

Student Assessment Systems Backgrounder

An essential aspect of any comprehensive assessment system is that it is guided by a clear articulation of the purpose of education, articulation of which drives the design of the system. NEA proposes the following purpose statement language for ESEA:

“The public education system is critical to democracy and its purpose, as reflected in this Act, is to maximize the achievement, skills, opportunities, and potential of all students by building upon their strengths and addressing their needs, and to ensure that all students are prepared to thrive in a democratic society and diverse, changing world as knowledgeable, creative, and engaged citizens and lifelong learners.”

While the goal of students being college and career ready may be a reasonable starting point, it is too narrow a goal for K-12 education overall. Countries that achieve significant results on international student assessments include such statements in their national documents on standards or curriculum. Without this broad, directive guidance, the components of assessment systems become overly diffuse or focused on small enabling skills and never seek to gauge the achievement of larger, essential education goals.

Fostering Complete Assessment Systems and a Whole Curriculum

The time is past when education reformers could focus only on administering single tests at the end of the year to determine and analyze student progress and school success. It is clear, looking at successful education programs in this country and abroad, that attention must be given to establishing a complete, balanced, and relevant assessment system. A complete assessment system should address broad education goals and include five major components that must all be addressed simultaneously:

- summative assessments;
- formative assessments;
- teacher capacity and development;
- effective data systems; and
- evaluation of effectiveness and proper use of assessment systems

It would be unwise to focus on summative assessment while leaving formative assessments for later attention. Indeed, the “state of the art” with respect to the development of formative assessments in the United States lags far behind its development of summative assessments. Other high-performing countries such as Singapore and New Zealand have placed an emphasis on formative assessments, and the achievement of their students attests to the wisdom of this decision. Likewise, improving assessments is useless without effective programs to improve teacher capacity and development around the use of assessment systems or data systems that can capture and organize assessment data in a way that improves instruction and student learning in a timely fashion. Finally, due to the proliferation of new assessment systems and implementation approaches, any complete assessment system requires an evaluation component at the national, state, local and site-based level to monitor the system’s effectiveness and to ensure that assessments are being used well and for intended purposes only.

The use of large-scale assessments in only reading and mathematics has resulted in a narrowing of the curriculum that ultimately denies a complete education in all disciplines to many students. This is especially true of students whose economic status does not give them access to out of school activities that might fill gaps, such as music or art education, who attend schools that are underfunded and unable to provide education in all disciplines, and who attend schools that are focused only on training students to do well on the two tests used for accountability. Therefore, complete assessment systems must address the whole curriculum, not only math and reading.

- *NEA proposes that all funding for education assessment be used to promote the development of complete assessment systems that include summative assessment, formative assessment, teacher capacity development, effective data systems, and evaluation of effectiveness and proper use of assessment systems.*
- *NEA proposes that ESEA provide funding that will enable states and/or districts to develop and use assessments for promoting and monitoring learning in disciplines beyond reading, math, and science. These assessments should support access to a complete education for all students.*

Addressing the Multiple Purposes of Assessment

A single assessment, regardless of how technically strong or broadly based, cannot serve multiple purposes that are as broad and disparate as “accountability” and “instructional planning.” If we have learned nothing else from the past 20 years of development of state-mandated assessments—or from several years of the spotlight on NCLB assessments—it is that such instruments are essentially instructionally inert. This is not a reflection on the low quality of the instrumentation, but of the fact that the assessments have been built specifically to serve accountability purposes. Instructionally useful instruments can be built; however, they cannot evolve from the classical assessments that we have been building for accountability purposes.

A complete system should incorporate the concept of assessment *of, for, and as* learning. This concept is explained and supported in assessment literature (Bennett & Gitomer, 2009). It is also embraced by several high-achievement countries such as Singapore, New Zealand, and Canada and is integrated into the assessment system description below.

Assessment *of* learning refers to summative assessments such as state accountability tests.

Assessment *for* learning is formative assessment that provides immediate information to students and teachers about whether students understand and can apply skills and knowledge. It includes tasks such as discussions, applications of skills, projects, interviews, and more.

Assessment *as* learning refers to the use of assessments to bring deeper understanding of education goals, concepts, skills, and the process of learning to both teachers and students. Discussion of student work by teachers and by students is an example of assessment *as* learning.

- *NEA proposes that Congress foster the development of assessment systems that serve multiple purposes, especially those that enhance and facilitate instruction and learning.*

Strengthening Assessment Tasks

Summative assessment, or assessment *of* learning, includes state accountability tests, interim assessments, and end of course tests. These assessments should consist mainly of rich, open-ended tasks that require application of skills and knowledge to solve problems, create projects, and think critically.

