

HIGH STAKES TESTING IN K-12 SCHOOLS AN ANNOTATED BIBLIOGRAPHY

Introduction

This bibliography begins an ongoing effort to develop a resource for those interested in research on high stakes testing in K-12 schools. The bibliography is a work in progress, is not exhaustive, focuses primarily on empirical research, includes mostly references in the past decade, and includes multiple perspectives on the issues.

Feedback or suggestions for other entries (please send complete citation) for this bibliography should be sent to Sandra Mathison at smathison@louisville.edu.

~~~~~  
**Note:** A number of people have contributed substantially to the preparation of this bibliography. They include Sandra Mathison, University of Louisville; Melissa Freeman, Kristen Wilcox, Lynee Sauer, University at Albany, SUNY. Preparation of this publication was supported under Grant # ESI-9911868 from the National Science Foundation. The contents do not necessarily reflect the position or policies of NSF.  
~~~~~

Bibliography

Ahearn, E. M. (2000). *Students with disabilities in state assessments: The NCEO state reports. Synthesis brief.* (ERIC Document Reproduction Service No. ED445433).

This document summarizes the 1999 National Center on Educational Outcomes (NCEO) report on special education outcomes and provides an overview of the changes in assessment policies and practices for students with disabilities over the past decade. The report focuses on three major issues in the assessment of students with disabilities: rates of participation in assessments, alternate assessments, and the reporting and use of assessment results. The report found that while states have made significant progress in all three areas, only 23 states were able to provide participation data for students with disabilities in 1999 and the rates of participation varied from 15% to 100%. Issues influencing participation include attaching high stakes to test performance and the lack of exposure for students with disabilities to the content of the tests.

- Category: Student Achievement/Special Populations
- Keywords: Academic standards; accountability; disabilities; elementary/ secondary education; educational outcomes

Airasian, P. W. (1987). *State mandated testing and educational reform: Context and consequences.* *American Journal of Education*, 95(3), 393 – 412.

This article traces several social changes from the 1960s to 1980s that account for the development of new roles and expectations for state mandated standardized testing as a reform mechanism in American education today. Airasian lists eleven characteristics of the new testing movement that describes how it differs from previous uses of tests in educational reform. He offers a set of propositions for understanding the current context of testing.

- Category: Historical/Political/Legal Contexts
- Keywords: Educational reform; social change; standardized tests

Airasian, P. W. (1988). Symbolic validation: The case of state-mandated, high-stakes testing. *Educational Evaluation and Policy Analysis*, 10(4), 301 – 13.

Policies and innovations such as state-mandated, high-stakes testing are not the only possible solutions to the perceived decline in student performance in the United States, but they have received the strongest social support. This article looks at the power of symbolism using state-mandated, high-stakes testing programs as an example of the way in which public values and perceptions give legitimacy and support to certain policies and innovations over others. Three types of symbolic appeal associated with high-stakes, state-mandated testing programs are identified. Such tests have gained wide support because they symbolize order and control, a focus on important outcomes, and a return to basic moral values. The author concludes by considering the impact the belief in testing might have on the public's overall awareness of educational issues.

- Category: Historical/Political/Legal Contexts
- Keywords: Symbolic validation; public perceptions; cultural values

Allington, R. L, & McGill-Franzen, A. (1992a). Does high-stakes testing improve school effectiveness? *ERS Spectrum*, 10(2), 3 - 12.

The authors question the relationship between the increased uses of high-stakes testing and stronger accountability measures of school achievement. Part of the equation, they state, are other factors such as each individual school's retention and special education policies. To explore how these policies influence the reading achievement levels reported by schools on high-stakes tests, they conduct case studies of seven elementary schools in New York State. The study involves comparing students' reported achievement on the third-grade statewide reading test to the achievement of all students who would have taken the test if they had not been previously retained or identified as handicapped. The schools are further identified as having low, moderate, or high uses of retention and special education placements. The authors argue that the reading achievement levels provided by the schools are not an accurate description of student achievement or reading levels. They suggest that the accountability profiles provided by New York State obscure and possibly reinforce questionable educational practices.

- Category: Student Achievement/ Special Populations
- Keyword: New York; accountability measures; reading levels; retention policies; special education placement; case study

Allington, R. L., & McGill-Franzen, A. (1992b). Unintended effects of educational reform in New York. *Educational Policy*, 6(4), 394 – 414.

New York State is one of the first states to provide a public report of the proportion of children in each school who have achieved the minimum proficiency level in tested areas. There is, however, little evidence reported of the effects of particular programs or reform strategies on test scores. This study looks at the relationship between retention, remediation, and the proportion of students identified as handicapped on test scores in eleven elementary schools. The authors found that during periods of high-stakes assessment there was a significant increase in students being identified as handicapped or retained at a lower grade level.

- Category: Student Achievement/ Special Populations
- Keywords: New York State; Comprehensive Assessment Report (CAR); student achievement; reading; retention; special education

Barksdale-Ladd, M. A., & Thomas, K. F. (2000). What's at stake in high stakes testing: Teachers and parents speak out. *Journal of Teacher Education*, 51(5), 384 - 397.

This study poses the questions: "What perceptions do teachers hold about mandated standards and related tests?" and "How do teachers make instructional decisions given these mandates?" 59 teachers from two states, one southern and one northern, are interviewed. Most of them are students in graduate literacy programs. Portions of interviews are included in this article and reflect teacher frustration related to current policies and assessment practices. Parents' perceptions are also addressed in this study.

- Category: Teachers and Teaching
- Keywords: Teacher/parent perceptions; instructional decisions; accountability

Bernal, E. M., & Valencia, R. R. (2000). The TAAS case: A recapitulation and beyond. *Hispanic Journal of Behavioral Sciences*, 22(4), 540 – 556.

This is a review of legal decisions regarding the TAAS, especially the effects of the Texas Assessment of Academic Skills (TAAS) exit-level tests and on promotion and graduation decisions, especially for poor and minority students. The authors suggest much has been learned from the plaintiffs' positions and that proposals such as the Wellstone/Scott bill could curtail some of the problems associated with high-stakes testing. They offer concrete suggestions for creating fair testing and assessment practices, and question policy makers' agendas in using the TAAS for maintaining the

value of a high school diploma but at the expense of poor and minority students.

- Category: Historical/Political/Legal Contexts; Student Achievement/Special Populations
- Keywords: TAAS; retention; graduation; Senate Bill 4; Wellstone/Scott Bill; psychometrics

Borman, K. M., Kromrey, J., Katzenmeyer, W., & Piana, G. D. (2000, April). *How do standards matter? Linking policy to practice in four cities implementing systemic reform*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED445062)

This paper reports on an early assessment of the impact of the National Science Foundation's Urban Systemic Initiative (USI), a multi-year reform effort in 20 cities designed to increase student achievement in mathematics and science. The reform initiative uses both standards-based curriculum and instructional approaches as well as a constructivist approach to teaching and learning. Four different cities are described: Chicago, Miami, Memphis, and El Paso. The report focuses on some initial observations from El Paso on teaching, learning, and assessment related to the use of high-stakes tests. One observed outcome of the use of high-stakes tests is the practice of teaching to the test. Two consequences of this practice are discussed: (1) teachers were able to use test score data to target individual students' needs, and (2) teachers were limited in their ability to enable the student to develop problem-solving strategies necessary for good performance.

- Category: Curriculum and Instruction
- Keywords: Teaching strategies; teaching to the test; constructivism

Bracey, G. (2000). *High stakes testing*. (CERAI-00-32) University of Wisconsin-Milwaukee, Center for Education Research, Analysis, and Innovation. <http://www.uwm.edu/Dept/CERAI/edpolicyproject/cerai-00-32.htm>

Bracey gives an historical analysis of why and how high-stakes testing has taken form—from post-WW II through present-day politics. He contends that public concern and nervousness regarding schools and achievement has ultimately led to current testing practices. Bracey also examines how testing reform is affecting minority students, curriculum, pedagogy, as well as how it is misrepresenting school and individual achievement and progress.

- Category: Curriculum and Instruction; Student Achievement/ Special Populations
- Keywords: Minority students; public concern; student achievement

Brandt, R. (1989). *On misuse of testing: A conversation with George Madaus*. *Educational Leadership*, 46(7), 26 – 29.

In this interview, George Madaus, Director of The Center for the Study of Testing, Evaluation, and Educational Policy, discusses his views and standpoint on high-stakes testing. The article covers Madaus' research on the multiple misuses of standardized testing, as well as possible approaches to testing alternatives.

- Category: Historical/Political/Legal Contexts
- Keywords: Alternative assessment

Bussert-Webb, K (2000). Did my holistic teaching help students' standardized test scores? *Journal of Adolescent & Adult Literacy*, 43(6), 572 – 573.

The teacher-author discusses changes in her pedagogy—changes she refers to as "holistic teaching"—and how those changes have resulted in scoring improvements on the Texas Assessment of Academic Skills (TAAS). Bussert-Webb points out that her changes in teaching are an effort to reach the entire child—not to simply teach testing material.

- Category: Teachers and Teaching
- Keywords: TAAS; teaching to the test

Camilli, G., & Bulkley, K. (2001, March 4). Critique of "An Evaluation of the Florida A-Plus Accountability and School Choice Program." *Education Policy Analysis Archives*, 9(7). <http://epaa.asu.edu/epaa/v9n7/>

The Florida A-Plus accountability system uses scores from the Florida Comprehensive Assessment Test (FCAT) and student referral and dropout rates to assign schools one of 5 grades (A, B, C, D, F). An earlier evaluation of the accountability system, *An Evaluation of the Florida A-Plus Accountability and School Choice Program*, reported a high correlation between the threat of school vouchers and improved test scores. This critique takes a second look at that evaluation and suggests this correlation may be due to other factors such as sample selection, regression to the mean, how gain scores were combined across grade levels or how schools were used as units of analysis. These authors conclude that the evidence provided in the evaluation cannot support the conclusions that school vouchers are responsible for higher scores.

