

**Assessment and Accountability for
Improving Schools and Learning:**

**Principles and Recommendations for
Federal Law and State and Local Systems**

Expert Panel on Assessment

June 2007

The Forum on Educational Accountability (FEA) is a working group of national education, civil rights, religious, disability, civic and labor organizations that have endorsed the *Joint Organizational Statement on No Child Left Behind (NCLB) Act*. FEA convened the Expert Panel on Assessment to provide guidance on implementing the principles of the Joint Statement. The panel's mission was to use the Statement as its starting point and develop recommendations for changes to ESEA/NCLB. The report represents the views of the panelists. FEA has published it to provide policymakers and the public with concrete and detailed approaches for implementing the ideas in the Joint Statement.

This publication is available online at www.edaccountability.org.

This report may be freely reproduced, provided that proper credit is given and no profit is made from reproduction or dissemination.

Published June 2007

Copyright © 2007 by the Forum on Educational Accountability

All Rights Reserved

Assessment and Accountability for Improving Schools and Learning: Principles and Recommendations for Federal Law and State and Local Systems

Expert Panel on Assessment and Accountability
June 2007

Members of the panel:

Jamal Abedi
Pete Goldschmidt
Brian Gong
Margo Gottlieb
Alba A. Ortiz
Pedro Pedraza
James Pellegrino
Pat Roschewski
Jim Stack

Monty Neill of FairTest and Marcie Dianda and Beth Foley of the National Education Association provided staff assistance to the panelists.

Forum on Educational Accountability
www.edaccountability.org

Table of Contents

Executive Summary

Introduction

Principle I: Equity and Capacity Building for Student Learning

Principle II: Comprehensive State and Local Assessment Systems

Principle III: Assessment and Accountability for Diverse Populations

Principle IV: Fair Appraisal of Academic Performance

Principle V: Fair Accountability Decisions

Principle VI: Use of Assessment and Accountability Information to
Improve Schools and Student Learning.

Appendix 1: Members of the Panel

Appendix 2: Joint Organizational Statement on No Child Left Behind (NCLB)
Law

Assessment and Accountability for Improving Schools and Learning: Principles and Recommendations for Federal Law and State and Local Systems

**Expert Panel on Assessment
Convened by the Forum on Educational Accountability**

June 2007

Executive Summary

Public education underlies much of what is great about America—our shared commitment to democratic ideals, our social inclusiveness, our economic prosperity, the opportunities available to those seeking a better life, and our highest ideals of “liberty and justice for all.” Today, there is no question that education at all levels must be improved if it is to fulfill these promises to the next generation of Americans.

The education and assessment experts who wrote this report represent a broad array of stakeholders concerned about the reauthorization of the Elementary and Secondary Education Act (ESEA), of which the No Child Left Behind Act (NCLB) is the latest version. The panelists seek to ensure that the next version of the law supports the intended purpose of NCLB as outlined in Section 1001: “Ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education.”

The panel concludes that significant changes must be made to NCLB in order to meet this goal. This report advances a set of recommendations grounded in six guiding principles and provides a vision for an inclusive, beneficial, and fair assessment and accountability system within a strong, equitable, and steadily improving educational system. The panel urges Congress and state policy makers, as well as educators and the public, to consider these recommendations as the law is reauthorized and implemented.

Principles and Primary Recommendations

Principle I: Equity and Capacity Building for Student Learning

Help states, districts, and schools fulfill their educational responsibilities to foster student learning and development by ensuring that all students have equitable access to the resources, tools, and information they need to succeed and by building capacity to improve teaching and learning.

1. Ensure all students have access and support to succeed in a rich curriculum.
2. Provide the equitable opportunities to learn needed to reach the ambitious goals for student achievement.

3. Focus on developing local capacity through incentives and support.
4. Match needed flexibility with increased local responsibility for implementing the law in ways that meet its goals and intents.

Principle II: Comprehensive State and Local Assessment Systems

Construct comprehensive and coherent systems of state and local assessments of student learning that work together to support instruction, educational improvement and accountability.

1. Provide incentives for states and districts to develop comprehensive and coherent assessment systems that inform instruction and decision-making in ways that state tests alone cannot and do not. Coherent and comprehensive assessment systems provide evidence of student and school performance in relation to rich and challenging educational goals, using multiple indicators of student learning from a variety of sources at multiple points in time.
2. Provide states incentives and supports to include high quality local assessment systems in meeting ESEA's accountability requirements, alone or by augmenting state assessments. Fund pilot projects in which interested states demonstrate how they can meet ESEA's accountability requirements through standards-based, locally-developed assessments of students' learning or by integrating local assessments with state assessments. Fund expansion of the number of supported projects as states indicate interest. Provide incentives for states to work together.
3. Provide tools for states and districts to self-evaluate and improve the coherence and effectiveness of their local comprehensive assessment systems. The assessment and instructional components should work together to support instructional improvement and educational accountability.

Principle III: Assessment and Accountability for Diverse Populations

Shape the design, construction, and application of assessment systems so they are appropriate for an increasingly diverse student population.

1. Design assessments based on principles of universal design, but ensure that the unique factors that impact the performance of subgroups (e.g., English language learners (ELLs), students with disabilities (SWDs), students from major racial and ethnic groups, or economically disadvantaged students) are specifically addressed in the assessments that are used to measure the academic achievement of these students and reporting of results.
2. Require states to provide research-based recommendations for selecting and using appropriate accommodations for ELLs and SWDs to ensure that these students have access to valid assessments of their content knowledge.
3. Require states to validate assessment systems for each subgroup.
4. Support research to address major issues that complicate the design of appropriate assessment systems for subgroups.
5. Provide incentives for states to work together to shape the conceptual design and construction of local and state assessments of academic achievement according to the characteristics of each specified subgroup. Federally fund research to address the most pressing technical issues related to assessments and accountability decisions for English language learners and students with disabilities.

Principle IV: Fair Appraisal of Academic Performance

Use multiple sources of evidence to describe and interpret school and district performance fairly, based on a balance of progress toward and success in meeting student academic learning targets.

1. Encourage states and districts to use multiple sources of evidence drawn from their comprehensive and coherent systems of classroom-, school- and district-based assessments to summarize and appraise student performance.
2. Encourage states to describe school performance in terms of status, improvement, and growth, using the states' multiple sources of evidence.
3. As states evaluate their assessment systems, conduct ongoing studies of the validity of the descriptions and interpretations of student and school performance to ensure the quality of core data analysis and reporting.

Principle V: Fair Accountability Decisions

Improve the validity and reliability of criteria used to classify the performance of schools and districts to ensure fair evaluations and to minimize bias in accountability decisions.

1. Encourage states to include all subjects – not just reading, math and science – in their comprehensive assessment systems, but use compensatory processes to ensure that the inclusion of more subjects does not become another means for schools and districts to fail accountability requirements.
2. Encourage states and districts to use multiple sources of evidence drawn from their comprehensive and coherent assessment systems to make accountability decisions about the quality of school and district performance and determine which schools and districts need what forms of assistance.
3. Retain the ESEA requirement for gathering and reporting disaggregated information by subgroups based on the comprehensive assessment system.
4. Use collective research from the states to establish realistic and challenging federal guidelines for rates of growth or improvement towards the goal of reaching specified learning targets.
5. Replace the current rules for AYP classifications with reliability and validity criteria that each state must apply when designing its accountability classification system so that it is fair and minimizes bias.
6. Use accountability decisions to inform assistance to schools.

Principle VI: Use of Assessment and Accountability Information to Improve Schools and Student Learning

Provide effective, targeted assistance to schools correctly identified as needing assistance.

1. Encourage states and districts to use multiple sources of evidence from state and local assessments and other forms of evidence to inform actions such as interventions and technical assistance.

2. If a school (or district) is identified as not making sufficient progress towards improvements or in outcomes, the district (or state) would investigate causes and undertake a series of interventions tailored to address particular needs.
3. Assistance may include providing professional development, developing partnerships with parents and families, improving curriculum, and attracting and retaining high quality teachers and administrators.
4. NCLB mandates for governance changes should be removed or, at most, made an option for possible action only after implementation of recommendations 2 and 3 (above).
5. The accountability systems should ultimately be judged on their consequences for the quality of the educational system and the learning of its students.

Note: The Forum on Educational Accountability (FEA) convened the Expert Panel on Assessment to build on the *Joint Organizational Statement on No Child Left Behind (NLCB) Act*. The panel's mission was to use the Statement as its starting point to develop recommendations for changes to ESEA/NCLB. This report therefore represents the views of the panelists. The full panel report is available at www.edaccountability.org, along with other FEA materials and the Joint Statement.

Members of the Expert Panel: Jamal Abedi, Pete Goldschmidt, Brian Gong, Margo Gottlieb, Alba A. Ortiz, Pedro Pedraza, James Pellegrino, Pat Roschewski, Jim Stack. Staff assistance provided by Monty Neill, Marcie Dianda and Beth Foley.

Introduction

Overall, the law's emphasis needs to shift from applying sanctions for failing to raise test scores to holding states and localities accountable for making the systemic changes that improve student achievement.

-- Joint Organizational Statement on NCLB

Public education underlies much of what is great about America—our shared commitment to democratic ideals, our social inclusiveness, our economic prosperity, the opportunities available to those seeking a better life, and our highest ideals of “liberty and justice for all.” Today, there is no question that education at all levels must be improved if it is to fulfill these promises to the next generation of Americans.

The federal No Child Left Behind (NCLB) Act dominates current educational policy discussions. As the U.S. Congress begins the reauthorization of the Elementary and Secondary Education Act (ESEA), of which NCLB is the current version, the debate about how to retain what is working and fix what is not has become more intense and more important. What are the law’s strengths? What are its shortcomings? What should Congress change to improve the law?

The education and assessment experts who wrote this report want to make sure that the next version passed by the Congress supports the intended purpose of NCLB as outlined in Section 1001: “Ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education....”

The panel concludes that significant changes must be made to NCLB to meet this goal. This report advances a set of recommendations grounded in six guiding principles and provides a vision for an inclusive, beneficial, and fair assessment and accountability system within a strong, equitable, and steadily improving educational system. The panel urges Congress and state policy makers, as well as educators and the public, to consider these recommendations as the law is reauthorized and implemented.

The panel drafted the guiding principles within a framework of three foundational belief statements about the importance of the accountability required by No Child Left Behind:

1. Accountability decisions for student learning are the responsibility of states, and each state accepts that responsibility.
2. Accountability decisions made within each state must be valid based on appropriate and accurate data.
3. Accountability decisions within states must benefit students and support student learning.

Therefore, an effective education system must be rooted in more than state compliance with federal accountability requirements. Although sanctions imposed by the federal or state government may force schools to engage in a great deal of activity, the educational system America needs requires more than compliance — it requires contributions from every teacher,

parent, local school board member and student, along with the engaged support of the community. The reauthorized federal law must actively empower stakeholders to enact positive changes in classrooms and schools, and it must foster a culture of local responsibility, as well as continue helpful forms of external accountability.

The Forum on Educational Accountability (FEA) convened the Expert Panel on Assessment to build on the *Joint Organizational Statement on No Child Left Behind (NLCB) Act*. The panel used the Statement (see Appendix 2) as its starting point to develop recommendations for changes to ESEA/NCLB; therefore, this report represents the views of the panelists.