We should encourage efforts to develop assessment tasks that can be scored for learning in multiple disciplines such as math-science and social studies-writing. This can increase cost- and time-effectiveness of assessments as well as support the incorporation of real world tasks into assessments. The tasks can be administered through technological platforms when valid. However, there should be small projects that require students to use multiple media to demonstrate achievement. This type of task allows students to demonstrate learning in several possible ways and thus conforms to the principles of Universal Design for Learning (see section below).

- *NEA proposes that summative assessments include a variety of tasks such as performances, projects, and portfolios.*

Effective and Balanced Timing of Assessments

NEA strongly advocates that students be assessed for accountability by an outside instrument, such as a federally mandated state test, no more than once in grades 4–6, once in grades 7-9, and once in grades 10-12. For young children, there should be no such assessment before grade 3. It is possible to ascertain school effectiveness by testing students once while they are in that school. It is not necessary to test every student every year, since individual scores on such tests are too limited for judging individual progress or determining diagnostic needs. Individual progress can be determined only by multiple, in-depth measures such as combinations of summative assessment tasks administered throughout the school year. Teachers and schools require a more balanced assessment system that allows them to develop other measures of student learning as well as administer other types of assessments and tasks beyond large-scale, standardized assessments. Diagnostic assessments need to provide valid, reliable information based on opportunities to apply skills and knowledge multiple times in order to pinpoint the specific strengths and needs of individual students. These are most useful when directly tied to classroom instruction as well as standards. Summative assessments, at least large-scale assessments similar to those currently in use, are not appropriate for diagnosing individual students.

High school summative assessments should allow students to demonstrate their abilities to apply knowledge and skill in complex situations, such as performance tasks, projects and portfolios.

The NAEYC, IRA, and NCTE have a joint statement indicating that large-scale, paper and pencil tests are not appropriate for young students below grade 3. At these early levels, teachers should use formative and diagnostic assessments to track students' progress toward achieving standards.

Summative assessments should be criterion-referenced tests aligned with standards. Criterion-referenced assessments make sense when the assessment is to determine whether students are meeting or progressing toward meeting standards. Norm-referenced assessments, designed to spread students along a normal curve, are not appropriate for this purpose. The criterion-referenced test should, however, allow for students to demonstrate achievement of criteria in multiple ways in order to honor the guidelines of universal design for learning. These assessments should consist of complex tasks administered three or four times a year and reported as cumulative scores (Bennet, 2009). This design eliminates summative assessments that provide only a single snapshot of student achievement. It also allows for the use of more complex tasks, since the assessment is not administered in one sitting and therefore not subject to the time constraints of a one-time, end of year assessment or a two-time administration solely to mark growth. Funding for innovation in assessment should also include support for development and use of performance tasks and/or projects as summative assessment tools.

Summative assessments should be developed, piloted, and scored at least in part by teachers. Current technology makes this highly feasible. This involvement leads to a deeper understanding of expectations by teachers and provides knowledge based on classroom experience with students to the process of assessment development.

- *NEA proposes that states use large-scale, summative assessments for accountability only once in each of the grade spans 4-6, 7-9 and 10-12. Students in grade 3 could be assessed by tasks that can be commonly agreed on, but that are designed with administrative flexibility that acknowledges the developmental level of these students.*
- *NEA proposes that different summative assessment tasks be administered several times through the school year in order to avoid relying on a single snapshot of student learning and to allow time to employ complex tasks.*
- *NEA proposes that teachers be substantively involved in all aspects of the development and scoring of summative assessments.*

Improving Formative Assessments

Assessment systems should have a formative component or assessment *for learning*. The essential aspect of formative assessment is that it provides direct feedback to students as well as teachers and occurs at a point where additional instruction or learning activities can be identified and used to address learning weaknesses or next steps. Both teachers and students can use formative assessment data to plan learning. These assessments take place at the classroom level and should be generated by teachers, chosen from assessment resources, and adapted by teachers to use with specific students, chosen by students, or some combination of these. In countries such as Singapore, New Zealand, Britain, and the Netherlands a bank of formative assessment tasks is available to teachers via the Internet. Formative tasks can be paper and pencil, performances, demonstrations, and projects. In order to tap complex, higher order skills and content, they should be as rich as possible, allowing students to demonstrate and analyze their own learning related to standards and the content underpinning the large-scale assessments. They should include performance tasks and projects.

Formative assessments can also be quick checks that allow teachers and students to confirm achievement, note progress, and set next steps in learning. Teachers must be allowed total flexibility in determining which tasks to use, when to use them, and how to adapt them. Tasks should be aligned with standards and summative assessments but need to go beyond these, especially if those components are limited as they are currently. These assessments must take into account changing needs of individual students and support teacher response to those needs in timely manner (Black & Williams, 1998). Ultimately, the means of using these assessment tasks to predict performance on summative assessments can be developed.

