

American Evaluation Association
Position Statement on
HIGH STAKES TESTING
In PreK-12 Education

High stakes testing leads to under-serving or mis-serving all students, especially the most needy and vulnerable, thereby violating the principle of “do no harm.” The American Evaluation Association opposes the use of tests as the sole or primary criterion for making decisions with serious negative consequences for students, educators, and schools. The AEA supports systems of assessment and accountability that help education.

Recent years have seen an increased reliance on high stakes testing (the use of tests to make critical decisions about students, teachers, and schools) without full validation throughout the United States. The rationale for increased uses of testing is often based on a need for solid information to help policy makers shape policies and practices to insure the academic success of all students. Our reading of the accumulated evidence over the past two decades indicates that high stakes testing does not lead to better educational policies and practices. There is evidence that such testing often leads to educationally unjust consequences and unsound practices, even though it occasionally upgrades teaching and learning conditions in some classrooms and schools. The consequences that concern us most are increased drop out rates, teacher and administrator deprofessionalization, loss of curricular integrity, increased cultural insensitivity, and disproportionate allocation of educational resources into testing programs and not into hiring qualified teachers and providing sound educational programs. [i] The deleterious effects of high stakes testing need further study, but the evidence of injury is compelling enough that AEA does not support continuation of the practice.

While the shortcomings of contemporary schooling are serious, the simplistic application of single tests or test batteries to make high stakes decisions about individuals and groups impede rather than improve student learning. Comparisons of schools and students based on test scores promote teaching to the test, especially in ways that do not constitute an improvement in teaching and learning. Although used for more than two decades, state mandated high stakes testing has not improved the quality of schools; nor diminished disparities in academic achievement along gender, race or class lines; nor moved the country forward in moral, social, or economic terms. The American Evaluation Association (AEA) is a staunch supporter of accountability, but not test driven accountability. AEA joins many other professional associations in opposing the inappropriate use of tests to make high stakes decisions. [ii]

Violations of AEA Guiding Principles and Other Professional Standards

Evidence of the impact of high stakes testing shows it to be an evaluative practice where the harm outweighs the benefits. Many high stakes testing programs

- invoke a fallible single standard and a single measure, a practice specifically condemned by the Standards on Educational and Psychological Testing

- are implemented and used to make high stakes decisions before sufficient validation evidence is obtained and before defensible technical documentation is issued for public scrutiny
- are employed without credible independent meta-evaluation
- are flawed, both in technical adequacy and in accuracy of scoring and reporting
- promulgate undue centralized control of what is taught and how
- channel educational offerings to satisfy monolithic, narrow, test-defined state standards rather than address the differential needs of students in local schools
- draw schools into narrow conceptions of teaching and education that leave children deprived of the history, cultural perspective, personal experience, and interdisciplinary nature of subject matter
- narrow the curriculum to tested subjects, usually reading, writing, and mathematics and marginalize non-tested subjects, which often include the fine arts, physical education, social studies and science
- stimulate teachers and principals to manipulate test scoring and standards, change students' answers, send slow learners away on testing day, or otherwise invalidate test scores
- consume a disproportionate amount of student and teacher time that takes away from other valued school goals and activities, e.g., spending as much 30% of the school year preparing specifically for tests
- assume that all children, including English language learners and special education students, learn in the same ways at the same rate and that they can all demonstrate their achievements on standardized tests
- focus attention on particular students, such as those scoring just below the cut-off score and ignoring those who score well above and below the cut off score
- encourage, in direct and indirect ways, students who may not pass the test to stay home on testing days or to drop out of school altogether
- measure, for the most part, parental income and race, and therefore perpetuate racism, classism and anti-working class sentiment
- contribute to an atmosphere of distrust, fear, divisive competition, and hysteria that is antithetical to teaching and learning
- contribute to an atmosphere that pits teachers against teachers, students against students, and schools against schools in a bid for financial rewards and to avoid financial retribution
- are used unjustly to fire and discipline teachers and principals

Expectations for Improved Evaluation Practice

Recognizing that the assessment of student achievement requires policy makers, practitioners, legislators, test publishers, evaluators, media personnel, and citizens to meet high technical and ethical standards, the American Evaluation Association posits

- that both the wisdom and experience of professional teachers and fully validated standardized testing are important for sound educational decision making

- that important evaluation decisions should be made on the basis of multiple criteria and multiple high quality measures validated for specified uses
- that test publishers involved in test development or implementation should be responsible for validation of representative high stakes uses for which the tests are designed, and that test publishers should publicly object and refuse future contracts with users when the publishers' tests are misused
- that measurement specialists and advisors involved in high stakes testing programs consider not only technical and theoretical but also consequential issues, such as the welfare of students, educators, schools, and society
- that contractors for testing services, state or local, should demand appropriate validation studies and consistent high quality services from test publishers and testing service providers
- that educational policy makers and practitioners should use well conceived systems of assessment and accountability that include multiple measures, and continuously strive for better representations of what is taught and achieved
- that state and local governance of education should draw on a wide range of perspectives as to what is best for students, schools and society
- that important decisions (for example, grade to grade promotion/retention, graduation, certification, classification, monetary rewards/sanctions) about students, teachers, and schools should not be made on the basis of any single test or test battery, no matter how many times it may be taken
- that there is an urgent need to initiate and fund evaluations of high stakes testing in all states and school systems where such policies have been enacted [\[iii\]](#)
- that it is imperative that findings from these evaluations be provided to policymakers, parents, teachers, students and the public about the consequences, positive and negative, of such policies
- that evaluators of high stakes testing, and programs in which high stakes testing is prominent, should draw consistently on standards for utility, feasibility, accuracy, and propriety as found in the Joint Committee Standards for Program and Personnel Evaluation, AEA Guiding Principles, and Standards for Educational and Psychological Testing
- that government and educational institutions should avoid any legislative programs or mandates for high stakes testing that violate professional testing and evaluation standards

The most serious problem with high stakes testing is its insistence that education be evaluated in a narrow way. The practice of high stakes testing in America is an effort to treat teaching and learning in a simple and fair manner, but in a world where education is hugely complex with inequitable distribution of opportunity. When we increase the standardization of education, we need challenges from multiple viewpoints as to the costs and benefits for the children in our schools. Education requires decisions as to how children, teachers, and schools will be sustained, improved and promoted, but high stakes testing oversimplifies the decisions to be made. We declare our

obligation to follow the principle of "do no harm," and that requires us to examine consequences in real situations for all people affected, not just authorities. Current high stakes testing policies and practices fail to provide the mechanisms of review, meta-evaluation, and validation demanded by our professional standards.

Process used in developing this statement:

This statement is the result of more than a year's work in which a Task Force on High Stakes Testing in K-12 Education was appointed by the American Evaluation Association President James Sanders to explore the need for and to draft such a statement for the approval of the Association's Board of Directors. The Task Force began its work with a town hall meeting at the AEA annual meeting in 2000 with several presentations on the issues in high stakes testing and discussion by those present. Over the next year the Task Force members debated, drafted, debated and redrafted the statement. Another town hall meeting was held at the AEA annual meeting in 2001 to solicit feedback. At this same time a draft was shared with the Board of Directors for their feedback but not formal approval. Immediately after that meeting, a notice seeking feedback was posted to the AEA listserv EVALTALK. With the feedback from these sources the statement was once again revised. The AEA Board of Directors approved the position statement as submitted by the Task Force in February 2002.

Task Force members endorsing the statement are:

Linda Mabry, Sandra Mathison, James Sanders, Robert E. Stake, Daniel Stufflebeam, and Charles Thomas

Funding and support:

Development of this position statement was partially supported by the National Science Foundation (NSF) under NSF grant number 0130605. Any opinions, findings, conclusions, or recommendations expressed in this statement are those of the American Evaluation Association and do not necessarily reflect the official views, opinions or policy of the NSF.

End notes:

[i] There is a large and growing body of research on high stakes testing much of which illustrates its deleterious effects. However, the research is not univocal. To provide the reader with more information about the state of our knowledge please refer to the High Stakes Testing in K-12 Schools Annotated Bibliography, available at <http://www.eval.org/hstlinks.htm>.

[ii] AEA joins many other professional associations, teacher unions, parent advocacy groups in opposing the inappropriate use of tests to make high stakes decisions. These include, but are not limited to the American Educational Research Association, the National Council for Teachers of English, the National Council for Teachers of Mathematics, the International Reading Association, the College and University Faculty Assembly of the National Council for the Social Studies, and the National Education Association.

[iii] Evaluations of and research on high stakes testing practices and policies that focus on both intended and unintended outcomes should be routinely conducted. To this end, we offer the following incomplete list of issues, those that may be neglected, as ones that should be considered in research and evaluation studies.

- how gaps in educational achievement between minority and non-minority students are effected
- the amount of student and teacher time taken away from other valued school goals and activities

- the extent to which improvements in test scores are reflections of actual and valued student learning
- the extent to which curriculum is narrowed, and how, by the test
- the impact on non-tested subjects
- the impact on English language learners, special education students, and high mobility students, and students with special talents
- drop out rates
- the fairness, accuracy, validity, reliability and credibility of the measures of content and thinking skills that students are expect to master
- the extent and form of cheating to increase scores
- the accuracy, fairness, and disclosure of scoring procedures, cut score setting, and methods of aggregation
- incidence of disciplinary action or termination of teachers as a direct result of high stakes testing
- monetary and non monetary costs of high stakes testing practices and policies
- ethical issues, such as access to student records and student, teacher, and parent rights to know.