- Category: Student Achievement/ Special Populations; Test Development and Administration
- Keywords: Florida; FCAT; accountability; test results; validity

Carnoy, M., Loeb, S., & Smith, T. L. (2000, April). *Do higher state test scores in Texas make for better high school outcomes?* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

This paper gives a detailed overview of the Texas Assessment of Academic Skills (TAAS) and evaluates the impact of the TAAS by examining trends in statewide test

scores, as well as analyzing data from high schools to determine, among other things, if rising tests scores coincide with rising dropout rates. Some of the patterns examined in this paper include: enrollment trends; 9th to 12th grade progression; 9th grade retention; and college plans (on the part of students.) The authors also look very closely at how the TAAS affects white students vs. minority students and how the TAAS affects districts depending on where they are located - urban, suburban, rural, etc.

- Category: Student Achievement/ Special Populations
- Keywords: Texas; TAAS; minority students; dropout rate; passing rate; enrollment rate

Catterall, J. S. (1989). Standards and school dropouts: A national study of tests required for graduation. *American Journal of Education*, 98(1), 1 – 34.

This study focuses on tests students must pass to graduate from high school and what effect they may have on reduced academic aspirations and drop out rates. Catterall attempts to bridge the gap between the teacher belief that the tests do not present much of a barrier to school completion because of their rudimentary nature and the student belief that test failure causes doubt about chances of graduating. The study is based on a series of interviews with educators and school administrators and on data collected face-to-face from over 700 high school students. The study found a correlation between failing the required graduation test and the raising of doubt about graduating.

- Category: Student Achievement/Special Populations
- Keywords: Dropout; Standards; exist tests; low achievers

Cheng, L. (2000). *Washback or backwash: A review of the impact of testing on teaching and learning.* (ERIC Document Reproduction Service No. ED442280)

This article discusses the phenomenon of washback or the influence of testing on teaching and learning. The assumption of washback is that tests should determine what is valued and therefore what is taught. Tests have long been used to shape curricula, control entry to systems, and impose educational methods. This article outlines the origin of washback, negative and positive washback, and its function.

- Category: Curriculum and Instruction; Teachers and Teaching
- Keywords: Measurement-driven instruction; teaching and learning; accountability

Cimbricz, S. (2002, January 9). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives*, 10(2). <http://epaa.asu.edu/epaa/v10n2.html>

The belief that state-mandated assessments drive teaching is widespread. Cimbricz

conducts a review of the empirical literature to explore the relationship between state testing and teachers' beliefs and practices. She found that while state-mandated testing does influence what teachers do, so do other factors suggesting that there is no consistent or predictable pattern of influence. Other factors influencing teachers' work are teachers' knowledge of a subject matter, their views on learning, their status in the school organization, and their personal philosophies of education. Cimbricz recommends further studies to better understand how teachers interpret and apply state mandates.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: Assessment driven reforms; teacher beliefs

Cizek, G. (1996). Setting passing scores. *Educational Measurement: Issues and Practice*, 15(2), 20 – 31.

Cizek offers a procedural definition of standard setting focusing on the process of rationally deriving, consistently applying and describing procedures on which judgments can be made. Guidelines, models, methods, new modes of assessment and validity evidence in standard-setting are discussed. Test-centered, examinee-centered and compromise models are described. This article calls for measurement specialists to develop and refine procedures for setting standards on assessment.

- Category: Test Development and Administration
- Keywords: Validity; standards setting; assessment models

Clark, R. W., & Wasley, P. A. (1999). Renewing schools and smarter kids: Promises for democracy. *Phi Delta Kappan*, 80(8), 590 - 96.

This article explores some of the dilemmas faced by schools across the nation under two dominant reform approaches:

- (1) creating higher standards and aligning those to high-stakes assessments, and
- (2) promoting privatization by creating charter schools and/or contracting out to private management firms.

The authors discuss reasons why neither strategy will live up to proponents' ambitious claims and will not produce expected results in student learning and achievement. Standardized tests cannot uphold new performance goals and charter schools serve limited numbers of students siphoning off pedagogical reform energies that could benefit kids left behind. The authors argue in favor of rigorous, innovative performance assessments.

- Category: Student Achievement/ Student Populations
- Keywords: Educational standards; charter schools; privatization; reform; performance assessment

Clarke, M., Haney, W., & Madaus, G. (2000, January). High stakes testing and high school completion. *The National Board on Educational Testing and Public*

Policy, 1(3).

This report concludes that high stakes testing is associated with increased high school dropout rates. The evidence to support this claim ranges from examining historical data to studying ethnicity. The authors conclude that high stakes practices directly or indirectly effect dropout rates. But the authors say that more research specific to dropout rates and high stakes testing is needed to reach conclusive results.

- Category: Student Achievement/ Special Populations
- Keywords: Dropout; ethnicity; MCT (minimum competency testing); socioeconomics; Florida; Texas; retention

Clinchy, E. (2001). Needed: A new educational civil rights movement. [Electronic version]. *Phi Delta Kappan*, 82(7), 492 – 98.

Looking at the issue of equal educational opportunity from a civil rights perspective, Clinchy provides a historical and social account of how public schooling in the United States has failed to meet that standard. Evidence is provided on how schools are financing the current high-stakes testing reform movement. Clinchy argues the new standards-based movement is a violation of basic human educational rights especially in terms of parental right to choose and the responsibility of schools to develop each child's full educational potential. He proposes a new educational civil rights movement to bring about a fair, equal, and democratic system to American schools.

- Category: Historical/Political/Legal Contexts
- Keywords: Educational civil rights; equal educational opportunity; parental choice; diversity

Cohen, J., & Rogers, J. (Eds.). (2000). *Will standards save public education? (with a foreword by Jonathan Kozol)*. Boston, MA: Beacon Press.

With an opening essay by Deborah Meier, a staunch opponent of standardized education and holding an alternative view of standards, this collection of short essays by a variety of well-known educators opens up a critical dialogue on the role of standards and standardized testing in US schools.

- Category: Curriculum and Instruction; Historical/Political/Legal Contexts
- Keywords: Standards; standardized curriculum; democracy; essays

Coleman, A. L. (1998). Excellence and equity in education: High standards for high-stakes tests. [Electronic version]. *Virginia Journal of Social Policy & the Law*, 6(1), 81 – 114.

Discusses the latest wave of standards-based testing as having moved from a measurement of minimum or basic skills to one of high standards learning for all. Coleman examines state educational reform efforts and issues regarding the fairness

of testing practices as shaped by due process principles and anti-discrimination laws. Then he explores how the congruence or non-congruence between specific state standards, curriculum, instruction and tests affect the legal implications of educational decisions made based on such tests. He advocates for a more careful assessment of the design, administration and use of tests and their alignment with standards, curriculum, and instruction.

- Category: Historical/Political/Legal Contexts; Curriculum and Instruction; Test Development and Administration
- Keywords: Standards-based reform; policy decisions; civil rights; equity; discrimination; psychometrics

Coleman, A. L. (2000). Fair testing: How schools should protect students' rights in high-stakes testing. *American School Board Journal*, 187(6), 32 – 35.

In this commentary, Coleman describes the legal challenges facing districts and states that have adopted state-mandated tests as the basis for making important educational decisions affecting students. He provides several guides for educators and school board members for assessing the legal and ethical quality of statewide assessment practices.

- Category: Historical/Political/Legal Contexts
- Keywords: Legal issues; students' rights; state assessment plan

Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement*, 57(3), 373 – 399.

This article examines pitfalls in the way new forms of high-stakes assessments use conventional analyses and interpretation of scores. Of concern is the use and misuse of appropriate measures of standard error or uncertainty of result. Student performances are part of larger measures of classrooms and schools. Therefore, the authors argue that concern over whether to treat individual scores as infinite measures or as measures limited to particular contexts needs careful consideration.

- Category: Test Development and Administration
- Keywords: Generalizability; reliability; standard error; score interpretation; student achievement

Darling-Hammond, L., & Wise, A. E. (1985). Beyond standardization: State standards and school improvement. *Elementary School Journal*, 85(3), 315 – 336.

This study examines how test-based standards are affecting the teacher-learner

relationship. 43 elementary and secondary teachers from three Middle Atlantic districts were interviewed on their perceptions of how these policies affect their work. Five effects are described: altered curriculum emphasis, teaching students how to take tests, teaching students for the test, having less time to teach, and feeling under pressure. Competency-based education, competency-based teacher certification, and testing for certification are discussed as well as appropriate teacher evaluation and accountability measures. The authors suggest drawing from the effective schools' research to untangle the accountability dilemma.

- Category: Curriculum and Instruction; Teachers and Teaching
- Keyword: Teacher perceptions; competency-based education; competency-based teacher certification

DeStefano, L., & Metzger, D. (1991). High stakes testing and students with handicaps: An analysis of issues and policies. In Stake, R. E. (Ed.), *Advances in Program Evaluation 1* (pp. 267 – 288). Greenwich, CT: JAI press.

This paper examines state-level policy concerning minimum competency testing programs and students with handicaps through an analysis of court cases. The paper finds two recurring themes: (1) the need for guidelines to promote consistency across teachers, schools, and districts, and (2) the need for research and evaluation to determine the impact of practices and policies on students and programs.

- Category: Student Achievement/Special Populations
- Keywords: Minimum Competency Testing (MCT); Individualized Education Plan (IEP); Legal issues in MCT; SSATI

Domenech, D. A. (2000). My stakes well done. *School Administrator*, 57(11), 14 - 19.

Domenech argues the way tests scores are being used and interpreted is undermining the whole standards movement. Tests should be used diagnostically and locally to improve the educational quality of students.

- Category: Historical/Political/Legal Contexts
- Keywords: Test scores; standards; test administration

Donlevy, J. (2001). High-stakes environments and effective student-teacher relationships: Some lessons from special education. *International Journal of Instructional Media*, 28(1), 1 – 9.

Discussed in this article is the nationwide proliferation of high stakes testing in public education. Three basic premises behind school and schooling are discussed: economic, developmental, and social necessity. The author believes public schools are now being driven by standardization, competition, and accountability, rather than student achievement. Donlevy also addresses how the high stakes movement is

affecting students in special education classrooms.

- Category: Student Achievement/ Special Populations
- Keywords: High stakes; standards; economy; development; social development; special education

Dorn, S. (1998, January 2). The political legacy of school accountability systems. *Education Policy Analysis Archives*, 6(1). <http://epaa.asu.edu/epaa/v6n1.html>

This article explores the ways in which current discussions around testing and test results are shaping our practice of education and public policy debates. The author discusses the political legacy of statistics in shaping political policies, some reasons for the popularity of tests as an accountability mechanism, and the assumptions and political costs of statistical accountability. He concludes by stating that two direct effects of such accountability approaches: (1) damage to our collective ability for broad discussions on the aim of schooling, and (2) public impatience with school reform has been provoked due to the current narrow basis by which schools are judged. Dissenters of high-stakes testing have provided counter-arguments based on renewed teacher professionalism and autonomy but the author provides several reasons why this is not an adequate counter-argument to high stakes testing. He cites the decline of professionalism, the contradiction of professionalism with traditions of democratic control, and the failure of arguments for professionalism to address the public's dissatisfaction with public education. Suggestions for reversing the destructive tendencies of statistical accountability systems are offered.

- Category: Historical/Political/Legal Contexts
- Keywords: Accountability; statistics; public policy; professionalism

Downing, S. M., & Haladyna, T. M. (1996). A model for evaluation of high-stakes testing programs: Why the fox should not guard the chicken coop. *Educational Measurements: Issues and Practice*, 5(1), 5 – 12.

High stakes tests have consequences beyond those who take them. At stake are the institutions that design, develop, implement, and score the tests. This article examines the responsibility and validity of testing programs. The authors suggest that external evaluations of high-stakes testing programs are necessary to assure validity and public protection. They offer several criteria for assessing external review programs.

- Category: Historical/Political/Legal Contexts
- Keywords: Accountability; validity; external evaluation

Duron, S. (2000). *An annotated bibliography on assessment and LEP students*. National Clearinghouse for Bilingual Education (NCBE) Center for the Study of Language and Education. <http://www.ncbe.gwu.edu/ncbepubs/reports/highstakes/bibliography.htm>

An annotated bibliography on assessment and LEP students with a special focus on the issues surrounding the use of high-stakes tests. Eighty-six entries are included along with an index for search and retrieval purposes.

- Category: Student Achievement/Special Populations
- Keywords: Limited English proficiency; accommodations

Falk, B. (2000). *The heart of the matter: Using standards and assessment to learn*. Portsmouth, NH: Heinemann.

Falk's aim in this book is to explain the complex issues involved in high stakes testing and to separate what is useful and what is harmful. Suggestions as to how to make worthy standards explicit and meaningful in teacher/school-developed curricula, how to create performance assessments that embody these standards, and how to use standards and performance assessments to enhance professional learning, are made. Falk's purpose is to empower teachers to act and speak out on behalf of children.

- Category: Curriculum and Instruction; Teachers and Teaching
- Keywords: Goals 2000; DRP Score; IQ; NAEP; NAEYC; NCTM; NCME; SAT; Title I

Firestone, W. A., & Mayrowetz, D. (2000). Rethinking "high stakes": Lessons from the United States and England and Wales. *Teachers College Record*, 102(4), 724 – 749.

Observations in case study schools and semi-structured interviews with teachers and administrators conducted in Maryland and Maine as well as in comparable schools in England and Wales form the basis of this cross-comparative study. Six themes about high stakes testing are highlighted and discussed.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: Maine; MEA; Maryland; MSPAP; England; Wales; international and national comparison

Flores, B. B., & Clark, E. R. (1999). High-stakes testing: Barriers for prospective bilingual education teachers. *Bilingual Research Journal*, 21(4), 335 – 57.

The need for bilingual and minority teachers continues to grow in the United States but teacher education programs have difficulty recruiting and retaining such candidates. One of the barriers for prospective bilingual education teachers is high-stakes testing. This article reviews the current use of entry and exit tests for prospective teachers, the implications of these tests for prospective bilingual teachers, and the relationship between actual teaching performance and performance on teacher competency tests. The data from the current tests in Texas is used as an example to

show the existence of discrepancies between actual teaching performance and passing the competency exams. The authors suggest that alternative assessment strategies should be employed with prospective bilingual teachers.

- Category: Teachers and Teaching
- Keywords: Prospective teachers; bilingual education; teacher competency exams; Texas

Goodson, I., & Foote, M. (2001, January 15). Testing times: A school case study. *Education Policy Analysis Archives*, 9(2). <http://epaa.asu.edu/epaa/v9n2.html>

This article chronicles one alternative public high school's resistance to the imposition of state mandated standards and tests. The article outlines the school's argument for maintaining its alternative forms of assessment and the commissioner of education's response.

- Category: Historical/Political/Legal Contexts; Curriculum and Instruction; Student Achievement/Special Populations
- Keywords: High school; resistance; alternative assessment; case study

Gordon, S. P., & Reese, M. (1997). High-stakes testing: Worth the price? *Journal of School Leadership*, 7(4), 345 – 368.

This article reports a study of perceptions of over 100 teachers of the Texas Assessment of Academic Skills (TAAS). Teachers completed open-ended surveys on how they prepare students for the TAAS and the effects of the test on students, teachers, and schools. Interviews of twenty survey respondents were conducted to gather in-depth data on teacher perceptions of the TAAS. The study found that high stakes testing has become the object rather than the measure of teaching and learning with negative side effects on curriculum, teacher decision-making, instruction, student learning, school climate and teacher and student self-concept and motivation.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: TAAS; at-risk students; student motivation; teacher motivation

Grant, S. G. (2000, February 24). Teachers and tests: Exploring teachers' perceptions of changes in the New York State testing program. *Education Policy Analysis Archives*, 8(14). <http://epaa.asu.edu/epaa/v8n14.html>

There is considerable attention on defining higher standards for students but little is known about how teachers are responding to the challenge of helping their students meet them. This study uses focus group data collected over two years to explore how elementary and secondary New York State teachers are responding to the state mandated tests and perceive the state mandated changes. Three themes are discussed: the nature and content of the tests, the value of professional development

opportunities to learn about the standards and tests, and the rationales for and consequences of the state tests.

- Category: Teachers and Teaching; Curriculum and Instruction
- Teacher perception; professional development; elementary and secondary education

Gratz, D. (2000) High standards for whom? *Phi Delta Kappan*, 81(9), 681 – 87

Gratz believes that if the standards movement is to last, it must serve to enhance educational rather than political prosperity. Gratz points to two historic trends that have fueled the standards movement: economic concerns and disparity regarding student achievement. He also addresses current concerns associated with the high stakes movement, including: competition; accountability; prevention from graduating and being promoted; increased stress; high quantities of nightly homework; time constraints; and student inequity. He maintains that the increased negative consequences triggered by high-stakes testing may lead to jeopardizing student's developmental needs.

- Category: Student Achievement/ Special Populations
- Keywords: Negative consequences; politics; stress

Guskey, T.R. (2001). High percentages are not the same as high standards. *Phi Delta Kappan*, 82(7), 534 - 536.

Guskey asserts that setting cutoff percentages on tests is a complex process that goes beyond statistical formulas. He argues that cutoffs must be based on both teachers' judgments of the importance of the concepts addressed and consideration of the cognitive processing skills required to complete the task. Guskey makes the point that raising standards or increasing expectations for student learning cannot be accomplished by arbitrarily raising the cutoff percentages for performance levels or different grade categories. What is needed, he argues, is thoughtful examination of the tasks students are asked to complete and the questions they are asked to answer in order to demonstrate their learning.

- Category: Test Development and Administration; Teachers and Teaching
- Keywords: Standards; TAAS; cutoff percentage

Haertel, E. (1999). Validity arguments for high-stakes testing: in search of the evidence. *Educational Measurement: Issues and Practice*, 18(4), 5 - 9.

This article is the Presidential Address given at the Annual Meeting of the National Council on Measurement in Education, Montreal, April 21, 1999. The author discusses three issues about validity in high-stakes testing. First, he provides an overview for how validity is currently determined. Second, he describes and presents a detailed validity argument for a large-scale testing program. Third, he suggests several

strategies for studying and providing different perspectives as part of an ongoing evaluation of test validity. The purpose of listening to various stakeholders' assumptions about testing is to develop stronger validity arguments for high-stakes testing.

- Category: Test Development and Administration
- Keywords: Test validity; validity arguments

Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2 – 7.

Until recently test scores were used for a limited set of purposes. The increased uses of standardized achievement scores for accountability purposes have increased opportunities for test score pollution. Test score pollution is based on contextual factors that alter test performance regardless of the construct the test intends to measure. Two major sources of test score pollution (student preparation and test administration practices) are described. The authors believe that such pollution is pervasive in American education and suggest ways to combat test score pollution.