Some assessments function as both summative and formative assessments. During a project, for example, a student should receive continuous feedback on progress and needs while the final project evaluation is summative.

- *NEA proposes that ESEA support the development and use of formative assessments by providing funding and by including the development and use of those assessments in requirements for school effectiveness.*

Applying Universal Design for Learning Principles

Universal Design for Learning (UDL) is the principle that specific needs of individuals based on language and learning differences need to be considered so students have fair and valid opportunities to demonstrate their learning. UDL anticipates special needs of students and creates curriculum, instruction, and assessments specifically designed to facilitate access. In the process, *all* students benefit. To appropriately assess students with disabilities and English language learners, educators should: (1) use the principles of UDL in developing assessments for all students to increase accessibility, i.e., multiple measures for demonstrating learning; (2) ensure that appropriate accommodations are available for students who need them; and (3) include measures of growth toward grade level targets, such as growth models that represent student progress over time.

Accessible assessments should be developed for all students using the principles of UDL that provide for proactive design that minimizes the need for accommodations (e.g., time, font size, and language complexity). [For information on UDL, see the guidelines published by the National Center on UDL at <http://www.udlcenter.org/aboutudl/udlguidelines>].

When students do require accommodations, they should be provided with the widest range of assessment accommodations feasible. To do this, tests must be designed with a clear specification of the constructs and skills that are being assessed and validated using a variety of accommodation options. This allows teachers to identify which target skills are necessary to successfully participate in the assessment and which accommodations might be needed for each student. [Recently published by the National Accessible Reading Assessment Projects (NARAP), a set of principles for creating accessible reading assessments can help guide

the development of future reading assessments. This document is available at <http://www.narap.info/publications/reports/NARAPprinciples.pdf>.]

Test design should consider how construct elements affect accessibility. For example, word choice, alternate answer choices, graphics, and cognitive demand can dramatically interfere in test item difficulty and yet have very little to do with the skill being assessed. Test items should require authentic demonstration of skills and knowledge for all students, including students with disabilities.

Alternate assessments can be effective vehicles for measuring student skills, knowledge, and growth. However, the administration time for alternate assessments should not decrease the instructional time that these students need.

- *NEA recommends the use of Universal Design for Learning principles when developing and implementing assessment systems, and when considering the accessibility and availability of accommodations and alternate assessments within such systems.*

Addressing the Needs of Students with Disabilities

Students with disabilities are currently subjected to assessments that leave them frustrated and demoralized due in large part to the fact that accessible, meaningful assessments are not readily available to them. This must change. NEA strongly supports fully including students with disabilities in state and local assessment and accountability systems. It is critical that assessments for students with disabilities be valid, reliable and meaningful. Special education professionals should be engaged in all aspects of standards and assessment development, including scoring.

Administering assessments to students with disabilities must comport with the IDEA and the recommendations of a student's Individualized Education Program (IEP) team. The IEP itself should not be used as an assessment. Since IEP teams determine which assessment students will take and what accommodations will be provided, educators need to be provided with professional development on the use of accommodations in instruction and assessment. In particular, classroom teachers and members of IEP teams need to understand the impact of the accommodation recommendations and alternative assessments defined in students' IEPs.

Additionally, funding must be provided to help states develop more appropriate assessments for students with disabilities and other special populations.

- *NEA recommends the development of assessments that are appropriate, valid, and reliable, as well as designed for the full inclusion of and administration to students with disabilities and that comport with a student's Individualized Education Program and the Individuals with Disabilities Education Act.*

Addressing the Needs of English Language Learners

It is essential to ensure that the unique factors that impact the performance of English language learners are specifically addressed in assessments used to measure and report the academic achievement of these students. Assessments must be sensitive to the various forms of diversity, including cultural, both within and across subgroups such as ELL students with learning disabilities. It cannot be assumed that assessment accommodations adopted for one subgroup will be effective or valid for other subgroups.

It is difficult to determine the effectiveness of assessments in native languages since there are many different languages represented in this country. Also, the validity of using first language for assessments of ELL students depends on the language of instruction and the level of students' fluency in English.

- *NEA proposes that ELL students be assessed in a valid and fair manner by addressing the unique factors that impact the performance of ELLs and ELLs with learning disabilities. States should be required to validate assessment systems for ELL, and provide research-based recommendations for selecting and using appropriate assessments and accommodations for ELLs, to ensure that these students have*

appropriate and multiple pathways to demonstrate content knowledge, skills, and abilities in accordance with intended learning standards and instructional goals.