The six principles provide the basic structure for this report. Each section states the principle, followed by primary and subordinate recommendations. Problems with the current NLCB in light of the principle are then presented. Each of the sections concludes with a more detailed explanation of the recommendations and discussion of how their application will solve the identified problems and support high quality education, assessment and accountability. A limited set of references follows. The report ends with an appendix with the author's biographies (Appendix I) and an appendix containing the *Joint Organizational Statement on NCLB* (Appendix II).

Jamal Abedi
Pete Goldschmidt
Brian Gong
Margo Gottlieb
Alba A. Ortiz
Pedro Pedraza
James Pellegrino
Pat Roschewski
Jim Stack

Principle I: Equity and Capacity Building for Student Learning

Help states, districts and schools fulfill their educational responsibilities to foster student learning and development by ensuring that all students have equitable access to the resources, tools, and information they need to succeed and by building capacity to improve teaching and learning.

Primary and Subordinate Recommendations

1. Ensure all students have access and support to succeed in a rich curriculum.

- a. Provide incentives for schools to develop and maintain a balanced, enriched curriculum, recognizing that schooling is more than reading and math, and knowledge is more than what can be measured by test scores. Decrease the pressures from accountability that lead to excessive narrowing of the curriculum in scope or depth.
- b. Ensure alignment of assessments with standards so that assessments adequately assess the full range of core learning.
- c. Fund increased research to support the implementation of educational reform across the states and territories. Such implementation does not mean moving towards national standards, as standards should remain under state control.

2. Provide the equitable opportunities to learn needed to reach the ambitious goals for student achievement.

- a. Make “opportunity to learn” part of the needs analysis required for every school/district. Students should be held accountable for their learning only after they have had adequate opportunities to learn. Schools and districts must be held accountable to provide those opportunities – and policymakers must hold themselves accountable for addressing inequities in access to necessary resources such as skilled teachers, appropriate curricula and materials, and adequate facilities.
- b. Move from “reallocating the pie” and demanding more efficiency from the same resources to a major investment in our educational infrastructure, including human capital. Policymakers may need to devise methods for holding educational systems accountable for effective and efficient uses of additional resources.

3. Focus on developing local capacity through incentives and support.

- a. Build state and local capacity by ensuring participation of all key stakeholders in the improvement process.
- b. Provide the resources states and districts need to leverage their capacity to make ongoing technical assistance and high-quality professional development available to schools and school staff.

- Invest in targeted professional development of educators to address critical assessment needs, including assessment literacy (including summative and especially formative assessments); skills in data analysis, interpretation, and use in educational decision-making; and local assessment infrastructure.
- c. Invest in state and district capacity to develop, refine, and improve educational standards and assessments so that the accountability system advances important outcomes and rests on information derived from state-of-the-art assessments as well as the use of that information by skilled teachers.
- 4. Match needed flexibility with increased local responsibility for implementing the law in ways that meet its goals and intents.**
- a. Support local feedback and self-evaluative mechanisms so that assessment and accountability systems can be improved and unintended harmful consequences can be identified and corrected.
- b. Improve the administration of the law by the U.S. Department of Education (USDE) to support greater transparency, clarity, and timeliness in communications about regulations and guidance. The law should be administered in ways that support best practice and innovations rather than attention to compliance.

Problems with the Law

Passage of the No Child Left Behind Act (NCLB), the current authorization of the long-standing Elementary and Secondary Education Act (ESEA), saw an immediate and sizeable jump in ESEA funding. Those increases subsequently flattened so that funding has been stagnant, leaving the law significantly under-funded. Allocations for the law's Title I allows less than one-half of eligible schools to receive funds.

As NCLB came into effect, state education was stagnating or in decline. In almost every state, moderate to severe funding inequities between schools persist. Low-income students typically attend under-resourced schools. Those students are also more likely to suffer severe social and economic problems -- including low and erratic family incomes, racial discrimination, frequent moves, inadequate nutrition and medical and dental care, and community instability -- that impact their preparedness for schooling. The schools they attend typically have fewer qualified teachers and more substitute teachers; lack books, libraries, computers and laboratories; occupy dilapidated buildings; and have large class sizes.

NCLB expects states, districts and schools to enable all students to score “proficient” on state assessments, but fails to ensure that students and schools are enabled to meet the demands placed on them. The law's framers may have intended that NCLB would spur states to improve funding equity and adequacy, but this generally has not happened.

NCLB thus devolves to an apparent theory of action in which fear of sanctions will compel educators to intensify efforts to raise student achievement. Indeed, educators are working hard to

meet the law's demands. However, the stated goal that all children will reach the proficient level by 2014 is unattainable. A key consequence is that improvements in reading and math scores are coming at the expense of narrowing the curriculum and focusing on what is tested. Rather than improve system capacity to ensure all children receive a high-quality education, the efforts too often focus on compliance to avoid sanctions.

The U.S. Department of Education's (USDE) implementation has been marked by shifts in tone and approach, a lack of transparency, and different “deals” for different states. This erratic approach reduces the incentives for localities and states to take responsibility, within reasonable federal guidelines, for education systems that serve all children well.

A review of the law, its implementation and consequences leads us to conclude that it:

- fails to ensure adequate opportunities to learn;
- is too rigid in its conceptualization and application;
- sets forth goals that are impossible to meet and far too narrow; and
- inhibits rather than strengthens local capacity to improve the quality of education.

Improvements to the Law

Implementation of the recommendations in this Principle will redirect the federal role toward assisting districts and schools to serve all children well. They will reconstitute the unrealistic and often damaging requirements found in NCLB. Many of these recommendations are consonant with the other organizations', some of which have addressed particular aspects in more detail. The remainder of this report focuses on assessment and the uses of assessment evidence in accountability and school improvement, including the need to establish challenging yet reasonable goals for improvement in schools, systems and student learning.

References

- Center on Educational Progress has issued a series of reports on the implementation of NCLB. www.cep-dc.org.
- Kozol, J. (2005). *The shame of the nation: The restoration of apartheid schooling in America*. New York: Crown.
- Linn, R. (2006). *Toward a more effective definition of adequate yearly progress*. <http://www.law.berkeley.edu/centers/ewi/research/k12equity/Linn.htm>
- Meier, D. and Wood, G. (Eds.) (2004). *Many children left behind: How the No Child Left Behind Act is damaging our children and our schools*. Boston: Beacon Press.
- Neill, M., Guisbond, L., and Schaeffer, B. (2004). *Failing our children*. Available at http://www.fairtest.org/Failing_Our_Children_Report.html
- Nichols, S.L., and Berliner, D.C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard University Press.
- Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap*. New York: Teachers College Press.
- Sunderman, G.L., Kim, J.S., and Orfield, G. (2005). *NCLB meets school realities: Lessons from the field*. Thousand Oaks, CA: Corwin Press.

Valenzuela, A. (Ed.) (2005). *Leaving children behind: How "Texas-style" accountability fails Latino youth*. Albany, NY: State University of New York Press.

Principle II: Comprehensive State and Local Assessment Systems

Construct comprehensive and coherent systems of state and local assessments of student learning that work together to support instruction, educational improvement and accountability.

Primary and Subordinate Recommendations

- 1. Provide incentives for states and districts to develop comprehensive and coherent assessment systems that inform instruction and decision-making in ways that state tests alone cannot and do not.** Coherent and comprehensive assessment systems provide evidence of student and school performance in relation to rich and challenging educational goals, using multiple indicators of student learning from a variety of sources at multiple points in time.
 - a. Provide funding for school districts to develop rich local assessment systems that address state content standards and incorporate classroom-based and common (developed and used by groups of teachers at the school and district levels) or standard school- and district-based measures.
 - This would provide resources to enable development of assessments at all levels, particularly addressing important standards that cannot be assessed well with current large-scale assessments, such as reasoning, communication, problem-solving, research, oral communication, and applied learning. Comprehensive assessment systems would address these areas through employing multiple appropriate assessment practices and tools, including: teacher observations; tests that include multiple-choice, short and longer constructed response items; essays; tasks and projects; laboratory work; presentations; and portfolios. It would also include development of assessments for specific subgroups, including English language learners (ELLs) and students with disabilities (SWDs).
 - b. Provide states additional resources to support local system implementation, including professional development, technical assistance and technology infrastructure.
- 2. Provide states incentives and supports to include high quality local assessment systems in meeting ESEA's accountability requirements, alone or by augmenting state assessments.** Fund pilot projects in which interested states demonstrate how they can meet ESEA's accountability requirements through standards-based, locally-developed assessments of students' learning or by integrating local assessments with state assessments. Fund expansion of the number of supported projects as states indicate interest. Provide incentives for states to work together.
- 3. Provide tools for states and districts to self-evaluate and improve the coherence and effectiveness of their local comprehensive assessment systems.** The assessment and instructional components should work together to support instructional improvement and educational accountability.

4. Design state and district reports that include all components of their comprehensive assessment system. State reporting should reflect the presence and weighting of the results from each measure that comprises its assessment system, including those unique for ELLs and SWDs. District level reporting should mirror the results of measures, both local and state, used in its accountability system.

Problems with the Law

In response to the 1994 reauthorization of ESEA, most states simply used standardized tests as the sole means of assessing student progress. Most then expanded and intensified the uses of those tests to comply with NCLB. The result is that NCLB relies overwhelmingly on large-scale, statewide accountability tests as the primary source of information about students' learning. The law and the regulatory guidance make little provision for valid and reliable assessment results from classrooms, schools and districts, though the law does permit the use of such assessments if they are part of state systems. Expansive state assessment systems that serve accountability purposes dominate, while local assessments that support teaching and learning atrophy.

Large-scale state accountability tests can provide a valuable one-point-in-time snapshot of students' performance in relation to key state academic standards. Carefully constructed locally developed measures, by comparison, are needed to assess complex key standards such as research, communication and problem-solving that are essential aspects of strong education. Local assessments also are the only way to generate ongoing assessment information that schools can use to monitor students' learning and revise instruction, as needed, throughout the school year. Reliable and valid local assessments help schools to continually improve students' achievement.

While NCLB recognizes the need for "diagnostic assessments" to meet students' individual needs, large-scale statewide accountability tests, by definition, do not provide this information. Standardized once-a-year tests help ensure that students are not systematically held to lower standards, but they do not provide adequately detailed information about individual students' specific strengths and needs, and the test results are not provided in a timely manner in order to inform teaching and learning. In addition, by design—and in contrast to classroom-, school-, and district-based assessments—they are often insensitive to particular curricula or instructional approaches.

In contrast, assessments developed for classrooms, schools, and districts are often embedded into curriculum and instruction and are administered throughout the school year. As a result, classroom-based assessments measure actual student learning, match and more fully address the curriculum being taught, and provide timely diagnostic information to improve instruction and programs.

Reliance on state large-scale tests has resulted in some excessive narrowing of the curriculum, clearly an unintended and negative outcome. Attaching high stakes to the typical standardized state test and low or no stakes to other means of assessment tends to narrow both instructional and assessment practices in classrooms, replacing an emphasis on critical thinking with a focus

on basic skills. A growing body of evidence indicates that NCLB provides strong incentives to focus attention disproportionately:

- on the subjects that are assessed by statewide tests, to the detriment of those that are not; and
- on those students who are below, but within reach, of proficiency targets, to the detriment of very high and very low achievers.

These are predictable outcomes from a system that has focused almost exclusively on math and reading scores. The system narrows the educational process and loses the interest of students who might well have been engaged by instruction in the neglected fields or by varied means of instruction. As a further consequence, socially valuable non-academic knowledge and skills are given increasingly short shrift in schools, particularly schools serving the economically disadvantaged, minority and limited English proficient students, and students with disabilities who are the primary focus of the law's concern.

Improvements to the Law

ESEA will be improved dramatically if it advances *comprehensive assessment systems* that map out appropriate roles for state and local assessments, meet accountability demands, and help shape instructional programs and classroom learning. A comprehensive assessment system would be comprised of state and local components benchmarked to meaningful standards.

Local assessments include teacher-based assessments (their own or ones adapted from other sources) conducted in their own classrooms. They also include “common” assessments, which are developed or adopted by groups of teachers at the school and district levels. Local assessments may include specific instruments or practices (e.g., tests and projects, teacher observations or discussions with students) and compilations of evidence of learning (e.g., portfolios, exhibitions and learning records).

Local assessments perform two legitimate, important, different, and complementary roles. First, they provide information to guide teaching and local and classroom instructional decision-making that state tests cannot and do not provide, and as such, they can operate parallel to the state assessment system without actually being integrated into it.

For example, “formative” assessments – ongoing assessments used to shape instruction and provide feedback to students — have been shown to be powerful tools to assist student learning and provide greater assistance to low-performing students. The use of such formative assessments must be primarily in the hands of teachers. Not all “summative” assessments (i.e., those that provide periodic summary evaluations of student learning such as at the end of a semester or year) are or should be designed for use in accountability; much summative evidence is properly the domain of teachers, students and schools.

A second, equally appropriate use of local assessments is to build some of them into the state’s accountability system to help states meet ESEA’s accountability mandates. They can provide locally-based valid and reliable summative information about students’ learning. The use of locally developed assessments as acknowledged components of the ESEA accountability system

would assign to states the responsibility to assist local districts in developing assessments with sufficient fairness, reliability and validity to meet established accountability requirements.

Questions of how to assemble and weight multiple components from different sources could be addressed in the state pilot programs we recommend. However, no one test should be weighted so heavily that educators effectively ignore other assessments because the one test is what really “counts.”

The local assessments would be in substantial part standards-referenced; that is, they would assess knowledge and skills described in state and/or local content standards. The standards, in turn, should be clear, concise, and limited to grade-or subject-specific content. Local assessments should be valid and fair for all students (see Principle III).

Local assessments also have a vital role in school improvement and in supporting students who are at risk academically. Used throughout the school year, particularly as needed by teachers, they provide local educators with rapid and regular feedback useful for making instructional adjustments in a timely fashion. These assessments should not simply mirror or seek to predict outcomes on state large-scale tests, as is the current trend with the “benchmark” or “interim” tests that have become popular, particularly in urban districts. These can have the effect of narrowing curriculum and instruction even more as teachers devote additional time to preparing students for benchmark testing.

Local and classroom-based assessments play an effective role in improving teaching and learning and therefore educators must be provided the opportunity to strengthen their assessment skills. Such skills include selecting and creating assessments of various sorts to meet the multiple purposes of instruction, accountability and improvement, as well as using an assessment repertoire to meet the needs of each and every student. It also involves techniques of observation and documentation, and helping students learn to self-evaluate. Thus, ESEA should ensure that professional development and school improvement activities include a significant opportunity for teachers to develop comprehensive assessment knowledge and skills.

In addition, local assessments provide more information to parents, policymakers and the public than do state assessments. To this end, resources, including time, should be allocated to enable states and districts to explain the purposes of assessment and the results of those assessments to the public.

Two remaining questions: Can locally-developed assessments alone enable states to meet ESEA’s accountability requirements? And, can classroom-based evidence be used in such a system?

Nebraska is the one state that developed a system that uses only local assessments for accountability. The types of assessments vary across the state, ranging from augmented standardized tests to classroom-based assessments. The USDE repeatedly challenged Nebraska's assessment program, but has now given it “approval pending” status, with full approval expected in summer of 2007. In addition, several states have implemented or announced local assessment information as a requirement for high school graduation or high school diploma endorsement,

including Wyoming and Rhode Island. These efforts should be recognized and encouraged for school accountability purposes as well.

We believe that current assessment technology is sufficiently advanced to support the development and use of locally-developed, standards-based assessments that accurately determine what students know and can do. Quality assurance reviews can be designed and implemented for systems using local assessments.

Teacher developed assessments have long been noted as often idiosyncratic or of low quality. These might suggest the need for common (school or district) assessments rather than teachers' individual assessments. However, teachers in Nebraska have been creating individual assessments of sufficient reliability and validity to meet USDE acceptance. Additionally, various programs have enabled teachers to compile evidence from classroom work that have been used to reliably and validly judge student achievement.

In Nebraska school districts, assessments are embedded into the curriculum and administered at the point of instruction. Local assessments must meet state requirements for technical measurement quality. Traditional measurement techniques used in large-scale assessment have had to be adjusted and modified to apply to classroom-based assessment. The data are reported to the state at the end of the school year for public accountability, including NCLB.

Some fear use of classroom-based information for accountability could damage the quality and undermine important uses of classroom assessments, especially if accountability continues to focus on high-stakes sanctions. Others think that only from classroom evidence can sufficiently rich evidence be obtained to show how well students are learning a comprehensive curriculum, that state and local tests will inevitably be too narrow and thereby undermine not only assessment but also curriculum and instruction. This issue could be resolved in the recommended state pilot projects in which classroom-based evidence is used when it meets relevant technical requirements. The consequences of accountability uses of classroom assessments could then be monitored and evaluated.

References

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-73.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *Assessment for learning – putting it into practice*. Buckingham: Open University Press.
- Brookhart, D.M. (2003). Developing measurement theory for classroom assessment: Purposes and uses. *Educational Measurement: Issues and Practice* 23 (June): 17-25.
- Commission on Instructionally Supportive Assessment (2001). *Building tests to support instruction and accountability: A guide for policymakers*. Report commissioned by NEA, NAESP, NASSP, NMSP, & AASA. Available at: <http://www.nea.org/accountability/buildingtests.html>.
- Gallagher, C. (2007). *Reclaiming assessment: A better alternative to the accountability agenda*. Heinemann, Portsmouth, New Hampshire

- Koretz, D. (2006). *The pending reauthorization of NCLB: An opportunity to rethink the basic strategy*. Available at <http://www.law.berkeley.edu/centers/ewi/research/k12equity/Koretz.html>.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- National Research Council (2003). *Assessment in support of learning and instruction: Bridging the gap between large-scale and classroom assessment*. Committee on Assessment in Support of Learning and Instruction, Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Wilson, M., & Bertenthal, M. (2005). *Systems for state science assessment*. Washington, DC: National Academy Press.
- William, D. (2007). Keeping learning on track: formative assessment and the regulation of learning. In F. K. Lester Jr. (Ed.), *Second handbook of mathematics teaching and learning* (pp. 1053-1098). Greenwich, CT: Information Age Publishing.

Principle III: Assessment and Accountability for Diverse Populations

Shape the design, construction, and application of assessment systems so they are valid and appropriate for an increasingly diverse student population.

Primary and Subordinate Recommendations

- 1. Design assessments based on principles of universal design¹**, but ensure that the unique factors that impact the performance of subgroups (e.g., English language learners (ELLs), students with disabilities (SWDs), students from major racial and ethnic groups, or economically disadvantaged students) are specifically addressed in the assessments that are used to measure the academic achievement of these students and reporting of the results.
 - a. When developing assessments, consider the specific characteristics of each subgroup, in conjunction with standards. Assessments must be sensitive to various forms of diversity, including cultural, both within and across subgroups. It cannot be assumed that assessments or accommodations developed or adapted for one subgroup will be effective and valid for other subgroups. For example, the issues to be addressed in assessments and accommodations for ELLs and SWDs are not the same.
 - b. Align and integrate standards and assessments that are specifically crafted for a subgroup (such as ELLs or SWDs) into the overall assessment system.
 - c. Incorporate available research, evidence and principles of fairness and equity for subgroups into assessment systems. (For example, use results from empirical research to

¹ In these recommendations, universal design refers to the design and delivery of products and services that are usable by people with the widest possible range of functional capabilities (Assistive Technology Act of 1998). Educators apply universal design principles in creating classroom and school environments and in designing and selecting instructional approaches, practices, and materials so as to accommodate students with diverse abilities and backgrounds. English language learners, students with disabilities, and economically disadvantaged groups can benefit from assessments that are culturally responsive, linguistically clear, and free from construct irrelevant sources of variability.

The intent of universal design is to create classroom instructional environments that meet the needs of the widest range of students in order to minimize the need for special adaptations or modifications. Principles of universal design include, for example, flexibility, choice, elimination of unnecessary complexity, appropriate space, and environments that promote communication and interaction. However, it is recognized that when students struggle, despite instruction and environments based on principles of universal design, teachers must differentiate instruction to meet the unique needs of individual students or subgroups. This same concept applies to development and design of student assessments (i.e., assessments should be designed to accommodate the widest possible range of students, but, despite this, some student subgroups will need additional considerations).

indicate when ELLs may be tested in English on content-based assessments based on their level of English language proficiency.)

- d. Provide the opportunities and resources necessary to ensure that all subgroups have meaningful access to the content that is based on state standards.
- e. Require multiple forms of evidence in the assessment of all subgroups, particularly for ELLs and SWDs, including results of classroom-based assessments and performance of ELLs in the native language and/or in English, consistent with the language(s) in which they receive instruction or are best able to indicate their learning.
- f. Understand the diversity within the ELL student population (such as linguistic and cultural differences; continuity of educational experiences inside and outside the U.S.) and act accordingly.

2. Require states to provide research-based recommendations for selecting and using appropriate accommodations for ELLs and SWDs to ensure that these students have access to valid assessments of their content knowledge.

- a. While the principles of universal design should be applied to the assessment system for SWDs and ELLs, base selection of assessments or accommodations on the specific needs of the students being tested.
- b. Provide specific guidance for selection of assessments and/or accommodations for students with dual classifications (e.g., twice exceptional: ELLs with reading disabilities).

3. Require states to validate assessment systems for each subgroup.

- a. Include large enough numbers of students from specific subgroups in the validation process.
- b. Control factors that negatively impact assessment outcomes for subgroups so that variables that are not the primary interest in assessments of achievement do not affect assessment results. (For example, a test *in* English is a test *of* English for ELLs; therefore, English language proficiency may affect students' ability to demonstrate their academic achievement in English.)
- c. Require that states develop accountability systems that incorporate both growth and status measures. For example, emphasize growth when students are acquiring English language proficiency since language is a developmental process, and then shift the emphasis to a mix of status and growth (as described in Principle IV) when students have achieved the necessary proficiency (as determined through validation studies) to learn academic content taught entirely in English.

