress toward college and career readiness. They also provide a way to measure growth, by showing how students move forward along the progressions.

To be valid, learning progressions must be based on evidence. At this point, there is research that shows progressions in learning in some areas, but not all. Additional research would be vital to support the development of assessments that yield valid information on student progress toward college and career readiness. The Institute for Education Sciences should make research to develop stronger, evidence-based, learning progressions in core subject areas a high priority.

- **Evaluate the implementation and impact of new assessment systems.**

New assessments based on standards for college and career readiness represent a bold departure in American education. States and state consortia need to know whether the systems are achieving their desired results or whether there are any unforeseen unintended consequences. The federal government should sponsor a careful evaluation of the new systems to determine how they are being implemented, what impact they are having on school practices and school system policies, and what changes, if any, might be warranted.

This brief was written by Robert Rothman, a senior fellow at the Alliance for Excellent Education. Learning Point Associates assisted with the research.



Endnotes

- ¹ A. Porter, M. Polikoff, and J. Smithson, *Is There a de Facto National Curriculum?: Evidence from State Standards*, paper prepared for the National Research Council Workshop on Assessing the Role of K–12 Academic Standards in States, available at http://www7.nationalacademies.org/cfe/Porter_Smithson%20State%20Standards%20Paper_Tables.pdf (accessed October 28, 2009).
- ² See, for example, Center on Education Policy, *How State and Federal Accountability Policies Have Influenced Curriculum and Instruction in Three States: Common Findings from Rhode Island, Illinois, and Washington* (Washington, DC: Center on Education Policy, 2009).
- ³ J. Pellegrino, N. Chudowsky, and R. Glaser, *Knowing What Students Know: The Science and Design of Educational Assessment* (Washington, DC: National Academies Press, 2001).
- ⁴ W. J. Popham, *Unlearned Lessons: Six Stumbling Blocks to Our Schools' Success* (Cambridge, MA: Harvard Education Press, 2009).
- ⁵ M. Thurlow et al., *Trends in the Participation and Performance of Students with Disabilities*, Technical Report 50 (Minneapolis, MN: University of Minnesota, National Center on Education Outcomes, 2008).
- ⁶ Testimony at the U.S. Department of Education's Race to the Top Assessment meeting, Boston, MA, November 12, 2009.
- ⁷ J. W. Pellegrino, L. R. Jones, and K. W. Mitchell, *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress* (Washington, DC: National Academies Press, 1999).
- ⁸ E. Quellmalz and J. W. Pellegrino, "Technology and Testing," *Science* 323 (2009): 75–79.
- ⁹ General Accounting Office, *The Canadian Experience with Standards, Examinations, and Assessments* (GAO-PEMD 93-11) (Washington, DC: U.S. General Accounting Office, April 1993), p. 3, available at <http://archive.gao.gov/d44t15/148938.pdf> (accessed February 5, 2010).

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

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

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

The framework, shown graphically here, encompasses seven policy areas that represent key leverage points in ensuring a comprehensive, systematic approach to improving secondary education. The framework also captures three guiding principles that apply to all of the policy areas. Although the appropriate federal role varies from one issue area to another, they are all critically important to reducing dropouts and increasing college and career readiness.