- *ELL students should be included in appropriately designed assessment results as soon as practicable, but no later than three years after their entry into the public school system.*

Fostering Teacher Capacity and Development

Professional development to enable teachers to use assessments and assessment data is an essential component of a complete, effective assessment system. There are several forms of building teacher capacity related to an assessment system. One is the use of teachers to score summative assessments. This provides opportunities to develop deeper understandings of the assessments, how students respond to assessment tasks, and what might be done in the classroom to improve student learning. This type of professional development is currently rare in the United States but characteristic of assessment systems in other countries. Called “moderation” in countries such as Singapore, this activity requires that time be added to the school year solely for the purpose of teachers participating in scoring activities. There is evidence that developing and scoring assessments is an effective investment in professional development (Darling-Hammond & Rustique-Forrester, 2005).

The second type of professional development occurs when teachers meet to share and discuss students’ performance on formative assessments. This is assessment *as learning*. It leverages the capacity to make connections among assessments, standards, and curriculum. The ultimate implementation of such a system would lead to professional development for teachers with regard to the single most important “assessment” done by teachers – classroom grading.

A third component of professional development addresses the capacity to understand and analyze data from all types of assessments and collapse/integrate/use these multiple forms of data to inform instructional planning. This can be accomplished using combinations of presentations of information, making resources available to teachers, and providing for teacher discussion of assessment data.

- *NEA proposes that ESEA provide support in terms of funding and priorities for professional development that increases the capacity of teachers to design, score, analyze, use and integrate assessments.*

Developing Effective Data Systems

An effective data system is an essential component of a complete assessment system. An effective data system not only helps stakeholders keep track of student scores on summative assessments but enables teachers to use technology to choose among optional formative assessment tasks and resources, to keep track of data on formative assessments, and share resources and insights with other teachers.

Singapore spent more than \$100 million to develop their educational data system, but it appears to have helped that country promote high levels of learning and strong teacher capacity.

For an assessment system to be effective, it also should address contextual variables such as teacher capacity, school climate, community support, and school health and safety. Data on these variables are part of the accountability and assessment systems in Alberta, Canada; Queensland, Australia, and high achieving countries around the world. Incorporating these factors into the assessment system acknowledges the reality that what takes place in classrooms cannot be disconnected from the conditions and experiences of children outside of school.

Measures of teacher effectiveness should be based on records, such as those required to become a National Board Certified Teacher. This would allow the use of sound, validated empirical data on teacher expertise, rather than relying on linking student performance to specific teachers over a limited time without attention to other variables known to affect student achievement beyond the influence of teachers. School climate is another contextual variable that should be included in the data system. A school review system, such as those in Britain and New Zealand and similar to what is proposed by the Broader, Bolder Approach to

Education, would allow for guided observation of teachers in classroom and characteristics of school climate, health, and safety. Parental support and involvement can be assessed through the use of parent surveys.

- *NEA supports the development of effective data systems that capture and organize multiple types of data on student learning, teacher capacity, and school and district effectiveness.*
- *Assessments must be used only for the purposes for which they are designed.*
- *NEA proposes that ESEA foster school review systems developed by states, especially prior to implementing strategies to improve struggling schools.*

Transition to New Assessment and Accountability Systems

To avoid overlapping and conflicting accountability systems upon reauthorizing ESEA, Congress must immediately replace NCLB accountability labels and requirements with a new, strengthened accountability system. To address the obvious need for a transition to this new system, Congress should thoughtfully consider how and which assessments will be administered and used pending the implementation of new assessment systems under Race to the Top and other assessment reform efforts. Furthermore, we strongly believe NCLB-era assessment results should no longer be used for federal accountability purposes after ESEA is reauthorized. The cessation of the NCLB accountability timeline—and the all-too-often inaccurate school labels—is critical to allow states to begin developing more complete accountability systems comprising multiple measures of student learning. States also will use this time to pilot and ramp up new assessment instruments under the new accountability system so that they may be used as soon as possible.

- *NEA proposes that Congress prioritize allowing states and districts to strengthen their assessment and accountability systems by minimizing overlapping and conflicting NCLB-era requirements.*

Select References

Ahearn, E. (2009). *Growth Models and Students with Disabilities: Report of State Interviews*. Washington, DC: National State Directors of Special Education, Project Forum.

Bennett, R. (2009). The next generation of K-12 reading assessment. Presentation at the ETS Research Forum, Washington, DC, May 3, 2009.

Bennett, R. & Gitomer, D. (2009). Transforming K-12 assessment: Integrating accountability testing, formative assessment, and professional support. In c. Wyatt-Smith & J. Cumming (Eds.), *Educational assessment in the 21st century*. New York: Springer.

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

Darling-Hammond, L. and Wood, G. H. (2008) Refocusing accountability: Using performance assessments to enhance teaching and learning for higher order skills. The Forum for Education and Democracy

Johnstone, D.J., Altman, J., & Thurlow, M. (2006). *A state guide to the development of universally designed assessments*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes