



NASP Position Statement on Using Large Scale Assessment for High Stakes Decisions

A basic premise of standards-based reform is that all children can learn. The National Association of School Psychologists (NASP) recognizes that, although some students may require more time or varied instruction, all students should be provided the opportunity to reach a set of high educational standards. In recent years, federal legislation (such as the No Child Left Behind [NCLB] Act [ESEA] of 2002; IDEA) and civil rights statutes (such as the Americans with Disabilities Act and Section 504) have shaped standards-based reform. These mandates, particularly the No Child Left Behind Act, have prompted the development of state and district "large scale assessments"-testing all students in a given population to measure attainment of uniform, high academic standards. NASP supports intended positive consequences of these efforts, including improved teaching and instruction, higher achievement for all students, higher standards for students who have struggled to overcome low expectations and increased access to the general education curriculum for all students. In particular, NASP notes that students with disabilities, students from disadvantaged backgrounds and students who do not speak English as a first language have struggled to overcome low educational expectations for some time.

However, despite good intentions as well as documented positive results in some settings, NASP is concerned by the potential for unintended negative outcomes of large scale assessment applied to both systemic and individual student decisions. NASP urges caution in the use of large scale assessments for high stakes decision-making at all levels, from school to district to state. Furthermore, NASP strongly opposes the use of large-scale testing as the *sole determinant* for making critical, high stakes decisions about individual students and educational systems, including access to educational opportunity, retention or promotion, graduation or receipt of a diploma.

Concerns and Cautions Regarding Large Scale Assessment

Recognizing multiple purposes of large scale assessment. All states are required to establish large-scale assessment programs to measure student progress toward standards. However, different stakeholders want assessments to meet a variety of needs--educators want test results to inform instruction; taxpayers want to know that the money they spend translates into student learning; governors want assurances that their students are achieving at a level similar to or better than students in other states.

Yet, NASP and other experts acknowledge that tests should be designed for the specific *purpose* that they are intended to serve and for the *population* that they will measure (National Research Council, 1999). While some states have tried to meet the demands for accountability by modifying existing large scale assessments or

Please note that NASP periodically revises its Position Statements. We encourage you to check the NASP website at www.nasponline.org to ensure that you have the most current version of this Position Statement.

developing new tests, many other states continue to use *single* tests for multiple purposes of system accountability, school improvement, and measurement of individual student or group performance, regardless of their intended use and inherent limitations.

High stakes and negative consequences for systems and individuals. Tests are considered *high stakes* for students when the results are used to make critical decisions about the individual's access to educational opportunity, grade-level retention or promotion, graduation from high school, or receipt of a standard or alternative diploma. These kinds of decisions all have immediate as well as long-range impact on the student. In some states, high stakes also are attached to test results for school systems--teachers, administrators, and schools are rewarded or sanctioned based on student performance. NASP recognizes that, when high stakes are attached to test scores, there is greater potential for misuse of data and negative consequences:

1) Use of a single test score in making promotion/retention decisions. NASP and test development experts agree that it is not appropriate to use performance on a single test (or composite test battery) for making high-stakes decisions for individuals. Yet, increasingly, states are requiring schools and school districts to use state test scores to determine whether students should be promoted to the next grade level, resulting in higher numbers of retained students each year. Extensive research over many years indicates that student achievement rarely improves when repeating a grade and, further, demonstrates a strong relationship between retention and increased dropout rates.

2) Use of a single test score in graduation decisions. Many states have adopted exit exams for high school graduation, in some cases resulting in the denial of a diploma to thousands of students based on a single test, without regard to their classroom performance, teachers' recommendations, or access to adequate classroom resources, quality instruction, or pupil services support. Although states may allow students to take these tests several times, multiple administrations of the same type of measure do not improve the reliability of the scores or reduce the general limitations of such testing.

3) Use of test performance as a basis for systems level rewards and sanctions. There is strong political support for the use of assessment results for *system* accountability, as reflected in the new provisions of NCLB. Administrators and teachers are rewarded or sanctioned based on student test performance, despite having little or no influence on some factors that significantly impact student achievement, such as student mobility and parent involvement. In some schools, these consequences could negatively affect instruction for all students, including students with disabilities, by dramatically narrowing the curriculum to emphasize test content and encouraging the use of generally inappropriate "quick fix" approaches to student learning.

4) Impact on mainstream education. High stakes testing programs can also have unintended but negative effects on the education provided to *all* students by narrowing the curriculum and unduly emphasizing basic skills to the exclusion of the arts, technology, sciences and humanities; creating a culture of "teach-to-the-test"; increasing the psychological stress on children and families; and decreasing teacher job satisfaction. Further, schools may focus limited resources on efforts to directly improve *test scores*, rather than on strategies to improve *school climate* and student *learning*. Tests should inform instruction, not dictate what is taught.

5) Impact on referrals to special education. Some schools respond to increased pressure from accountability associated with high stakes testing by increasing the number of children they identify as needing special education supports. Systematic methods of collecting data on special education referrals and placement are critical in order to accurately monitor this trend across time and make comparisons within and among schools from year to year.

6) Impact on student mental health. When "failing" the test means failing the grade, failing to graduate, or even lesser consequences such as attending summer school or loss of certain privileges, students may experience long-term anxiety, low self esteem, depression, etc. At a more systemic level, class-wide and building-wide testing can put students, teachers and administrators at risk for anxiety and other forms of emotional distress. These consequences can impact not only test-taking but also learning and motivation.

Interpreting Results From Large Scale Assessments: Cautions and Considerations

NASP strongly urges districts and states to take great care when applying test results from large-scale assessments to high stakes decisions such as graduation, retention, merit pay, etc. School psychologists have expertise in assessment, and can play a key role in helping others to appropriately interpret and use results from large scale assessments. Factors that influence the accurate interpretation of standards test results including the following:

Who is assessed? There may be inconsistency in the groups of students included in the state assessment reports over time. For example, when students are retained or drop out, the group of students included in testing changes. Further, some states and districts continue to exclude some students with disabilities and/or limited English proficiency from their assessment systems, in violation of Civil Rights statutes. New mandates and funding incentives may further pressure states to exclude groups of students who might tend to score below standards or require extensive accommodations. Additionally, due to high student mobility in some areas, many students tested in one school in a given year may have received much of their instruction elsewhere. Measuring effectiveness of instruction across schools or over time is severely compromised with highly mobile populations.

What tests are used and what do they measure? Assessment programs vary in many ways across states, as some states compare individual student performance to a national group, while others compare individual student performance to established performance standards. Further, states differ in the content measured and how proficiency is defined and demonstrated. For example, some states may use "minimum standards" while others use "high standards." Although trends within states are more reliable for comparison than cross-state trends, even comparisons within a given state must be reported carefully to assure similar data and standards are used.

What accommodations were provided? States have different rules about the kinds of accommodations that can and cannot be used for students with disabilities and students with limited English proficiency. It is important to know not only that students were given appropriate accommodations, but also the kinds of accommodations given, how reliably these accommodations were implemented, and if accommodations were provided across all testing situations. The interpretation of test data may be unreliable when accommodation practices are inconsistent.

How are test results used? While following recognized standards for test development and standardization will help to assure reliable and valid results, administrators and other school personnel should exercise extreme caution when applying results of large scale assessments to decision making about individual students. Myriad factors can impact the performance of any one student at a single point in time, significantly reducing the reliability of test scores. Therefore decisions regarding the promotion, graduation, placement or referral of individual students should be based on multiple sources of individually obtained data rather than the results of large scale assessment.

Recommended Guidelines for Large Scale Assessments

School psychologists must take an active role in promoting the appropriate use of large scale assessments and test results. The National Association of School Psychologists recommends the following principles to guide large scale assessment programs at the school, district, state and federal levels:

Including All Students

- To support the necessary inclusion of all students in standards-based assessment programs, schools must appropriately implement accommodations, modifications, or alternate assessments when necessary, as determined by the Special Education Team. In addition, all reports should include the results for all students tested, clearly identifying which students are included in each data set while protecting the personal identities of individuals. Further, interpretation of student progress should be based upon high but realistic expectations. While all students can learn, rates and styles of learning differ widely even within the "normal" range of ability.

- NASP urges careful consideration when interpreting the results of large scale assessments for individuals or groups of students with disabilities or limited English proficiency, as these tests may not adequately reflect the content or level of their instruction or address realistic instructional goals. Other strategies, such as curriculum-based measurement and progress monitoring of skill performance, should be used in combination with large scale assessments to reliably and validly measure progress in the general education curriculum for these students.

Decision Making for Individual Students

- NASP opposes the use of performance on a single test or standards test battery as the sole determinant in any decision about promotion, retention, instructional placement, scholarship or graduation for any individual student. Rather, NASP supports the use of multiple measures of academic achievement, including grades, curriculum-based procedures and teacher evaluations, as well as parent input, in making such decisions.
- If tests are used as components of high stakes decisions such as graduation, NASP urges consideration of options such as phasing in testing requirements over time, providing opportunities for retesting, or developing procedures for appeals and waivers, as well as other sources of information about student performance.
- Consequences of high stakes testing for individual students should not be posed as either-or choices, but as indicators of the need for and implementation of educational strategies such as early intervention, programmatic changes, or specific evaluation of learning problems.
- If an assessment is to be used as part of a high stakes decision (e.g., promotion, graduation), the test content must be aligned to curriculum and instruction. Students must have adequate opportunity to learn the material covered by the test.

Test Design and Selection

- NASP asserts that tests must meet professional standards of technical adequacy and must be reliable and valid for the purpose for which they are being used. Further, tests designed to measure progress towards standards must be appropriately aligned with standards, curriculum and instruction, and opportunity to learn. School psychologists should provide consultation to districts and policymakers to assure that technical issues tied to test construction and selection are addressed.
- Tests should be critically reviewed to determine whether they are designed and developed to be accessible and valid for the widest range of children and youth, including students with disabilities and students with limited English

proficiency. When tests are so designed, fewer accommodations will be needed and scores can be more validly compared.

- States should distribute information about the appropriate use of large scale assessments and caution educators and parents about the limitations of such tests.

Training

- Policymakers must fund and support ongoing staff development and training opportunities for educators as required by NCLB. The rapid pace of implementing inclusive large scale assessments has not allowed sufficient opportunities for training. In many cases, parents, educators, and administrators are making decisions related to testing and inclusion without clear procedures in place or an understanding about the effects of their decisions.

Evaluation and Research

- All standards testing programs must have a systematic evaluation plan to assess appropriate selection and implementation of procedures as well as student and system outcomes. Evaluation must consider the match between the use of a test and its design; differences in performances across groups of students and possible sources of bias; the degree to which all students are included; compliance with intended accommodations, modifications and alternative procedures; and the intended and unintended consequences of the testing program for individual students, staff, schools, districts and states. Evaluation plans should be continuously monitored and modified as needed.
- NASP supports ongoing research to address many unanswered questions about large scale assessment and to assure development and implementation of accurate, fair and useful measures of student and system progress.

Funding

- Policymakers must ensure that mandated state-wide or district-wide assessment programs have sufficient funds and timelines to develop, implement, maintain, and evaluate a high quality process. Large scale assessment is a complex and costly endeavor when appropriately designed and implemented. However, such assessment is even more costly when inadequate or inappropriate procedures are used.

Summary

It is the position of the National Association of School Psychologists that standards-based tests be used as global indicators of student and program progress, and to

highlight the need for additional resources, not to determine educational placement or graduation eligibility for an individual child, or to establish rewards or sanctions for any personnel, school or district. Policymakers are urged to carefully monitor and evaluate the actual consequences of large-scale assessment programs and to implement essential guidelines for the development and application of these accountability systems.

- Adopted by the NASP Delegate Assembly, April 12, 2003

Key Supporting Resources

Almond, P. J., Lehr, C. A., Thurlow, M. L., Quenemoen, R. (*in press*). Participation in large-scale state assessment and accountability systems. Chapter in *Large-Scale Assessment Programs for All Students: Development, Implementation and Analysis*.

American Educational Research Association (2000). AERA position statement concerning high-stakes testing in PreK-12 education. Available: www.aera.net/about/policy/stakes.htm

Elliott, S. N., Braden, J. P., & White, J. L. (2001). *Assessing one and all: Education and accountability for students with disabilities*. Washington, DC: Council for Exceptional Children.

Heubert, J. P., & Hauser, R. M. (1999). *High stakes testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press. Available: www.nap.edu/books/0309062802/html/index.html

National Association of School Psychologists (2002). *Large scale assessments and high stakes decisions: Facts, cautions and guidelines*. Bethesda, MD: Author.

National Research Council (1999). *Testing, teaching, and learning: A guide for states and school districts*. Washington DC: Author.

Quenemoen, R. F., Lehr, C. A., Thurlow, M. L., Massanari, C. B. (2001). *Students with disabilities in standards based assessment and accountability systems: Emerging issues, strategies, and recommendations*. (Synthesis Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Ysseldyke, J. E. & Bielinski, J. (2001). *Critical questions to ask when interpreting or reporting trends in the large-scale test performance of students with disabilities*. Washington, DC: Council for Chief State School Officers, State Collaborative on Assessment and Student Standards.

Additional Resources

Council for Exceptional Children (www.cec.sped.org)

National Association of School Psychologists (www.nasponline.org)

National Center for Educational Outcomes, University of Minnesota
(www.coled.umn.edu/NCEO)