4. Support research to address major issues that complicate the design of appropriate assessment systems for subgroups. These include:

- A universal definition for ELLs ;
- Appropriate identification of ELLs and SWDs;
- Psychometric properties of English language proficiency assessments;
- Psychometric properties of both native language and English academic achievement assessments;
- Psychometric properties of assessments for the diverse types and severity levels of disabilities;
- Alternate assessments of academic achievement for ELLs, ELLs with disabilities, and non-ELLs with disabilities;
- Accommodations and modifications for specific groups and for subgroups within groups (e.g., by type or severity of disability);
- Alignment of assessments for SWDs with their Individualized Education Plans (IEPs);
- Criteria used for student participation in testing and the effects of arbitrary criteria applied to subgroups (e.g., percentage of students who can be exempted, limits on the number of times students can take native language achievement tests, or specification of when students have to be tested in English);
- Comparability of native language assessments, alternate assessments and regular assessments; and
- Language domains tested in Title I as compared with those tested in Title III for ELL students.

5. Provide incentives for states to work together to shape the conceptual design and construction of local and state assessments of academic achievement according to the characteristics of each specified subgroup. Federally fund research to address the most pressing technical issues related to assessments and accountability decisions for ELLs and SWDs.

- a. Provide incentives and technical assistance for states to improve current assessments, apply universal design, use multiple measures and growth models, as applicable to specific subgroups.
- b. Examine the extent of applicability of the principles of universal design to the design of state assessments.
- c. Establish criteria for ensuring that state assessments are relevant, equitable, valid, and of high quality for all students.
- d. Develop subject area assessments in languages other than English when students speaking a language form a significant proportion of the population.
- e. Research the validity of common district assessments designed for ELLs or SWDs as part of the state's comprehensive assessment system.

6. Phase in application of any consequences to schools contingent on the existence of valid and reliable academic assessments for these students.

- NCLB should respect the provision in the Individuals with Disabilities Education Improvement Act (IDEA) that the multidisciplinary team (a team of experts including teachers and parents) should determine whether the student will be included in district- and state-wide testing and the nature of that inclusion (same test, out-of-level, alternative, etc.). That would mean that the law would also eliminate arbitrary caps. The same can be said for ELLs, i.e., that placement committees should make decisions about their inclusion and the nature of inclusion in the assessment process based on valid evidence.

Problems with the Law

While the NCLB requirement to report by major subgroups brings highly needed attention to achievement issues for these students, it also introduces a whole new set of technical and ethical issues in the NCLB accountability system. English language learners and SWDs merit specific attention because they present unique challenges and require distinct approaches to their inclusion in assessment and accountability systems. It is thus important that descriptions of common concerns (e.g., the need for accommodations or alternate assessments) not be interpreted to mean that the same solutions can be applied to both groups, nor that the solutions diminish the rigor or validity of the system.

Although unintended, NCLB has advanced three accountability systems. One applies to the general school population; one that applies to ELLs; and one that applies to SWDs. English language learners face two sets of standards (English language proficiency and academic content), two forms of assessment (language proficiency and academic achievement) and two reporting systems. By definition, ELLs are not 'proficient' academically in English.

Because the needs of special student populations and diverse learners were not considered adequately when the law was drafted, schools that serve significant numbers of these students are in the difficult situation of having to demonstrate and receive credit for performance levels that meet the law's proficiency requirements in the absence of assessment systems appropriate for these learners. Consequently, schools with large numbers of students in several of these subgroups have a much greater chance of failing AYP requirements. For example, Linn (in press) indicated, "A school with a large enough number in, say, three racial/ethnic groups, students with limited English proficiency, economically disadvantaged students, and SWDs, would have a total of 29 hurdles to clear, four for each of the six groups plus the five that all schools have for the total student body." (p. 21) Therefore, mutually exclusive categories of subgroups need to be represented in analyzing and reporting assessment results.

Identifying factors affecting the performance gap between SWDs and ELLs with other students may also provide insight into assessment and accountability issues for other subgroups of students who may experience similar problems with the current assessment and accountability systems.

Issues concerning NCLB accountability for ELL students

Schools and districts with large numbers of ELLs are being punished for not meeting AYP when, by definition, these students are not proficient in the English language. By extension, they are not

proficient academically in English because their English language proficiency confounds achievement. For example, a number of studies have shown that it takes from five to seven years and even more for most ELLs to gain sufficient mastery of academic English to join English speaking peers in taking full advantage of instruction in English. During this time, learning must occur at a faster rate for ELLs to catch up with their proficient English peers, yet when that instruction is offered only in English, it cannot occur at the same rate as for a native speaker of English. However, the NCLB accountability system expects these students to reach the same achievement levels as their native English language peers within the same time frame.

For ELLs, proficiency in English is a developmental, longitudinal pathway defined by the same cohort of students as they progress through school. The relationship between English language proficiency and academic achievement varies as a function of such factors as students' foundation in their native language, continuity of schooling, and exposure to English, all of which NCLB does not take into account. As egregious, the law holds schools accountable for ELL students' at an absolute marker rather than examining their academic progress. Instead of looking at academic performance longitudinally, it examines ELLs' performance cross-sectionally and bases accountability decisions on test scores from different students each year.

In addition, using assessments developed for native speakers of English with ELLs is problematic since unnecessary linguistic complexity and lack of visual or graphic support often makes these assessments less reliable and less valid for ELL students. As stated in the *Standards for Educational and Psychological Testing*, tests that employ language are, in part, a measure of student language skills. ELLs may have the content knowledge to do well on assessments, but due to linguistic barriers, they are not able to demonstrate it. On the other hand, excluding ELLs from state and national assessments, as some have advocated, would have serious consequences on their academic outcomes (as is true for SWDs). Their academic progress, skills, and needs would not be appropriately assessed; the quality of instruction they receive would be affected; institutions would not be held responsible for their performance; and they would not be included in state or federal policy decisions. It is imperative to include ELLs in state assessment and accountability systems, but assessments must be valid, reliable, and yield meaningful results for these students.

Issues concerning NCLB accountability for students with disabilities

Some SWDs perform substantially lower than many of their non-disabled peers in some academic areas including math and reading. Therefore, reaching the same level of proficiency as their non-disabled peers can be challenging for these students and their educators.

NCLB does not acknowledge the tremendous variation in the nature and severity of students' disabilities, or persistent differences in student access to the general education curriculum. Standards and assessments for SWDs cover only academic goals, ignoring the social and vocational goals important for all students, including those with disabilities.

State assessments often have low reliability and validity for some subgroups. Issues concerning classification and assessment, as well as the appropriateness, effectiveness and validity of accommodations, have a major impact on assessment outcomes. Lack of sufficient

experimentally designed research on the effectiveness of accommodations and the validity and impact of accommodated assessments for these students makes the selection of appropriate accommodations difficult. On the other hand, assessments without using appropriate accommodations or modifications will not produce accurate measures of student achievement. (Note that the issues in this paragraph also often pertain to ELLs.)

Each of these factors alone and when interacting with each other may have profound impact on the fairness and validity of accountability systems for both SWDs and ELL students. Yet, NCLB requires most SWDs to perform at the same level as their non-ELL/non-disabled counterparts in the race toward proficiency for all students by the target year of 2014.

Improvements to the Law

ESEA would be improved dramatically if the application of consequences associated with accountability provisions for ELLs and SWDs were phased-in contingent on the existence of valid and reliable academic assessments for these students. For this to occur, the legislative architects of ESEA need to consider the body of research on the accessibility of assessments for SWDs and for students engaged in dual language development (i.e., native language and English as a second language) in the design of an accountability system that includes ELLs and SWDs. As important, they need to recognize that there are factors that may make assessment inaccessible for SWDs and ELLs. These factors should be identified through research and controlled within the design or administration of the assessment.

The academic achievement of ELLs who receive instruction in their native language should be assessed in that language with comparable, standards-based measures. Until they reach a threshold level of English language proficiency, ELLs cannot benefit from instruction and assessment in English even with appropriate accommodations.

Valid, reliable, and fair measures need to be developed for ELLs that demonstrate their academic improvement. In the interim, states should be required to document student's growth in English academic language proficiency, including how they serve those students and monitor their progress. Similarly, valid and reliable assessments need to be developed for SWDs that address their needs. It is important for ESEA to increase the assessment options for ELLs and SWDs. Multiple forms of assessment can present a more comprehensive picture of what ELLs and SWDs know and are able to do. Multiple forms of evidence, with documented reliabilities, such as hands-on performance, portfolio assessment, and performance-based assessment would potentially increase the validity and fairness of assessment for these students.

While one size fits all approaches to assessment are unrealistic, it is also impractical to develop many different assessments each tailored for a small group of students. It is thus important to adhere to the principles of universal design to create a comprehensive, aligned system, while, at the same time, considering the unique characteristics of students in the design of assessment and accountability systems. Using universal design, assessments would be developed, from the beginning, to provide access to the broadest possible range of students and to reduce the need for alternative assessments and accommodations. This goal may be accomplished, for example, by

minimizing item bias; providing simple, clear instructions and procedures; increasing visual support; and providing effective prompting and feedback on tasks.

In sum, the recommendations here include significant features that would make assessment more accurate and useful for all students. With universal design, many students with mild disabilities could take the traditional test since they would be crafted for a diverse student population. In addition, students will benefit from assessments that are developed with a sufficient number of SWDs and ELLs; assessments with a system of validated modifications and accommodations; alternative assessments for students with more significant disabilities; and native language assessments for students who were receiving native language instruction (i.e., students in bilingual education programs).

References

- Abedi, J. (2004). *Inclusion of students with limited English proficiency in NAEP: Classification and measurement issues*. Available at <http://www.cse.ucla.edu/products/Reports/R629.pdf>.
- Abedi, J., Bailey, A., Butler, F., Castellon-Wellington, M., Leon, S., and Mirocha, J. (2005). *The validity of administering large-scale content assessments to English language learners: An investigation from three perspectives*. Available at: <http://www.cse.ucla.edu/products/reports/r663.pdf>.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bailey, A. L. (Ed.). (2007). *The language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.
- Gottlieb, M., & Nguyen, D. (2007). *Assessment and accountability in language education programs: A guide for administrators and teachers*. Philadelphia, PA: Caslon Publishing.
- Hakuta, K., Butler, Y., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* (Policy Report 2000-1). University of California: Linguistic Minority Research Institute.
- Linn, R. L. (in press). Accountability system design. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in accountability testing. Yearbook of the national society for the study of education*, Vol. 104, Part I.
- Marion, S., & Pellegrino, J. (2006). A validity framework for evaluating the technical quality of alternate assessments. *Educational Measurement: Issues and Practice*, Winter 2006, 47-57.
- Wright, W. E. (2006). A catch-22 for language learners. *Educational Leadership*, 64(3), 22-27.

Note: more information on universal design can be found at www.cast.org.

Principle IV: Fair Appraisal of Academic Performance

Use multiple sources of evidence to describe and interpret school and district performance fairly, based on a balance of progress toward and success in meeting student academic learning targets.

Primary and Subordinate Recommendations

- 1. Encourage states and districts to use multiple sources of evidence drawn from their comprehensive and coherent systems of classroom-, school- and district-based assessments to evaluate student performance and inform actions.**
 - a. Provide models and technical assistance to help states collect and use the information from balanced comprehensive assessment systems to summarize and appraise student performance.
 - b. Provide resources to states to help them use information from the comprehensive assessment system in making high-stakes student decisions such as grade promotion and graduation.
 - c. Provide guidance for establishing uses and weights of multiple sources of evidence in a comprehensive system.
- 2. Encourage states to describe school performance in terms of status, improvement, and growth, using the states' multiple sources of evidence.**
 - a. Status is performance of students at a point in time in relation to learning targets (for example, established content and performance standards).
 - b. Improvement is change in performance of different groups of students over time (e.g., grade 3 one year compared to grade 3 the previous year).
 - c. Growth is progress over time of the same students toward the learning targets.
- 3. As states evaluate their assessment systems, conduct ongoing studies of the validity of the descriptions and interpretations of student and school performance to ensure the quality of core data analysis and reporting.**

Problems with the Law

Two major problem areas in the current law are its reliance on single measures (i.e., standardized tests) and its reliance on status measures only, rather than on a complementary mix of status, improvement, and growth measures. In addition, some states have inadequate data systems and procedures for evaluating the validity and consequences of their assessment and accountability systems.

Lack of multiple measures

NCLB includes language about the need for assessment information that serves multiple purposes, but it fails to acknowledge that these purposes cannot be fulfilled by the single mode of testing it advances – large-scale, statewide tests. Principle II argues for the increased use of locally developed assessments in ESEA. As important for making accountability decisions is the use of a wide range of assessments and sources of information about student learning. NCLB makes no provision for the use of multiple sources of evidence or multiple types or methods of assessment as part of a comprehensive and coherent system of evidence about students' learning that can be used to evaluate individuals, schools or systems.

Because the tests are administered annually, they cover an entire year's worth of content and typically contain only two or three questions per content area. It is impossible to get either a valid or reliable picture of student performance in relation to specific content standards with so few items. And as discussed in Principle II, reliance on the results of state tests as the sole measure of the accountability has resulted in the narrowing of curriculum and instruction.

In practice, NCLB has caused some states to make poor assessment – and consequently, instructional – decisions:

- Challenging assessments have been unintended casualties of NCLB. Some states have been reluctant to require additional assessments because results may increase the number of schools identified as not meeting adequate yearly progress (AYP). This is one major reason why the number of states administering writing assessments has dropped significantly since 2001 when NCLB was passed.
- Use of constructed response items (i.e., a non-multiple-choice test that requires a written or oral response) has emerged as one means to include several assessment methods within a test. Because they are more costly and take more time to develop and score, some states have dramatically reduced their use, even though such items are the most valid way within tests to assess whether students are able to generate answers, explain their reasoning, and engage in certain types of problem solving. However, the USDE has promoted more reliance on multiple-choice tests. As a result, today's typical state assessment consists of 50 questions, of which at least 45 are multiple-choice items, and the use of constructed-response items has diminished. Indeed, many states have only multiple-choice items for reading and math.

Another reason for the minimal use of constructed-response items is that NCLB requires that accountability reports be provided to districts and schools before the beginning of the following school year. Testing contractors cannot reliably process large numbers of constructed response items quickly. If the tests are administered late in the school year, as makes sense for accountability, then pressure develops to eliminate constructed-response items.

Obtaining results before the start of the next school year also is required for implementing the various stages of sanctions built into the law. Thus, the emphasis on imposing sanctions drives a narrowing of assessment methods with damaging consequences for curriculum, instruction, and student learning. The importance of the quick turnaround for accountability is also questionable. For example, the public school choice provisions require the quick turnaround. However, very few families have availed themselves of public school choice; fewer than two percent of eligible students have changed schools.

ESEA should shift from a focus on a limited set of sanctions to an effort to help all schools improve, reducing the imperative for fast turn-around of results on state tests.

With NCLB and the USDE focusing on scores on state tests for accountability determinations, there has been little financial or technical support to help states develop and implement assessment and evaluation systems that utilize multiple sources of evidence. Some states have made progress in this regard, including Nebraska with its state system of local assessments used for school accountability; and Rhode Island and Wyoming with their use of multiple forms of assessment in high school graduation decisions. But devising accountability systems that rely on multiple measures will require substantial financial support, particularly if those systems also seek to utilize growth and improvement, as well as status models.

Use of status measures only

Using the definitions of status, growth and improvement noted above, it is apparent that NCLB is constructed on a status model. Its goal is that all students score proficient on a state assessment by 2014. It is school and district standing in relation to that goal, measured by AYP: that counts the most. (We will address AYP and its consequences in more detail in principles V and VI.) The law made a modest provision for an improvement calculation in its “safe harbor” provision. Safe harbor means that if any subgroup fails to meet AYP provisions, the school or LEA will meet AYP if the percentage of students scoring below the proficient level declines by 10 percent compared to the same subgroup from the previous year. Safe harbor affects very few schools.

Lack of evaluation

Few states regularly evaluate their assessment systems. The evaluations that do occur tend to take fairly narrow views, focusing on a limited set of technical issues rather than also considering the broader questions of how well the assessments provide evidence on a rich set of desired learning outcomes and the impact of the assessments on curriculum, instruction and learning. While NCLB required a great increase in testing compared with the previous version of ESEA, and increased markedly the accountability consequences attendant on that testing, it provided no resources to evaluate assessment quality or the educational consequences of high-stakes accountability testing. Thus, evidence has accumulated in piecemeal fashion, though it has been sufficient to document weaknesses in NCLB as well as some harmful consequences of NCLB and high-stakes testing.

Improvements to the Law

ESEA will be improved dramatically if it advances an accountability system in which decisions about schools and districts are based on multiple sources of evidence documenting students’ learning. The development of state assessment systems that include local evidence, as described in Principles II and III, would enable the use of existing information in a cost-effective manner. Public confidence in the accuracy of accountability decisions would also increase as the frequency of incorrect conclusions about student achievement based on too little information decreases.

Other kinds of evidence, some of which are mentioned in the current law, include student grades, graduation/dropout rates, in-grade retention, percent of students taking honors/advanced classes and Advanced Placement exams, college enrollment rates, and employment histories after high school. Not all these sources would be appropriate in an accountability context, but nearly all can provide schools and districts with a more complete picture of students' learning and information to guide local improvement efforts.

We also believe that adopting a "multiple measures" approach would increase the likelihood that important longitudinal information about students would be aligned and examined. For example, aligning achievement data with college matriculation and relevant workforce information would allow for longitudinal analysis of student matriculation pre-K through college and into the workforce. Some states have these data, but the student identifiers in the data sets are not uniform or consistently present. Thus, important understandings of student performance trends and related opportunity structures are lost. States need to have a rich data system capable of collecting the data and individual student identifiers in place in order to analyze the data. Additionally, professional development and/or training for educators and administrators would be necessary in how to interpret the data and use it to modify instruction and curriculum offerings, and drive other school or student-based supports and interventions.

Using multiple sources of evidence of students' learning in a credible way also speaks to the diversity of our student population. For example, ELLs and SWDs require documentation of achievement and progress that is different from mainstream students due to their developmental and/or linguistic needs. (We discussed these students' assessment needs in more detail in Principle III.) But there also is enormous diversity within what appear to be homogeneous student populations. Therefore, educators need multiple forms of assessment and data sources so they can evaluate students appropriately and adjust instruction according to students' individual strengths, needs, interests, English language proficiency, and cognitive styles.

Effective means for gathering the various sources of evidence and issues of combining or weighting the sources of evidence in constructing composite indicators for use in evaluation and decision making are among the complex questions that require solutions if states are to construct fair, valid and helpful assessment and accountability systems. The forthcoming reauthorization of ESEA should provide substantial resources to help states make the move from reliance on single tests to use of multiple measures. The pilot projects recommended in Principle II would address the use and weighting of multiple measures, as well as the use of local assessments.

If the quick turnaround requirements did not exist, states could more readily employ richer assessments. For example, states could add constructed-response items and writing prompts and then provide results from the assessments on a more extended timeline. Alternatively, a state could choose to use a more limited set of measures in its large-scale assessment while ensuring that a variety of kinds of local evidence is included in the overall system. For many reasons, ESEA should shift from a focus on a limited set of sanctions to an effort to help all schools improve, reducing the imperative for fast turn-around of results on state tests.

School improvement requires meaningful and accurate evaluation. Effective evaluation and improvement steps depend on having adequate information. Because most evaluation and

improvement is done locally, multiple forms of locally-based information, supplemented with data from large-scale assessments, can best provide educators with sufficient evidence to guide improvement (as discussed in Principles V and VI). Educators also must learn how to use multiple sources of information for evaluation and improvement purposes; such professional development should be built into the school improvement processes outlined in ESEA.

Funds should be provided to enable states to make further progress in developing growth models and data systems to use in their accountability processes. Such financial support should include designing methods for using multiple sources of evidence within growth systems. (We will elaborate on the use of growth models in Principle V.)

Finally, the next authorization of ESEA should provide adequate funds to enable states to regularly evaluate their assessment systems. Evaluation of state assessment systems should include both internal, including an array of stakeholders, and external, independent judgments. It should be integrated into the development of more complex systems that incorporate classroom, local and state assessments, utilize multiple sources of evidence, and include status, improvement and growth models, all with a primary focus on gathering and using information to help improve student learning and school capacity to serve all children well.

References

- Abedi, J., Courtney, M., Mirocha, J., Leon, S., and Goldberg, J. (2005). *Language accommodations for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification*. Available at: <http://www.cse.ucla.edu/products/reports/r666.pdf>.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.) (2001) *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Note: Many of the references to previous principles also address issues raised in the discussion of this principle.

Principle V: Fair Accountability Decisions

Improve the validity and reliability of criteria used to classify the performance of schools and districts to ensure fair evaluations and to minimize bias in accountability decisions.

- 1. Encourage states to include all subjects – not just reading, math, and science – in their comprehensive assessment systems but use compensatory processes to ensure that inclusion of more subjects does not become another means for schools and districts to fail accountability requirements.**
- 2. Encourage states and districts to use multiple sources of evidence drawn from their comprehensive and coherent assessment systems to make accountability decisions about the quality of school and district performance and determine which schools and districts need what forms of assistance.**
- 3. Retain the ESEA requirement for gathering and reporting disaggregated information by subgroups based on the comprehensive assessment system.**
 - a. Students identified as ELL should be included in that subgroup for accountability purposes as long as that student remains in the same school or district in which they were identified; if the student remains ELL when moving to a new school or district, the inclusion pertains to the new school or district.
 - b. Students with disabilities whose disability makes it unlikely they will be able to meet the proficiency standard even with high quality instruction and appropriate accommodations may be able to take alternative or out-of-level (instructional level) assessments in order to accurately determine their achievement. Further research is necessary for the USDE to provide accurate guidelines on the allowable numbers of such students or criteria for the identification of students eligible for alternative assessment.
- 4. Use collective research from the states to establish realistic and challenging federal guidelines for rates of growth or improvement towards the goal of reaching learning targets.** Require each state to establish its rates within those guidelines.
 - a. Guidelines for rates of improvement should be supported empirically.
 - b. Guidelines for sufficient rates of improvement regarding graduation and grade promotion should be specified.
- 5. Replace the current rules for AYP classifications with reliability and validity criteria that each state must apply when designing its accountability classification system so that is fair and minimizes bias.** Federal reliability and validity criteria should address:
 - a. School and district resources and delivery/opportunity to learn standards (as discussed in Principle I);

- b. Use and weighting of multiple indicators (as discussed in Principle IV);
- c. Permissible ways to use status, improvement, and growth in constructing the system, which should involve some compensatory approaches;
- d. Methods for dealing with error and lack of reliability in school classifications.
- e. Provision of evidence that the classifications are valid, that schools are correctly classified, and that the classification is not a pass/fail system;
- f. Procedures by which subgroups are included in accountability decisions, including what indicators are considered; and
- g. How each state will ensure continuous improvement of its accountability system in terms of validity, reliability and educational consequences.

6. Use accountability decisions to inform assistance to schools.

Problems with the Law

Under previous principles, we have discussed the need for states to develop local and state assessment and accountability systems that rely on multiple sources of evidence. The absence of such systems today means that accountability decisions are made on the basis of too little information.

Positively, NCLB requires disaggregated data on key populations. While data should continue to be disaggregated, far more data than simply scores on a state test should be included in the disaggregated information presented to the public and used in accountability decisions. The nature of many of these required decisions is itself problematic in that they emphasize or open the door to punitive actions rather than to actions to ensure continuous improvement.

The key classification decisions within NCLB are tied to the determination of whether a school or district has made AYP. Experience with NCLB has made it clear that the goal of 100 percent student proficiency by 2013-14 is neither plausible nor realistic and that the law's AYP requirements are counterproductive to effective accountability and improvement.

The 100 percent student proficiency goal has generated a patchwork of proficiency targets that varies greatly from state to state, ranging from unrealistically high to embarrassingly low. Several states have even reduced their original definition of proficient student performance so they can at least come within striking distance of the goal that 100 percent of the students would meet this standard of performance by 2013-14. This is probably not what the law intended, but it is a predictable response from a system that is being pressured to deliver what it cannot. Moreover, numerous studies project that in most states all or almost all schools and districts will eventually fail to make AYP. Such gross classifications do not have educational value and are likely to have harmful consequences.

NCLB mistakenly bases judgments about schools' performance on single year aggregate summaries of performance (i.e., percent proficient). Inferences based on those results tend not to be valid. A school with a high proportion of students who are proficient is considered a "good" school, and a school with a low proportion of proficient students is not, but in both cases the law ignores potential confounding factors that affect performance and fails to recognize improvement or progress. For example, a student's score in the fourth grade is posited to be a result of that school's processes that occurred in fourth grade. However, possible confounding factors include such things as time (a student's fourth grade performance is also a function of his early learning) and language proficiency (results on a math assessment could be confounded with ability to read math text in English). Measuring student progress over time ameliorates some of these issues because performance changes take both time and improvement into account.

The law's reliance on annual performance targets over-emphasizes results that are linked to school enrollment characteristics and demographics, creating classifications of schools that are frequently unfair and not helpful. NCLB expects students who are behind to demonstrate faster growth in order to close existing achievement gaps, until all reach the "proficient" level. Thus, low-income or ELL students must show significantly faster achievement growth than their non-poor or non-ELL peers. Schools in poor school districts often fail to make AYP and are subject to sanctions, even when their students show very substantial increases in performance but do not reach annual proficiency targets. The schools get no credit for students who make substantial progress but do not reach the test cut-off score.

The law also creates a diversity penalty: Large schools with diverse student enrollments are disadvantaged since they can face sanctions if even one of their student subgroups fails to make AYP for two years, while academically comparable schools with less diversity have been shown to be less likely to fail. Thus, the presence of diverse student groups formerly viewed as assets and challenges to be met by the education system are now perceived as liabilities. In contrast, schools in wealthy districts with many high achievers and too few minority students to comply with the law's accountability requirements are often commended for their success, even though many students who score above the proficiency level may make little further progress.

AYP provisions also require schools to show gains for all subgroups of students in one or two years before the next stage of sanctions kicks in. This is an unreasonable timeframe since significant improvements in education that increase student achievement rarely show results that quickly. As a result, there is a strong incentive to increase scores on AYP-linked assessments through drill and practice related to the test content.

It becomes impossible to know whether score gains in particular schools are the result of real increases in learning or are merely inflated test results. Because such increases often are not real gains in knowledge and skill, they commonly do not show up on other tests of the same students, such as the National Assessment of Educational Progress (NAEP). While most state scores have indicated major gains in reading and math, NAEP grade four reading scores have not risen since 2002, while grade eight scores declined; and state scores in math have shown far larger gains than have NAEP state-level math results.

In all these instances, AYP misleads the public, yet it frequently but inappropriately commands the attention of educators. From a measurement perspective, AYP also raises concerns about the predictive validity of state assessments and proficiency levels. What, if anything, do they predict about students' future performance? For example, are students deemed "proficient" prepared to succeed in college or the workplace? Currently, their predictive validity is assumed but largely untested.

There are additional unintended consequences to the NCLB accountability structure. NCLB requires that states include results of math and reading tests in the accountability system. States can, if they choose, add other subjects. But if they do, the results on those assessments cannot be used to help a school make AYP. Rather, they can only become additional hurdles schools and districts must leap to avoid failure. Thus, states do not include assessments in other subjects in their AYP calculations, though a few plan to include results of the science tests.

Improvements to the Law

The reauthorized ESEA should set parameters within which states could design new accountability systems or modify their current one, and should provide resources to help states do so. States and districts can then generate decision rules that maximize the reliability and validity of classifications of schools, districts, and states and guide the most beneficial actions toward improving schools, systems and student learning.

ESEA would be improved dramatically if it increased the reliability of decisions about school performance by moving away from its exclusive reliance on annual aggregate summaries of performance that are based on results from a single end-of-year test. A better option would be to utilize multiple measures and a mix of status, improvement and growth evidence, thereby constructing a body of evidence to use in evaluation and decision-making.

States should be encouraged to add additional core subject areas to their accountability systems. In doing so, they should not be required to add additional state tests, but could rely on local evidence as discussed in Principle II. If they add subjects, they would be allowed to construct a method for including the various factors in the decision-making process without increasing the likelihood of facing sanctions.

Systems that utilize multiple sources of evidence from multiple subjects to make judgments based on growth or a combination of growth, improvement and status could operate in a variety of ways. We will briefly discuss two possibilities: the use of growth measures, and combining growth and status (or status improvement) measures. We emphasize we are discussing possible models to illustrate some of the options states could use to devise specific policies and programs within parameters set by ESEA.

First, the federal government could recommend rates of growth that states would follow in constructing growth-based systems. Such recommendations should be based on rates of improvement attained by a substantial portion of current Title I schools. This approach is substantially different from that implemented in the USED's "Growth Model Pilot," which has

approved a very small set of approaches that all inherit the basic flaws of the current NCLB accountability model.

For example, setting acceptable growth targets could involve the following steps:

- 1) Ascertain the average rate of increase over the previous three years in each Title I school in the state and rank the schools based on their rate of gain.
- 2) Based on this list, identify the school at the 75th percentile above the bottom (or the 50th percentile, or...).
- 3) The rate of gain by the school at that point shall become the target rate of increase for all Title I schools, while schools at or above this standard shall be expected to maintain or increase the rate of improvement. In this example, three-fourths (or one-half) of the Title I schools would have to increase their rate of improvement, while one-fourth (or one-half) would have to at least maintain their rate. Incentives could be established to encourage schools above the rate to further improve.

Further, a state could establish target rates of gain at multiple key points, e.g., basic, proficient, and advanced, across the proficiency continuum. In any case, when nearly all students in a school, and its subgroups of sufficient size, have attained proficiency, then that school and its subgroups might not be required to meet the rate of gain. Note that if state scores have recently inflated, it is likely the rate of growth shall soon stall, based on historical evidence of trends in score increases. In that case, established rates of growth may produce unrealistically optimistic projections.

An accountability model that incorporates growth would operate differently than the growth model pilot project announced by the USDE in November 2005. Under this pilot, states must implement growth models within the confines of NCLB's requirement that states must ensure that students are proficient by 2013-14. Such a requirement essentially means that rates of growth must average out to the current impossible expectations. As a result, few schools would show different results than they would with the original status model. Further, growth should continue to be part of the accountability system after 2013-2014 as many schools will continue to enroll less prepared, at-risk, students who will require some time to close achievement gaps.

Second, growth and status could be combined. For example, performance levels (status or improvement in status) could be emphasized at key grade levels, while growth could be emphasized in each and every grade. From a technical standpoint it is relatively straight-forward to create a system at the school level that incorporates growth and proficiency targets. A school district/building/state, based upon where they begin, would be expected to improve annually in reasonable increments for all subgroups, with each subgroup mirroring the upward trend.

In this model, schools would be expected to meet different annual performance targets based on their specific student enrollment. Growth targets would be conditioned on students' initial academic status. In this way growth estimates could account for variability in student baselines, and school growth rates could be adjusted accordingly. The measures that are used to assess students' initial skills could be locally- or state- developed, provided they are reliable and valid

and are administered and analyzed in ways that establish defensible baselines from which to calculate growth.

Under either model, a reasonable time period, such as three to five years, should be set as the period in which schools that are not at the set rate of gain would be expected to attain that rate. If a school has not done so by the end of this period, then districts or states would be required to intervene to move the school onto a positive growth track.

Incorporating growth into ESEA accountability is not without challenges. Growth targets are not in and of themselves necessarily egalitarian. For example, a school in a gateway community that consistently receives children of immigrants from economically depressed areas, or one with high mobility rates, could be unfairly burdened with greater growth requirements than another school. Fixed performance targets and expectations for student growth would need to be established in ways that are equitable, fair, and realistic. Targets should be empirically studied to ensure they are generally attainable and adjusted if they are too easy or too difficult.

Decisions that impact schools involve more than growth and status data, even when those are based on multiple sources of evidence. Information about the school's overall context must be gathered if decisions regarding assistance or sanctions are to be fair and educationally beneficial. Schools also should not face punitive sanctions for failing to succeed at tasks for which the federal government has not provided adequate resources, and they must be given reasonable time to adapt to changing circumstances (as discussed in Principle I).

We have called for the use of multiple sources of evidence and incorporating local assessment information in making decisions (Principles II, III, and IV) as well as for the use of growth measures. It will take ingenuity and persistence to design systems in which multiple forms of information from different sources combined into data that can be analyzed to fairly and accurately evaluate growth in student learning across all important dimensions of learning and at multiple levels of proficiency. A reauthorized ESEA should provide resources to help states and districts develop and implement such systems, as part of the pilot programs recommended in Principle II.

For categorizing or for making decisions, multiple sources must be combined using a formula or matrix with each component weighted. The federal government should establish guidelines states would use in constructing their formulas. For example, weights could be given to a statewide assessment and local assessments in core subject areas, and other data including graduation rates, grade promotion rates, attendance and similar data on student outcomes. Assessment data should contribute the largest share, combined between local and state assessments. Other measures, such as graduation rates, grade progression rates, or attendance, would contribute the balance. Measures that are not directly academic, such as improvements in the learning climate, that are known contributors to positive academic outcomes and that can be measured in ways that can demonstrate improvement, could also be included within "other measures."

It is likely to be quite challenging to incorporate the particular needs and challenges of SWDs and ELLs into sophisticated accountability decision systems. Clearly, such students must be

included in assessment systems, and the assessments used with such students must be appropriate for them (Principle III).

In general, disaggregated results should be reported for the various student population subgroups. There should be interventions if schools or districts do not meet the needs of these students. But the accountability problem is that it is not educationally wise to impose one-size-fits-all consequences given the vast differences in school contexts. Accountability decisions should factor in which subgroups do not make sufficient progress, how many subgroups, and by how much they fail, as well as the resources available to schools.

For any subgroup that does not make sufficient progress, a school, district or state should be required to establish methods for meeting that subgroup's needs. Technical assistance and interventions should be tailored to the needs of specific subgroups. Schools with multiple subgroups that are not making sufficient progress will need more comprehensive improvement efforts and support. (See, for example, the recommendations in *Redefining Accountability: Improving Student Learning by Building Capacity* from the Forum on Educational Accountability.)

A critical question is the comparative rates at which subgroups should be expected to make progress. NCLB currently defines closure of the achievement gap when all students reach the “proficient” level. The models proposed above continue to require that schools and districts move all students toward proficiency, hence closing the gap.

Gaps will only close if students who are behind progress more quickly than those who are not. There is a lack of evidence to demonstrate schools alone can ensure that their historically disadvantaged populations can progress more quickly than more advantaged populations. Expecting schools to accomplish this feat without markedly increased support is likely to continue the NCLB problem of causing harmful educational consequences resulting from educators' desperate attempts to meet NCLB mandates without the resources to do so.

Congress should fund extensive research and evaluation to develop knowledge on how best to accelerate the academic learning of the various subgroups. When ways to do so are established, then governments can consider ways to include such progress in its accountability structures.

In addition, more careful consideration must be given to subgroups of students who need additional time and support to meet the performance targets. For example, experts say it takes five to seven years for ELLs to become academically literate in English. They should be given the time needed to meet this goal, rather than be expected to become so in a few years. Further, much more evidence is needed about how the many different disabilities students have affect learning and what therefore are reasonable expectations for student progress given the nature and severity of their disabilities. Neither the one-percent nor the two-percent cutoffs for classifying SWDs has an adequate research base.

References

- Goldschmidt, P., Roschewski, P., Choi, K., Auty, W., Hebbler, S, Blank, R., and Williams, A. (2005). *Policymakers guide to growth models for school accountability: How do accountability models differ?* The Council of Chief State School Officers, Washington, DC.
- Choi, K, Goldschmidt, P. and Yamashiro, K. (2005). Exploring models of school performance: From theory to practice, *National Society for the Study of Evaluation*, v.104.
- Forum on Educational Accountability. (2007). *Redefining accountability: Improving student learning by building capacity*. Available at:
<http://www.edaccountability.org/CapacityBldg02-15-07.pdf>
- Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends*. Available at:
http://www.civilrightsproject.harvard.edu/research/esea/nclb_naep_lee.pdf.

Principle VI: Use of Assessment and Accountability Information to Improve Schools and Student Learning

Provide effective, targeted assistance to schools correctly identified as needing assistance.

- 1. Encourage states and districts to use multiple sources of evidence from state and local assessments and other forms of evidence to inform actions such as interventions and technical assistance.**
- 2. If a school (or district) is identified as not making sufficient progress toward improvements or in outcomes, the district (or state) would investigate causes and undertake a series of interventions tailored to address particular needs.** The district (or state) would:
 - a. Gather information to provide a deeper understanding of strengths and needs from multiple sources of data, such as school visits, self and external program evaluations, etc.
 - b. Consider school resources and student opportunity to learn (as addressed in Principle I).
 - c. Consider differential assistance for schools' status and growth performance. For example, a school may need different assistance if it is high status and low growth versus low status and high growth or low status and low growth.
 - d. Tailor assistance and interventions to specific needs of the school and of any subgroups that are not doing well. Focus both on steps towards improving schools and addressing immediate student needs.
 - e. Allow sufficient time for changes to take hold and be effective before taking any additional steps.
- 3. Assistance may include professional development, developing partnerships with parents and families, improving curriculum, and attracting and retaining high quality teachers and administrators.**
- 4. NCLB mandates for governance changes should be removed or, at most, made an option for possible action only after implementation of recommendations 2 and 3 (above).**
- 5. Accountability systems should ultimately be judged on their consequences for the quality of the educational system and the learning of its students.**

Problems with the Law

Under NCLB, decisions guiding accountability actions are made using too little information. The decision-making and action processes lack validity. Evidence about inputs to schools and about

school processes is only included if a district or state happens to include it when deciding what corrective actions to take. Interventions are often one-size-fits-all, though there are instances where states or districts make more careful evaluations of schools that are in the later stages of corrective action and then target actions to specific needs. Federal law should expect evaluations and actions to incorporate inputs, school processes, and multiple sources of evidence of student learning gathered over time, and for interventions to be targeted to specified needs.

An example of the problem is that all schools that fail to make AYP for one or more student subgroups face, in current law, the same escalating sanctions. Thus, the law treats a school in which one subgroup of students does not make AYP the same as a school in which multiple subgroups, and the school as a whole, do not make AYP.

The NCLB accountability process also fails to take into account that school improvement efforts take time and often do not show results for a few years. Under current law, new sanctions are very likely to take effect before previous actions have had a chance to take hold and demonstrate their effectiveness.

The various sanctions employed in the law are not based on evidence that they are reasonable solutions to often difficult problems. There is evidence that the changes in school governance specified in the law are often not effective: charter schools appear to do about as well as other schools; private management companies have been removed from many districts for failure to improve schools; reconstituted schools often do not do better, unless the student body itself has been replaced; and state takeovers have not produced improved learning outcomes. Wisely, many states are using the "other action" option the law provides. Even then, the presumption that governance changes are what is needed to improve schools is unproven.

The sanctions provisions need to be changed. The law should establish means and resources to strengthen school and district capacity to engage in continuous improvement. As the "Joint Organizational Statement on NCLB" put it, "*Overall, the law's emphasis needs to shift from applying sanctions for failing to raise test scores to holding states and localities accountable for making the systemic changes that improve student achievement*" (emphasis in original). This injunction by no means eliminates the requirements to assess or to hold schools accountable, but it significantly alters the context for assessment and the conceptualization of accountability.

Improvements to the Law

The use of multiple sources of evidence is a sound procedure in assessment, evaluation, classification and accountability decisions, and in interventions to improve educational practices and outcomes. It is essential that each step in the process – assessment, evaluation, classification and actions – be organized to maximize improvement and be subject to continuing scrutiny. The needs of the various subgroups of students must be considered at each step, and data systems must be developed that can meet the complex needs of diverse learners in diverse learning situations.

The law should provide states and districts with resources so they can carefully examine the assessments on which they intend to base their decisions. If the assessments measure what they

purport to measure and provide consistent results, states and districts can then consider the effect of their decision rules. That is, they can determine whether valid inferences based on assessment results relate to established standards for all students, with particular attention to specified subgroups. Systems also need to be careful so that changes in student characteristics do not affect school performance indicators, something they can guard against by gathering and incorporating demographic contextual factors into their accountability systems.

Once multiple sources of evidence from classroom, local and state assessments gathered over time are considered from status, improvement and growth perspectives, decisions can be made more reliably, validly and fairly. Any ensuing interventions can then be better tailored to the actual problems. Actions taken should be part of a continuous improvement effort. Assessments can serve a powerful role in the improvement process, both in providing rich evidence on which improvement efforts can be based and in the actual teaching and learning process.

In closing, we emphasize that the major purpose of assessment and accountability is improving significant learning outcomes and thus improving schools and districts and the professionals who work in them. As we point out in Principle I, resources and opportunity to learn are essentials that are too often in short supply. The accountability structures in NCLB fail to provide means for consistent and continuous improvement in the education of the whole child. Our hope is that the redesigned systems of accountability – assessment, evaluation, decision-making and action – will contribute to better and fairer results.

References

References for other Principles provided background information on this Principle.

Members of the Expert Panel on Assessment

Jamal Abedi, Ph.D., is a Professor at the Graduate School of Education of University of California, Davis and a research partner at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Dr. Abedi's research interests include studies in the area of psychometrics and test and scale development. Among his recent work are studies on the assessment and accommodations for English language learners (ELLs) and research on the opportunity to learn for ELLs. Results of his recent studies on the impact of linguistic factors on the assessment and accommodation for ELLs have been used and reported nationwide. Abedi is the recipient of the 2003 national Professional Service Award in recognition of his "*Outstanding Contribution Relating Research to Practice*" by the American Educational Research Association. Abedi's educational background is in psychometrics and research methodology.

Pete Goldschmidt, Ph.D., is an assistant professor in the Michael D. Eisner College of Education at the California State University Northridge and a Senior Researcher at the Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA. He is currently serving as a member of the NCLB Growth Pilot peer review panel as well as a member on the standing IES (Institute for Education Sciences) Read/Write grant proposal review committee. Dr. Goldschmidt is also advising the Chilean Ministry of Education in development of its growth/value added model for accountability.

Brian Gong, Ph.D., is Executive Director of the National Center for the Improvement of Educational Assessment, Inc., a non-profit organization dedicated to helping foster student achievement through improved assessment and accountability policies. Gong and his Center colleagues have worked with over half the states on designing and implementing valid and useful assessment and accountability systems, both prior to and subsequent to the passage of NCLB. Gong has worked extensively in accountability, including recently on designs of growth models for school accountability, validity frameworks for accountability, quality assurance for use of local assessment information in accountability, and procedures for ensuring meaningful alignment of alternate assessments with standards.

Margo Gottlieb, Ph.D., specializes in the design of assessment systems for English language learners in pre-K-12 settings, the evaluation of language educational programs, and the development of English language proficiency standards. Having been a teacher and administrator, Margo has provided technical assistance to governments, states, school districts, publishers, universities, and professional organizations nationally and internationally. She is Lead Developer for World-Class Instructional Design and Assessment (WIDA), a consortium of multiple states, at the Wisconsin Center for Education Research, Madison. She also serves as the Director, Assessment and Evaluation, for the Illinois Resource Center, Arlington Heights.

Alba A. Ortiz, Ph.D., is a Professor in the Department of Special Education, and Director of the Office of Bilingual Education, in the College of Education at The University of Texas at Austin. Her research and teaching focus on English language learners with language and learning disabilities, and prevention and early intervention for second language learners experiencing

achievement difficulties. At UT Austin, she is the holder of the President's Chair for Education Academic Excellence, an honor bestowed in recognition of her research, teaching, and service contributions in the fields of bilingual education and special education. A past president of the International Council for Exceptional Children, Dr. Ortiz is the co-chair of Exceptional Needs Standards Committee for the National Board for Professional Teaching Standards and a member of the National Institute for Child Health and Human Development Biobehavioral and Behavioral Sciences Subcommittee.

Pedro Pedraza, is an educational researcher at the Centro de Estudios Puertorriquenos, Hunter College, City University of New York, working in the area of K-12 public education. He is the founder and director of the National Latino/Education Research Agenda project and recently edited (with Melissa Rivera) *Latino Education: An Agenda for Community Action Research* (Mahwah, N.J. Lawrence Erlbaum Associates, 2005). He is currently investigating the role of arts in education in a successful alternative high school in Brooklyn, N.Y., El Puente's Academy for Peace and Justice, with funding from the Ford Foundation.

James Pellegrino, Ph.D., is Distinguished Professor of Psychology and Education at the University of Illinois at Chicago where he also Co-Directs the Learning Sciences Research Institute. His scholarly work focuses on the application of theory and research on learning to design highly effective instructional and assessment practices that support student achievement in areas such as mathematics and science. He is a member of the National Research Council's Board on Testing and Assessment and has directed the production of highly visible NRC reports, including those produced by the Committee on the Evaluation of National and State NAEP, the Committee on Learning Research and Educational Practice, and the Committee on the Foundations of Assessment. He serves on the Technical Advisory Committees for the Kansas and Wyoming state NCLB assessments. He was recently elected a member of the National Academy of Education.

Pat Roschewski, Ph.D., Director of Statewide Assessment in Nebraska, has worked for the last fifteen years in the building of district and classroom-based local assessment systems. She worked for 22 years as a teacher and served the same school district as director of curriculum and assessment. Since 2000, she has been working as Director of Statewide Assessment for the Nebraska Department of Education in collaboration with the University of Nebraska and has worked with school districts of many sizes in building local systems for standards, assessment, and accountability. Pat has worked with state and federal policy as the requirements of NCLB have been integrated into the Nebraska system. She has authored several articles about the Nebraska STARS process and has contributed to the growing body of classroom assessment research.

Jim Stack, Ed.D. is the former Director of Achievement Assessments for the San Francisco Unified School District. He received his doctorate from the University of the Pacific, Stockton, California, in the area of Curriculum and Instruction with a specialization in Bilingual Education. He has experience as a Bilingual teacher, a Bilingual Curriculum specialist, and a Bilingual Program Evaluator. Dr. Stack was the 2003 President of the California Educational Research Association. He has presented to and worked with teacher and administrator organizations both

nationally and internationally in the areas of instruction and assessment of English language learners.

Joint Organizational Statement on *No Child Left Behind (NCLB) Act*

List of signers updated June 5, 2007

The undersigned education, civil rights, religious, children's, disability, civic and labor organizations are committed to the No Child Left Behind Act's objectives of strong academic achievement for all children and closing the achievement gap. We believe that the federal government has a critical role to play in attaining these goals. We endorse the use of an accountability system that helps ensure all children, including children of color, from low-income families, with disabilities, and of limited English proficiency, are prepared to be successful, participating members of our democracy.

While we all have different positions on various aspects of the law, based on concerns raised during the implementation of NCLB, we believe the following significant, constructive corrections are among those necessary to make the Act fair and effective. Among these concerns are: over-emphasizing standardized testing, narrowing curriculum and instruction to focus on test preparation rather than richer academic learning; over-identifying schools in need of improvement; using sanctions that do not help improve schools; inappropriately excluding low-scoring children in order to boost test results; and inadequate funding. *Overall, the law's emphasis needs to shift from applying sanctions for failing to raise test scores to holding states and localities accountable for making the systemic changes that improve student achievement.*

Recommended Changes in NCLB

Progress Measurement

- 1. Replace the law's arbitrary proficiency targets with ambitious achievement targets based on rates of success actually achieved by the most effective public schools.**
- 2. Allow states to measure progress by using students' growth in achievement as well as their performance in relation to pre-determined levels of academic proficiency.**
- 3. Ensure that states and school districts regularly report to the government and the public their progress in implementing systemic changes to enhance educator, family, and community capacity to improve student learning.**
- 4. Provide a comprehensive picture of students' and schools' performance by moving from an overwhelming reliance on standardized tests to using multiple indicators of student achievement in addition to these tests.**
- 5. Fund research and development of more effective accountability systems that better meet the goal of high academic achievement for all children.**

Assessments

6. **Help states develop assessment systems that include district and school-based measures in order to provide better, more timely information about student learning.**
7. **Strengthen enforcement of NCLB provisions requiring that assessments must:**
 - **Be aligned with state content and achievement standards;**
 - **Be used for purposes for which they are valid and reliable;**
 - **Be consistent with nationally recognized professional and technical standards;**
 - **Be of adequate technical quality for each purpose required under the Act;**
 - **Provide multiple, up-to-date measures of student performance including measures that assess higher order thinking skills and understanding; and**
 - **Provide useful diagnostic information to improve teaching and learning.**
10. **Decrease the testing burden on states, schools and districts by allowing states to assess students annually in selected grades in elementary, middle schools, and high schools.**

Building Capacity

9. **Ensure changes in teacher and administrator preparation and continuing professional development that research evidence and experience indicate improve educational quality and student achievement.**
10. **Enhance state and local capacity to effectively implement the comprehensive changes required to increase the knowledge and skills of administrators, teachers, families, and communities to support high student achievement.**

Sanctions

11. **Ensure that improvement plans are allowed sufficient time to take hold before applying sanctions; sanctions should not be applied if they undermine existing effective reform efforts.**
12. **Replace sanctions that do not have a consistent record of success with interventions that enable schools to make changes that result in improved student achievement.**

Funding

13. **Raise authorized levels of NCLB funding to cover a substantial percentage of the costs that states and districts will incur to carry out these recommendations, and fully fund the law at those levels without reducing expenditures for other education programs.**
14. **Fully fund Title I to ensure that 100 percent of eligible children are served.**

We, the undersigned, will work for the adoption of these recommendations as central structural changes needed to NCLB at the same time that we advance our individual organization's proposals.

1. Advancement Project
2. American Association of School Administrators
3. American Association of School Librarians (AASL), a division of the American Library Association (ALA)
4. American Association of School Personnel Administrators (AASPA)
5. American Association of University Women
6. American Baptist Women's Ministries
7. American Civil Liberties Union
8. American Counseling Association
9. American Dance Therapy Association
10. American Federation of Labor – Congress of Industrial Organizations (AFL-CIO)
11. American Federation of School Administrators (AFSA)
12. American Federation of State, County and Municipal Employees (AFSCME)
13. American Federation of Teachers (AFT)
14. American Friends Service Committee (AFSC)
15. American Humanist Association
16. American Music Therapy Association
17. American Occupational Therapy Association
18. American Speech-Language-Hearing Association
19. American School Counselor Association
20. Americans for the Arts
21. Annenberg Institute for School Reform
22. Asian American Legal Defense and Education Fund
23. Asian Pacific American Labor Alliance
24. ASPIRA
25. Association for Supervision and Curriculum Development
26. Association of Community Organizations for Reform Now (ACORN)
27. Association of Education Publishers
28. Association of School Business Officials International (ASBO)
29. Association of Teacher Educators
30. Big Picture Company
31. Business and Professional Women/USA
32. Center for Community Change
33. Center for Expansion of Language and Thinking
34. Center for Parent Leadership
35. Center for Policy Alternatives
36. Change to Win
37. Children's Aid Society
38. Children's Defense Fund
39. Church Women United
40. Coalition for Community Schools
41. Citizens for Effective Schools

42. Council of Administrators of Special Education, Inc.
43. Coalition of Essential Schools
44. Commission on Social Action of Reform Judaism
45. Communities for Quality Education
46. Council for Children with Behavioral Disorders
47. Council for Exceptional Children
48. Council for Hispanic Ministries of the United Church of Christ
49. Council for Learning Disabilities
50. Cross City Campaign for Urban School Reform
51. Disciples Home Missions of the Christian Church (Disciples of Christ)
52. Disciples Justice Action Network (Disciples of Christ)
53. Division for Learning Disabilities of the Council for Exceptional Children (DLD/CEC)
54. Education Action!
55. Educate Not Incarcerate
56. Episcopal Church
57. Evangelical Lutheran Church in America
58. Every Child Matters
59. FairTest: The National Center for Fair & Open Testing
60. Forum for Education and Democracy
61. Gay, Lesbian, Straight Education Network (GLSEN)
62. Gender Public Advocacy Coalition
63. Hmong National Development
64. Holmes Partnership
65. Indigenous Women's Network
66. Institute for Language and Education Policy
67. International Reading Association
68. International Technology Education Association
69. Japanese American Citizens League
70. Learning Disabilities Association of America
71. League of United Latin American Citizens (LULAC)
72. Mental Health America
73. Ministers for Racial, Social and Economic justice of the United Church or Christ
74. National Association for the Advancement of Colored People (NAACP)
75. NAACP Legal Defense and Education Fund (LDF)
76. National Alliance of Black School Educators
77. National Association for Asian and Pacific American Education (NAAPAE)
78. National Association for Bilingual Education (NABE)
79. National Association for the Education and Advancement of Cambodian, Laotian and Vietnamese Americans (NAFEA)
80. National Association for the Education of African American Children with Learning Disabilities (NAEAACLD)
81. National Association of Federally Impacted Schools (NAFIS)
82. National Association of Pupil Service Administrators
83. National Association of School Nurses
84. National Association of School Psychologists
85. National Association of Secondary School Principals (NASSP)

86. National Association of Social Workers
87. National Baptist Convention, USA (NBCUSA)
88. National Coalition of ESEA Title I Parents
89. National Coalition for Asian Pacific American Community Development
90. National Coalition for Parent Involvement in Education (NCPIE)
91. National Conference of Black Mayors
92. National Council for Community and Education Partnerships (NCCEP)
93. National Council for the Social Studies
94. National Council of Churches
95. National Council of Jewish Women
96. National Council of Teachers of English
97. National Education Association
98. National Education Taskforce
99. National Federation of Filipino American Associations
100. National Indian Education Association
101. National Indian School Board Association
102. National Korean American Service & Education Consortium (NAKASEC)
103. National Ministries, American Baptist Churches USA
104. National Pacific Islander Educator Network
105. National Parent Teacher Association (PTA)
106. National Reading Conference
107. National Rural Education Association
108. National School Boards Association
109. National School Supply and Equipment Association
110. National Science Teachers Association
111. National Superintendents Roundtable
112. National Urban League
113. Native Hawaiian Education Association
114. Network of Spiritual Progressives
115. Organization of Chinese Americans
116. People for the American Way
117. Presbyterian Church (USA)
118. Progressive National Baptist Convention
119. Protestants for the Common Good
120. Public Education Network
121. Rural School and Community Trust
122. Service Employees International Union
123. School Social Work Association of America
124. Social Action Committee of the Congress of Secular Jewish Organizations
125. Sikh American Legal Defense and Education Fund
126. Southeast Asia Resource Action Center (SEARAC)
127. Stand for Children
128. Teachers of English to Speakers of Other Languages, Inc. (TESOL)
129. United Black Christians of the United Church of Christ
130. United Church of Christ Justice and Witness Ministries
131. United Methodist Church, General Board of Church and Society

132. USAction
133. Women's Division of the General Board of Global Ministries, TheUnited Methodist Church
134. Women of Reform Judaism

For an updated list of signers go to: www.edaccountability.org