- Category: Test Development and Administration; Historical/Political/Legal Contexts
- Keywords: Test score pollution; accountability; cheating; test administration

Haney, W. (1993). Testing and minorities. In L. Weis and M. Fine (Eds.), *Beyond silenced voices: Class, race and gender in United States schools* (pp. 45 – 73). Albany, NY: State University of New York Press.

Many studies have shown that minorities tend to do worse on standardized tests than majority individuals suggesting that they are somehow less able. Haney offers an overview of the history of standardized testing as well as studies pointing to differential performance between minority and majority individuals. After defining statistical bias and fairness, Haney shows how the tests themselves are not generally biased against minorities but that educational systems that promote differential tracks and expectations are. Relying solely on standardized tests for promotion or other educational systems is therefore biased against minorities. Haney suggests using a combination of course grades and tests for such purposes.

- Category: Student Achievement/Special Populations
- Keywords: Minorities; differential performance; test bias; fairness

Haney, W. (2000, August 19). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). <http://epaa.asu.edu/epaa/v8n41/part1.htm>

Haney outlines evidence to show that the reported "miracle" of educational reform in Texas along with the implementation of the Texas Assessment of Academic Skills (TAAS) testing system is in actuality a myth and illusion. Haney outlines in 8 parts the

Texas story. After an introduction, part 2 provides an overview of educational reform in Texas from 1970 to the present. Part 3 provides an overview of reported positive results. Part 4 exposes several areas of negative impact of TAAS. Part 5 continues this analysis with a special look at the dropout rates. Part 6 describes educators' perspectives on the TAAS. Part 7 provides other compelling evidence on student achievement or lack of it in Texas. Part 8 outlines several lessons learned from similar reform efforts in other states. Additionally, the full text of Judge Prado's ruling in a Mexican American Legal Defense and Education Fund (MALDEF) lawsuit, *GI Forum v. Texas Education Agency*, is provided in the appendix along with documentation of summary arguments made by the two sides in the case. The author was an expert witness for MALDEF.

- Category: Curriculum and Instruction; Student Achievement/ Special Populations
- Keywords: Texas; TAAS; legal issues; drop-out rates; teacher perspectives

Haney, W., Madaus, G., & Lyons, R. (1993). *The fractured marketplace for standardized testing*. Boston: Kluwer Academic Publishers.

The book scrutinizes the commercial aspect of testing and the effects of the marketplace on the quality of tests and test use. The authors find that monopoly markets prevail in some segments of the marketplace while in others small numbers of firms have oligopolistic control. The analysis ends with data from 1992, but the authors argue that the most relevant result of the book lies in the lesson that more care must be taken to avoid continuing to rely on imperfect test instruments arising from a highly fractured test market.

- Category: Test Development and Administration
- Keywords: Test validity; standardized testing industry; costs of testing; legislation; test preparation

Harman, S. (2000). Resist high-stakes testing!: High stakes are for tomatoes. *Language Arts*, 77(4), 332.

Harman describes ways in which educators are disenfranchised and frustrated by the implementation of standardized testing and suggests one course of action is a protest and resistance strategy. Suggestions on methods of resistance, as well as organizations and individuals that can be contacted to promote this effort, are included.

- Category: Historical/Political/Legal Contexts
- Keywords: Resistance; authentic assessment

Herman, J. L., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, 12(4), 20 - 25, 41 - 42.

This study looks at the effects of standardized testing on teaching and learning processes in upper elementary classrooms in eleven districts in nine states. Survey responses received from 341 teachers reveal information about the pressure that teachers feel to improve test scores and the amount of time teachers spend on test preparation. The authors conclude that the presence of standardized testing has a substantial effect on the kinds of teaching and learning that go on in schools especially in schools that serve low-SES students.

- Category: Teachers and Teaching; Student Achievement/ Special Populations; Curriculum and Instruction
- Keywords: Teacher attitudes; instructional practices; social class

Hesch, R. (2000). Mass testing and the underdevelopment of inner-city communities. *The Alberta Journal of Educational Research*, 46(1), 49 – 64.

Hesch argues that the standardized testing movement contributes to the decline of inner-city communities. He examines how mass testing reduces teachers' abilities to respond to inner-city students needs and to utilize effective and culturally sensitive approaches to teaching and learning. Finally, he provides arguments for and against standardized testing as well as efforts by parents and teachers to resist standardized testing as a means to educational equality.

- Category: Historical/Political/Legal Contexts; Student Achievement/Special Populations
- Keywords: Manitoba; social class; inner cities; student achievement; culture; resistance

Hess, F. M., & Brigham, F. (2000). None of the above: The promise and peril of high-stakes testing. *American school board journal*, 187(1), 26 – 29.

The authors believe that current accountability practices cannot coexist within the traditional culture of education. Hess and Brigham weigh the costs and benefits of instituting high stakes reform and consider the potential to increase equity, the impact on the curriculum, and the ability to track and evaluate teachers and staff based on student performance. They are concerned that though the standards may improve education as a whole, they may also negatively impact poor and minority students, detract from teacher morale, and waste educational resources.

- Category: Curriculum and Instruction; Student Achievement/ Special Populations
- Keywords: Accountability measures; evaluation; poor students; minority students

Heubert, J. P. (2001). *High-stakes testing: Opportunities and risks for students of color, English-language learners, and students with disabilities*. The Civil Rights Project, Boston, MA: Harvard University.

<http://www.law.harvard.edu/civilrights/conferences/SpecEd/moreinfo.html>

This article examines both the opportunities and risks inherent within the standards and high-stakes testing movements for students within specific populations. The extent of graduation and promotion testing in the U.S and of high-stakes testing is discussed. Norms of appropriate test use and elements of sound testing policies are described. The article concludes with recommendations regarding the use of tests for high-stakes purposes including: refrain from using high-stakes tests until schools are actually teaching relevant knowledge; ensure that the high-stakes test is valid for its intended purpose; and avoid "either-or" decisions through effective early intervention.

- Category: Student Achievement/ Special Populations
- Keywords: Standards movement; promotion; graduation; drop-out rates; "world-class" standards

Heubert, J. P., & Hauser, R. M. (Eds.) (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.

This book addresses how the testing environment is affecting students' performance and achievement within a cultural context--socioeconomic and political. The book is divided into three parts: context, interpreting and assessing results, and ensuring appropriate uses of tests.

- Category: Historical/Political/Legal Contexts ; Student Achievement/ Special Populations
- Keywords: Tracking; assessment; policy; legality; special learners

Hilliard, A.G., III. (2000). Excellence in education versus high-stakes standardized testing. *Journal of Teacher Education*, 51(4), 293 - 304.

Hilliard believes that quality, skilled teachers are being sidelined in favor of a high stakes testing curriculum. Hilliard contends that high standards in public education are important and necessary; however, he does not believe that high standards equate to high stakes. He argues that abuses of minorities and the poor, especially through the constructs of IQ and intelligence, are still very much alive in our culture. Also examined in this article are the effects of teacher certification testing—overwhelmingly teachers who score lower on certification tests are the teachers working in the poorest, lowest scoring districts.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: Intelligence; minorities; social class; teacher certification tests

Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, 54(5), 482 – 93.

This study examines the ways the Texas Assessment of Academic Skills (TAAS) has effected the practices of professional educators and whether the presence of

high-stakes assessment was thought to threaten or compromise good teaching. The participants include teachers, curriculum supervisors, and reading specialists who are all members of the Texas State Reading Association (TSRA). A survey covering topics ranging from general attitudes, test preparation, to the effects of TAAS on students was sent to 750 participants. The analysis is based on the survey data from 200 returned surveys. The overall findings support other studies that report the negative effects of high-stakes testing on teaching and learning. Specific steps are offered to recapture excellence in teaching and to challenge high-stakes reform efforts, especially with regards to reading.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: Texas; TAAS; reading assessment; negative effects

Hollenbeck, K., Tindal, G., & Almond, P. (1988). Teachers' knowledge of accommodations as a validity issue in high-stakes testing. *Journal of Special Education, 32(3), 175 – 83.*

This pilot study determined the extent of knowledge possessed by general education and special education teachers regarding allowable accommodations on state assessment exams for students with disabilities. Also studied was “how accommodation choices influence the validity of decisions resulting from this assessment.” Test modifications and test accommodations are not synonymous according to the authors. Accommodations do not change the test, rather they are used as a tool for comparing students. Modifications on the other hand, do change the exam and the exam’s administration. The authors conclude that there is a need—at the local and state level—for pre-service and in-service training for educators regarding acceptable accommodations.

- Category: Student Achievement/ Special Populations
- Keywords: Accommodations; modifications; teacher knowledge; “teachers as measurement experts”; survey

Holman, L. J. (1995, April). *Impact of ethnicity, class, and gender on achievement of border area students on a high-stakes examination. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco (ERIC Document Reproduction Service No. ED384627)*

This study explores whether student characteristics such as ethnicity, socioeconomic status, and gender are predictive of their status on the Texas Assessment of Academic Skills (TAAS) competency exam. The sample for this study was 363 randomly selected 5th grade students from the El Paso Independent School District. The findings suggest that ethnicity and class are significant variables for predicting student test performance. Several recommendations are offered.

- Category: Student Achievement/ Special Populations

- Keywords: Texas; TAAS; ethnicity; socioeconomic class; gender; elementary students

Holmes, D. & Darmon, S. (2000). *LEP Students and High Stakes Assessment*. National Clearinghouse for Bilingual Education. <http://www.ncbe.gwu.edu/ncbepubs/reports/highstakes/index.htm>

This report summarizes what tests are administered to limited English proficient students across the country, the challenges of including LEP students in mandated high stakes testing, the accommodations made for LEP students, and case studies of testing practices and issues for LEP students in many different states.

- Category: Student Achievement/Special Populations
- Keywords: Limited English proficiency; testing accommodations

Howe, H. (2000). *High-stakes trouble*. *American School Board Journal*, 187(5), 58 – 59.

In this article Howe explains his response to an earlier published article regarding improving student performance for those children who come from poverty-stricken backgrounds. Howe believes that schools, by themselves, cannot increase performance through high stakes; education and attention relating to circumstances outside the academic arena must also be addressed.

- Category: Student Achievement/ Special Populations
- Keywords: Poverty; labeling

Huber, R. A., & Moore, C. J. (2000). *Educational reform through high stakes testing--don't go there*. *Science Educator*, 9(1), 7 – 13.

North Carolina's ABCs of Public Education is outlined in this report. The focus is science reform and instruction in grades K-8. Also included is an overview of the words which define the acronym "ABC": Accountability; Basic (subjects); and Control (of curriculum and instruction at the local level.) Rewards and punishments—the stakes— for teachers associated with the New ABCs are examined, as are issues of equity surrounding science education and issues of excellence. It is the belief of the authors that science educators, nationally, should take notice and take warning of the New ABC's program.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: North Carolina; New ABCs of Public Education; rewards and punishments; equity and excellence

Jacob, B. A., & Levitt, S. D. (2001). *Rotten apples: An investigation of the prevalence and predictors of teacher cheating*. <http://economics.uchicago.edu/download/teachercheat61.pdf>

As the stakes associated with standardized testing increase, there is a rising concern that teachers and administrators may feel compelled to cheat either in the way they administer the test or by altering student responses afterwards. In this study, Jacob and Levitt develop an algorithm for detecting teacher cheating on standardized tests. Using data from the Chicago Public Schools, the authors examine the question-by-question answers provided by students in grade 3 through 7 on the Iowa Test of Basic Skills (ITBS) for the years 1993-2000. Cheating is determined through two measures: unexpected test score fluctuations and unusual answer strings for students within a classroom. The authors found over 1,000 separate instances of classroom cheating which represents 4-5% of all classrooms in that district.

- Category: Test Development and Administration
- Keywords: cheating; ITBS; Chicago

Jenkins, J. A. (1993, April). *Can quality program evaluation really take place in schools?* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA. (ERIC Document Reproduction Service No. ED397067)

Addressed in this paper is the accuracy of high stakes assessments. Jenkins explores whether high stakes assessment results are misleading due to extraneous factors. Some of the major problems with high stakes assessment he addresses are: lack of formal education regarding educational assessment; tests not being properly administered; possible lack of student motivation; students not prepared w/ test taking skills; and attention being paid only to scores. Jenkins includes changes that must be implemented before assessment reform will be successful.

- Category: Test Development and Administration
- Keywords: Test administration; measurement; student motivation

Jennings, J. F. (1998). *Why national standards and tests?: Politics and the quest for better school.* Thousand Oaks: Sage.

This book provides a behind-the-scenes look at how Congress and the Executive Branch have dealt with the issue of national standards and tests. Jennings reviews the major debates about the appropriateness of national standards and tests for all American schools. The author includes an account of how business and government leaders encouraged setting higher standards. Jennings finds that the debates between 1989 and 1997 showed that the idea of national standards and testing is controversial, but asserts that national action is needed in schools and setting higher standards has helped the country develop better schools.

- Category: Historical/Political/Legal Contexts
- Keywords: Goals 2000; Goals 2002; educational reform; national education goals; ESEA

Jett, D. L., & Schafer, W. D. (1993, April). *High school teachers' attitudes toward a statewide high stakes student performance assessment*. Paper presented at the Annual Meeting of the American Education Research Association, Atlanta, GA. (ERIC Document Reproduction Service No. ED357373)

A survey was sent to 1,220 Maryland high school teachers on their attitudes toward the Maryland Writing Test. The authors conclude that overall (97%) teachers have favorable attitudes toward the test but that English language arts teachers place significantly more importance on the test than do teachers of mathematics, science, and social studies. Teachers generally agree, however, on characteristics they consider most important and least important suggesting that fairness is of prime importance to all high school Maryland teachers.

- Category: Teachers and Teaching
- Keywords: Maryland; Maryland Writing Test; high school; teachers' attitudes

Johnson, E., Kimball, K., Olson Brown, S., & Anderson, D. (2001). A statewide review of the use of accommodations in large-scale, high-stakes assessments. *Exceptional Children*, 67(2), 251 - 264.

Current standards-based reform efforts aim to educate all children including special education students. Advocates for students with disabilities support the inclusion of students with disabilities in state-mandated assessments and while most states provide accommodations for such inclusion, the psychometric, legal, and practical challenges of such inclusion are not well researched. This review finds that accommodations are provided but in ways that are inconsistent across districts and/or states. Specifically, the accommodations procedures used for the Washington Assessment for Student Learning (WASL) for 4th and 7th grade in 1998 are examined. These researchers conclude that the accommodations provided do not appear to place students in need of accommodations at an advantage over other students but do raise questions for discussion as to the ways various accommodations are determined and implemented.

- Category: Student Achievement/ Special Populations; Test Development and Administration
- Keywords: WASL; students with disabilities; test accommodations; test bias

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81(3), 199 - 203.

The public school system in North Carolina and their high-stakes testing environment are focused on in this piece. The focus of this plan is improving student performance through high stakes testing; school-based accountability (through a reward system);

and increased local control. Researchers conducted a statewide survey of the opinions of teachers effected by this new legislation. The survey addresses how pedagogy has changed since the implementation of the latest reforms.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: North Carolina; accountability

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64(3), 425 – 61.

The validity of test-based decisions is dependent on the appropriateness of the passing score and its relationship to the performance standard the test is designed to measure. Therefore validity is dependent on the interpretation of test scores not on the test item itself. This article outlines various approaches to the validation of performance standards, the setting of passing scores, and the assumptions upon which these are based. Kane then outlines the kind of evidence used to validate score interpretations providing the strengths and weaknesses of each.

- Category: Test Development and Administration
- Keywords: Performance standards; passing scores; score validation; arbitrariness of standards

Kannapel, P. J., Coe, P., Aagaard, L., Moore, B. D., & Reeves, C. A. (2000). Teacher responses to rewards and sanctions: Effects and reactions to Kentucky's high-stakes accountability program. In B. L. Whitford & K. Jones (Eds.), *Accountability, assessment, and teacher commitment: Lessons from Kentucky's reform efforts* (pp. 127 – 146). Albany, NY: State University of New York Press.

Under the Kentucky Education Reform Act (KERA) schools are expected to achieve predetermined levels of improvement on state-mandated assessments or face sanctions. On the other hand, teachers who raise scores beyond the goal set by the state are eligible for financial rewards. This longitudinal study tracked the implementation of KERA in 20 Kentucky schools in four rural school districts from 1990-2000. School observations as well as open-ended interviews were used to gather how accountability measures were influencing school practices and how educators and parents perceived KERA. School effects such as more writing in classrooms and more use of open-response questions on classroom tests were noticed. Another effect was the addition of an arts and humanities curriculum rarely covered in rural schools but assessed on the Kentucky Instructional Results Information System (KIRIS). However, the primary effect was that educators were more interested in bringing test scores up than in promoting higher levels of learning and that the use of rewards and punishments may have decreased rather than increased teacher motivation.

- Category: Teachers and Teaching; Curriculum and Instruction

- Keywords: Kentucky; KERA; KIRIS; teacher rewards; accountability measures; student achievement

Kaplan, L. S., & Owings, W. A. (2001). How principals can help teachers with high-stakes testing: One survey's findings with national implications. *NASSP Bulletin*, 85(622), 12 – 24.

Teachers are intrinsically involved in the high-stakes testing movement. Therefore, understanding how they perceive the high-stakes testing efforts and how they understand their role and the principal's role in these efforts is important. In this study, 700 Virginia teachers were surveyed on their perceptions of the effects of high-stakes testing on teaching and learning. The results suggest that teachers have mixed feelings about the high-stakes testing movement. Teachers were more likely to show support for standards-based, high-stakes testing when instructional best practices exist, when assessment is built into instruction, and when weaker students are given support and opportunities to learn. Teachers were less likely to support standards-based, high-stakes testing when they felt their leaders were ambivalent and/or when they valued more authentic learning experiences for students.

- Category: Curriculum and Instruction; Teachers and Teaching
- Keywords: Teachers' perceptions; survey; Virginia

Ketter, J., & Pool, J. (2001). Exploring the impact of a high-stakes direct writing assessment in two high school classrooms. *Research in the Teaching of English*, 35(3), 344 - 393.

This study examines the impact of a direct writing assessment on two high school English classes. The Maryland Writing Test (MWT) has been a mandated graduation requirement since 1990. Through observations and interviews, this study examines how the test influences the teachers' beliefs about writing instruction as well as their strategies for teaching writing. It also explores how students who did not pass the test respond to these instructional strategies. The authors conclude that an emphasis on the test minimizes teachers' abilities to respond effectively to individual students' inability to understand the test questions and reinforce a narrow, limited writing curriculum.

- Category: Teachers and Teaching; Student Achievement/ Special Populations
- Keywords: Maryland; writing instruction; narrowing the curriculum

Kher-Durlabhji, N., Lacina-Gifford, L. J., Carter, R. B. & Jones, R. (1995). Preservice teachers' views on standardized testing practices. *Research in the Schools*, 2(1), 35 – 40.

This study focuses on student teachers' perceptions of test preparation practices used in schools, specifically in terms of the frequency and appropriateness of their use. Responses on close-ended questionnaires from 268 pre-service teachers are analyzed

in terms of whether or not participants felt likely to use a particular strategy and whether they thought it was appropriate to do so. The authors found a strong correlation between the likelihood of use and the ethicalness of a strategy.

- Category: Teachers and Teaching
- Keywords: Pre-service teachers; score enhancing strategies; ethics

Kiger, D. (1997). Using staff values to develop assessment principles for high-stakes testing. *ERS-Spectrum*, 15(2), 35 – 40.

This study examined the degree to which staff in a particular district agreed with certain assessment principles and deemed them important. Kiger speaks to the importance of aligning assessment guidelines with the “values of the school districts,” since testing results are being used not just to evaluate student performance, but also to evaluate staff and curriculum performance. The study illustrates that staff supported principles and the district’s current assessment program were not necessarily consistent. Also discussed in this study were the benefits of incorporating informed community members (stakeholders) in assessment development.

- Category: Curriculum and Instruction
- Keywords: Assessment development; staff perceptions; districts; community

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M. (2000, October 26). What do test scores in Texas tell us? *Education Policy Analysis Archives*, 8(49). <http://olam.ed.asu.edu/epaa/v8n49/>

One of the reasons the high-stakes testing program in Texas, the Texas Assessment of Academic Skills (TAAS), has received so much attention is because of the reported large gains made by students. This study examines to what extent these reported scores are providing an accurate picture of student achievement in Texas. Comparison with student gains on the National Assessment of Educational Progress (NAEP) is provided as one way to assess the validity of Texas students’ achievement gains. The authors conclude by raising serious questions about the validity of the gains in TAAS scores.

- Category: Test Development and Administration; Student Achievement/ Special Populations
- Keywords: Texas; TAAS; NAEP; gains; validity

Kohn, A. (2000a). *The case against standardized testing: raising the scores, ruining the schools*. Portsmouth, NH: Heinemann

Kohn addresses the arguments in favor of standardized testing in a question-answer format. He places politics and decisions based on them as the driving force behind standardized testing. Kohn argues that recent research explains how little test results reveal about learning and how little they “close the gap” for low-income and minority

students. He concludes with an appeal to teachers, parents, and students to take action in rethinking standardized testing and its role in education.

- Category: Curriculum and Instruction; Student Achievement/Special Populations
- Keywords: minority students; standards; SAT; MAT; TAAS; ACT; CAT

Kohn, A. (2000b). *Burnt at the high stakes. Journal of Teacher Education, 51(4), 315 - 327.*

In this article Kohn examines how and why standardized testing patterns have evolved. He discusses fairness concerns, as well as effects on instruction and students. Also studied are the different components of standardized testing: formats (norm-referenced and multiple choice); time constraints; frequency of exams; and age of students being subjected to high stakes testing environments. Kohn also takes an intensive look at the current system of rewards and punishments accompanying the high stakes movement by examining the outcomes and repercussions associated with such a system. It is Kohn's contention that the standards movement is especially harmful to poor, minority students.

- Category: Student Achievement/ Special Populations; Curriculum and Instruction
- Keywords: Rewards, punishment, minority students, fairness; norm-referenced tests

Koretz, D. M. (1992). *What happened to test scores, and why. Educational Measurement: Issues and Practice, 11(4), 7 - 11.*

The author addresses two issues: "what has happened to the achievement of American students in recent decades, and what do we know about the causes of trends in scores?" Koretz discusses trends from World War II on, including the declines on achievement tests during the 1960s and 1970s and the upturn from roughly 1974 through 1980. The author investigates the trends' pervasiveness, timing or the "cohort effect", and differences among subject areas to find clues to understanding the broad patterns of trends in achievement on standardized tests. Koretz concludes that as the importance of test scores in the public debate continues to grow the misuse of test data will increase.

- Category: Test Development and Administration
- Keywords: SAT; Achievement trends; Minority students; NAEP

Koretz, D. M., Linn, R. L., Dunbar, S. B. & Shepard L. A. (1991, April). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED340730).

The purpose of this study was to determine whether an increase in achievement test

scores extended beyond performance on that particular test to other comparable tests. The sample for this study is 840 third graders from 36 schools in a large, high-poverty and high-minority urban district. The districts' test score results for two tests in 1990 are compared to a third test given by the researchers. Overall, the results suggest that students do not generalize well from one test to another, but that they do generalize better in reading than in math. The study also raises concerns about the effects of high-stakes tests on the quality of instruction.

- Category: Student Achievement/ Special Populations; Curriculum and Instruction
- Keyword: Student performance; elementary; math; reading

Landman, J. (2000). *A state-mandated curriculum, a high-stakes test: One Massachusetts high school history department's response to a very new policy context.* (ERIC Document Reproduction Service No. ED440915)

This is a case study of one high school's history department's response to the new Massachusetts' history and social studies standards and tests. Over a period of a year, several teachers and administrators were interviewed so that responses were given before and after test administration and before and after test score releases. The study describes the teachers and administrators responses and the changes implemented in the history department to address the state-mandated standards. Landman highlights the positive and negative consequences of the history department's attempts to align their curriculum with the Massachusetts' state standards.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: Massachusetts; MCAS; history; social studies; teacher perception; case study

Langenfeld, K., Thurlow, M., & Scott, D. (1997). *High stakes testing for students: unanswered questions and implications for students with disabilities. Synthesis report 26.* (ERIC Document Reproduction Service No. ED415627)

This report addresses the current culture of high stakes testing—specifically how students with disabilities are effected. How high stakes testing effects curriculum, teaching, learning, and school environments is closely examined in this paper. Also included is a section comparing the costs and benefits associated with testing programs. The authors conclude that an increase in test scores does not automatically mean students are learning more.

- Category: Student Achievement/ Special Populations; Curriculum and Instruction
- Keywords: Students with disabilities; attitudes and school climates

Linn, R. L. (2000). *Assessments and accountability. Educational Researcher, 29(2), 4 – 16.*

Tests are designed and used in a variety of ways making it more difficult to interpret and trust the results that are reported by states and districts. Linn first provides an overview of five decades of assessment-based educational reforms and then discusses salient features of the current standardized testing movement especially regarding validity and credibility of results. Linn concludes that the current use of test results for accountability purposes has negative effects for student learning. He proposes seven strategies to enhance the validity, credibility and positive impact on learning in test-based reform programs.

- Category: Historical/Political/Legal Contexts
- Keywords: Test use; validity; credibility; student achievement

Lomax, R. G., West, M. M., Harmon, M. C., Viator, K. A., & Madaus, G. F. (1995).

The impact of mandated standardized testing on minority students. *Journal of Negro Education*, 64(2), 171 - 185.

This study, supported by the National Science Foundation (NSF), sought to examine the impact of standardized tests and testing in math and science on minority students. The findings draw upon three data sources: 1) a national survey of teachers; 2) studies in six urban districts; and 3) an analysis of widely used tests and test material. The study includes content reviews of tests administered in grades 4, 8, and high school, national surveys sent to teachers of grades 4-12 (the surveys inquired about district demographics and asked for opinions regarding standards), and site interviews. Results of this study yielded a decisive disparity between both the quality and the delivery of instruction to minority students.

- Category: Student Achievement/ Special Populations
- Keywords: Educational inequity; minority students; math; science

Mabry, L. (1999). Writing to the rubric: Lingering effects of traditional testing on direct writing assessment. *Phi Delta Kappan*, 80(9), 673-679.

Mabry analyzes the contradictions between the direct assessment of student achievement in writing in classrooms and the state-mandated performance assessment. In particular, she contends that scoring rubrics are essential in large-scale and standards-based performance assessments in writing since they promote reliability assessments, but that the consequence is standardized writing as well. This in turn standardizes the teaching of writing and in the end the use of rubrics jeopardizes the teaching and learning of writing.

- Category: Curriculum and Instruction; Test Development and Administration
- Keywords: rubrics; teaching to the test; pedagogy

Madaus, G. F. (1985). Test scores as administrative mechanisms in

educational policy. *Phi Delta Kappan*, 66(9), 611 - 617.

Serious problems are associated with the use of testing as the principal mechanism of educational reform. Policymakers determined to reform education quickly have dismissed warnings about the arbitrary nature of cutoff scores as irrelevant. This article reviews the evolution of standardized testing from districtwide programs of the 1940's to instruments of state and national policy in the late 1960's and beyond. The shift in test usage is related to the minimum competency movement and the British O- and A-level examinations. Effects on curriculum, instruction, students, and teachers are discussed. The author concludes with the admonition that we are rapidly sliding back into 19th-century Utilitarianism.

- Category: Historical/Political/Legal Contexts; Curriculum and Instruction; Teachers and Teaching
- Keywords: Reform; minimum competency movement; British examination system; accountability

Madaus, G.F. (1988) *The influence of testing on the curriculum*. In L. N. Tanner & K. J. Rehage (Eds.), *Critical issues in curriculum, Part I* (pp. 83 - 121). Chicago: National Society for the Study of Education,

This chapter is a comprehensive description of testing in schools and the power of testing to influence curriculum, teaching, and learning. A number of commonly used terms in discussions of high stakes testing are defined, including low and high stakes testing, norm and criterion referenced scores, internal and external testing programs. The current issues related to the impact of high stakes testing were anticipated by this descriptive and historical account.

- Category: Curriculum and Instruction; Teachers and Teaching
- Keywords: Impact of testing; reform; tests as administrative controls

Madaus, G. F., & O'Dwyer, L. M. (1999). *A short history of performance assessment: Lessons learned*. *Phi Delta Kappan*, 80(9), 688 - 95.

This article offers a historical perspective on performance assessment, dividing it into the pre-modern, modern, and postmodern periods, ranging from 210 BC to the present. Performance assessment was traced from the Han Dynasty of ancient China to the craft guilds of 13th-century Europe. In the early 19th-century, changes in assessment were geared toward increased efficiency, moving from qualitative to quantitative judgments, making assessments more easily administered, objective, reliable, and inexpensive as the number of examinees increased. Machine-scorable, standardized, multiple-choice tests were the assessments of choice for policymakers in the U.S. until the late 1980's. In the mid-1980s the "construction of knowledge in a socio-cultural context" paradigm resulted in a return to performance assessment, also referred to as authentic assessment. Translating performance testing into the realities of high-stakes testing programs has proved difficult, due to issues of manageability,

standardization, subjectivity, unreliability, and expense.

- Category: Historical/Political/Legal Contexts
- Keywords: Accountability; authentic assessment; efficiency movement

Mathison, S., Ross, E. W. & Vinson, K. D. (2001). Defining the social studies curriculum: The influence of and resistance to curriculum standards and testing in social studies. In E. W. Ross (Ed.), *The social studies curriculum: Purposes, problems, and possibilities* (pp. 87 – 102). Albany, NY: SUNY Press.

The authors analyze the way in which educational standards and high stakes testing have effected teaching and learning in the social studies. An analysis of the political context for the current emphasis on standards based educational reform illustrates the power of the liberal—conservative consensus to provide a strong rhetorical strategy for the acceptance of standards and high stakes testing.

- Category: Historical/Legal/Political Contexts
- Keywords: social studies; standards based educational reform; politics of testing

McColskey, W., & McMunn, N. (2000). Strategies for dealing with high-stakes tests. *Phi Delta Kappan*, 82(2), 115 – 120.

Teaching students strategies for doing better on high-stakes, state-mandated tests do not always mean that they are being helped to achieve a higher level of learning. This article provides a set of short-term test-taking strategies and a set of strategies committed to long-term learning in the hope of assisting educators in being able to assess their own efforts in the classroom.

- Category: Curriculum and Instruction
- Keywords: Instructional strategies

McDonnell, L. M. (1994). Assessment policy as persuasion and regulation. *American Journal of Education*, 102, 394 - 421.

A growing body of evidence suggests that assessment is a policy that allows officials at higher levels of government to exert leverage over what happens in classrooms. Negative consequences include a widening gap in opportunities available to different kinds of students, a narrowing of the skills and content taught, a centralization of educational decision-making, and a deprofessionalization of teachers. Policymakers rely on the same assessments for multiple purposes and fail to recognize the tension between informational or persuasive objectives of assessment policy and its regulatory or accountability uses. Strategies for amalgamating professional and political perspectives include recognition of the limits of assessment data as objective sources of information, attachment of capacity-building instruments to assessment policies, and deliberation in government.

- Category: Teachers and Teaching; Historical/Political/Legal Contexts
- Keywords: Policy instruments; accountability; California; CLAS

McDonnell, L. M., & Choisser, C. (1997). *Testing and teaching: Local implementation of new state assessments* (CSE Tech. Rep. No. 442). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

This is a report of a study on the link between state-mandated assessments and teaching practices in Kentucky and North Carolina, two states implementing very different assessment approaches. In 1993-1994, data was gathered through personal and phone interviews with 139 teachers and administrators with follow-up interviews a year later with 60 of the initial participants. A daily log was also kept and collected by 23 teachers in each state to compare their instructional practices with desired state goals. The authors found that while the states had different goals, there was little difference in how the teachers in each state implemented those goals in their instructional practices. The authors concluded that the link between assessment and instruction is neither rapid nor uniform.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: Kentucky; North Carolina; teaching practices; curricular goals

McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. NY: Routledge.

Based on the premise that standardization reduces the quality and quantity of what is taught and what is learned in schools, this book traces the educational costs of reform on three Texas magnet schools. The impact of Governor Mark White's decision to involve H. Ross Perot in educational reform is viewed through the "collateral damage" to students and teachers. Curricular content is trivialized and teachers are de-skilled as a result of the emphasis on accountability, a "one-size-fits-all" approach, and a test-driven curriculum. While more and more students are passing the TAAS, fewer and fewer are actually reading. The gap between minority and privileged students widens. McNeil notes that Texas illustrates what can happen when good intentions, limited resources, and the desire for a quick fix intersect.

- Category: Curriculum and Instruction; Teachers and Teaching; Student Achievement/ Special Populations
- Keywords: Texas; TAAS; magnet schools; accountability; inequity; proficiency; minorities

Mehrens, W. A., & Popham, W. J. (1992). How to evaluate the legal defensibility of high-stakes tests. *Applied Measurement in Education*, 5(3), 265 – 283.

As the stakes rise and unfavorable decisions are made about graduation, retention,

employment and/or licensure based on a test, so does the possibility that a legal suit will be brought against test developers and users. This article examines such court cases along several dimensions to determine important and necessary standards for tests to withstand legal scrutiny. Several suggestions for test construction and use are offered.

- Category: Historical/Political/Legal Contexts; Test Development and Administration
- Keywords: Legal issues; validity; reliability; cut scores

Miller, M. D., & Legg, S. M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice*, 12(2), 9 – 15.

The authors contend that one reason for the trend incorporating alternative assessment is partly due to current standardized testing practices—which measure achievement within too narrow a scope. This article explores the possibilities for introducing alternatives into statewide curricula. Examined in this article are (2) key components of using and interpreting alternative assessment practices:

1. psychometric properties (basis for use and interpretation)
2. consequences (reactions by student, teachers, administrators, those affected)

Also explained in this article are validity, fairness, cost, and reliability, as well as the consequences to alternative assessment.

- Category: Test Development and Administration
- Keywords: Alternative assessment; psychometrics; performance

Moore, W. P. (1994). The devaluation of standardized testing: One district's response to a mandated assessment. *Applied Measurement in Education*, 7, 343 – 367.

This study examines teacher testing-related attitudes and practices in a court-ordered achievement testing setting. As part of a desegregation order, standardized tests were required as a measurement of effectiveness of the effort. The study provides background on the history of test use in the desegregation effort. The author explains, however, that desegregation is not assumed to be “*the* causative condition,” but “contributes to the linkage between mandated reform and the need to show substantial improvement in student academic performance.” The study showed that teachers were strongly influenced by the test to change their instructional efforts and curriculum, they were dissatisfied with the climate of pressure and accountability and overall, the results verify that teachers engage in inappropriate testing practices in order to improve test scores.

- Category: Test Development and Administration; Teachers and Teaching; Student Achievement/Special Populations
- Keywords: Standardized tests; accountability; ITBS; teacher testing-related attitudes; desegregation

Moran, R. F. (2000). Sorting and reforming: High-stakes testing in the public schools. *Akron Law Review*, 34 Akron L. Rev. 107.

This paper traces the roots of high-stakes testing and the ongoing debate about the propriety of its use. While uniform standards may be appropriate for businesses, all students should not be expected to attain the same skills at exactly the same age or grade level. Students who can not fit into a standard mold are penalized by high-stakes testing. The Department of Education's Office of Civil Rights (OCR) has been called upon to investigate testing practices that appear to discriminate against Hispanic children, and courts have provided some protection from high-stakes testing, as in the leading case on high school exit examinations, *Debra P. v. Turlington*. The most significant recent legal challenge is a class action suit, *GI Forum v. Texas Education Agency*, wherein individual students and several advocacy agencies allege that the use of the Texas Academic Assessment System (TAAS) as a graduation requirement discriminated against black and Hispanic students, violating due process. The *GI Forum* decision indicates that in a conservative era, federal courts are increasingly unwilling to probe the workings of state educational systems.

- Category: Historical/Political/Legal Contexts
- Keywords: Texas; TAAS; Accountability movement; Office of Civil Rights; *Debra P. v. Turlington*; Equal Educational Opportunities Act

Moss, P. A. (1994). Validity in high stakes writing assessment: Problems and possibilities. *Assessing Writing*, 1(1), 109 - 128.

Literacy education has had a strong influence on educational measurement and is largely responsible for the increased emphasis on writing in high-stakes assessments. However, tension exists between these two areas. In order to increase the validity of assessments used for high-stakes purposes the tasks, conditions and scoring criteria need to be similar for all students. On the other hand, current literacy research suggests less standardized forms of assessment, those that favor more open and collaborative work between teachers and students, are favored. A compromise that is often reached is to use less standardized forms of writing assessment in the classroom and reserve the more standardized forms for high-stakes use. Moss suggests that rather than making room for both in the curriculum the result is often a narrowing of the curricula as teachers prepare students to take the test. Moss offers several recommendations for ensuring that good literacy and alternate assessment practices do not get buried beneath the pressure of high-stakes assessment practices.

- Category: Curriculum and Instruction; Test Development and Administration
- Keywords: Writing assessment; literacy practices; alternative assessment; portfolios

Moss, P. A., & Schutz, A. (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal*, 38(1), 37 - 70.

There is little discussion about the development and agreement of standards upon which the recent standards-based assessments are based. This article explores several issues regarding the assumed justification and agreement of standards. The authors use two theories provided by the discourse ethics of Jürgen Habermas and Hans-Georg Gadamer's philosophical hermeneutics to discuss issues surrounding the consensus-seeking process in the creation of educational standards. The authors conclude that conventional consensus seeking approaches in the development of educational standards stifle the diversity that exists in the public. They offer an approach that builds on the strength of diversity and dissensus in the development process.

- Category: Historical/Political/Legal Contexts
- Keywords: Standards-based assessment; educational standards; consensus building

Nash, J. B., & Calderon, M. (1994). *Principals' perceptions of community in low performing campuses in minority settings*. (ERIC Document Reproduction Service No. ED380527)

In this paper, principals of underachieving schools (by state standards) in Texas were interviewed regarding their perceptions of being labeled "low performing" and their attempts at removing their school from the "Priority 1" were traced. A low performing school, at the time of this study, was any school where 20% or less of its students were passing all tests, but were showing no sign of improvement. Staff and professional development programs offered by the case study schools were also outlined and detailed in this paper. When this paper was finally published none of the schools were still considered "low performing." This was attributed to the school community refocusing their leadership, staff, and mission. However, researchers were concerned that some schools may appear on the list again.

- Category: Curriculum and Instruction
- Keywords: Texas; TAAS; professional development; "low performing"; Priority 1

National Clearinghouse for Bilingual Education (1997). *High stakes assessment: A research agenda for English language learners. Symposium summary*. Washington, DC: Author. (ERIC Document Reproduction Service No. ED417589)

High stakes assessment of English language learners (ELLs) was the subject of a symposium sponsored by the U.S. Department of Education, Office of Bilingual Education and Minority Languages Affairs (OBEMLA). The report covers several concerns around testing English language learners and addresses such questions as: At what point does testing a child in a second language yield meaningful results?; What accommodations are appropriate for testing ELLs?; and What is the role native language assessment plays in high stakes testing? Recommendations for research are proposed.

- Category: Student Achievement/ Special Populations
- Keywords: Conference; English language learners; English Second Language; elementary secondary education

Natriello, G., & Pallas, A.M. (1999, November). *The development and impact of high stakes testing*. Paper presented at the Conference on Civil Rights Implications of High Stakes Testing, sponsored by the Harvard Civil Rights Project, Teachers College, and Columbia Law School. (ERIC Document Reproduction Service No. ED 443871)

In this paper, Natriello and Pallas review several reasons for the growth of the use of testing for high school graduation in the United States. Using New York, Minnesota and Texas as examples, they look at the impact of such requirements across racial and ethnic groups, including significant differences in passing rates for students from different ethnic and socioeconomic backgrounds. As a result they question whether the motivational consequences of high stakes testing are indeed uniformly positive across racial, ethnic, and class lines as some research has suggested. They offer some suggestions for improving the use of such tests for graduation purposes.

- Category: Student Achievement/ Special Populations; Historical/Political/Legal Contexts
- Keywords: Texas; New York; Minnesota; graduation requirements; ethnicity; disadvantaged youths; passing rates

Noble, A. J., & Smith, M. L. (1994a). *Old and new beliefs about measurement-driven reform: "The more things change, the more they stay the same"* (CSE Tech. Rep. No. 373). Los Angeles: UCLA, National Center for Research on Evaluation, Standards, and Student Testing.

This report focuses on one aspect of a larger study on the effects of the Arizona Student Assessment Program (ASAP) on teaching and learning. This paper looks specifically at the rationale and assumptions of including a performance-based test mandate in ASAP and its effects on teaching and learning. The authors conclude that there is an inherent contradiction between performance-based learning and assessment and mandated, behaviorist views of school reform that work against teachers developing the desired capacity and competence that the reform efforts are striving to achieve.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: Arizona; ASAP; performance-based assessment; cognitive-constructivist learning; mandates; professionalization of teachers

Noble, A. J., & Smith, M. L. (1994b). *Measurement-driven reform: Research on policy, practice, repercussion* (CSE Tech. Rep. No. 381). Los Angeles: UCLA, National Center for Research on Evaluation, Standards, and Student

Testing.

This paper looks specifically at the discrepancies and contradictions among policy makers' ideologies and rhetoric around student assessment in Arizona. Two contradictory positions, one process-oriented or constructivist, the other outcomes-oriented or behaviorist, are outlined. Six ideological inconsistencies in Arizona's policy makers' policies and processes are described and discussed.

- Category: Historical/Political/Legal Contexts
- Keywords: Arizona; ASAP; policy; performance-based assessment; political interests

Ohanian, S. (1999). *One size fits few: The folly of educational standards.* Portsmouth, NH: Heinemann

In her book, Ohanian focuses on fundamental problems inherent in the standards movement. The devastating effect of standardization and high-stakes testing on "nonstandard kids" is of primary concern. Relating stories of her former students, she warns of the danger of losing children because so many simply do not fit into the standard mold. The "standardistos" of the California Academic Standards Commission have failed their "nonstandard kids" to the extent that approximately one-third of the students entering ninth grade do not graduate four years later.

"Standardistos" from the business community, such as IBM CEO Louis V. Gerstner, routinely collect annual salaries and bonuses in excess of \$25 million, yet they emphasize the importance of high-stakes testing for students and merit pay for teachers as the means to educational reform. Meanwhile, urban students are forced to deal with overcrowded, dilapidated, and filthy classrooms, and teachers struggle to survive on inadequate salaries. The greater issues of poverty and social and economic reform are lost in the pro-standards movement. Ohanian concludes by offering encouragement to teachers who resist the "standardistos."

- Category: Curriculum and Instruction; Historical/Political/Legal Contexts
- Key Words: "Nonstandard kids," California Academic Standards Commission, merit pay, "standardistos"

Ohanian, S. (2001). *News from the test resistance trail.* [Electronic version]. *Phi Delta Kappan*, 82(5), 363 – 366.

In this article, Ohanian offers various state-to-state examples of different standards and assessments and how they are failing both students and teachers. She points to scenarios where teacher's careers have been jeopardized for discussing or outwardly criticizing testing practices and procedures. Actual test questions—broken down by state and grade level—are given in the article, pointing out their irrelevance and unfairness. Ohanian also sees herself as an example. She recounts the time police officers came to her door threatening to charge her with a felony if she refused to implicate parents who had organized to protest testing policies in Gwinnet County,

Georgia.

- Category: Historical/Political/Legal Contexts
- Keywords: Standards; protest; resistance

Orfield, G., & Kornhaber, M. L. (Eds.) (2001). *Raising standards or raising barriers?: Inequality and high-stakes testing in public education*. Washington D.C.: The Century Foundation Press.

This edited volume brings together articles from a variety of researchers to explore critically and comprehensively the limits and consequences of test-driven reforms in education. The nine chapters challenge the assumptions made by policy makers in support of high-stakes tests. They address such issues as: dropout and retention rates, the relation of tests to learning outcomes, student motivation, the impact of testing on curricular decisions and materials, and the impact standardized test use has on the civil rights of parents and students, especially minority parents and students.

- Category: Historical/Political/Legal Contexts; Student Achievement/Special Populations; Curriculum and Instruction
- Keywords: Civil rights; dropout rates; promotion; graduation; policies; rhetoric; test use

Palmer, D. K., & Garcia, E. E. (2000). *Voices from the field: Bilingual educators speak candidly about proposition 227*. *Bilingual Research Journal*, 24(1-2), 169 – 178.

This paper addresses the findings from a mini-study on the reactions of California bilingual educators to the implementation of Proposition 227. Proposition 227 passed into law a series of mandates that directly affect language minority children. One of these is the imposition of English-only standardized testing. Discussions with teachers revealed a particular concern over the now-required SAT-9 test in English beginning in second grade. Administrators also expressed concerns over the consequences of requiring SAT-9 testing in English and tying that testing to school evaluation. The study found that standardized testing was a main concern of all professionals, and they constantly connected testing and Proposition 227 as both being behind a statewide policy change toward English Only.

- Category: Student Achievement/Special Populations
- Keywords: SAT-9; Proposition 227; bilingual education; English Only

Paris, S. G. (1998). *Why learner-centered assessment is better than high stakes testing*. In N. M. Lambert & B. L. McCombs (Eds.), *How students learn: Reforming schools through learner-centered education* (pp. 189-209). Washington, DC: American Psychological Association.

In this chapter, Paris first makes a distinction between formative and summative

assessment. Formative assessments are usually connected to the curriculum and are used for diagnostic purposes to improve teaching and learning. Summative assessments are usually externally imposed and provide high-stakes measures for accountability. Paris believes that the emphasis on the latter is undermining good formative assessment practices in classrooms. Paris describes several unintended consequences of high-stakes' assessment practices such as a narrowed curriculum, test pollution, and cheating. He then provides a series of recommendations for implementing learner-centered assessment practices that support and encourage student learning and motivation.

- Category: Curriculum and Instruction; Teachers and Teaching
- Keywords: Formative assessment; summative assessment; learner-centered teaching; portfolios; authentic assessment; test pollution; cheating

Paris, S., & McEvoy, A. (2000). Harmful and enduring effects of high-stakes testing. *Issues in Education*, 6 (1/2), 145 – 160.

In this paper, the authors offer in-depth answers to reactions often voiced regarding high stakes testing. They also suggest methods for improving the current culture of high stakes—improvements that could be made at national, state, and local levels. Of primary concern is that testing procedures are being used and enforced as a way for policymakers to gain control over the educational system. It is the position of the authors that more research and monitoring be undertaken concerning the contexts of high-stakes testing.

- Category: Test Development and Administration
- Keywords: Accountability; high stakes reform; reaction/criticism; improving testing

Paris, S., & Urdan, T. (2000). Policies and practices of high-stakes testing that influence teachers and schools. *Issues in Education*, 6(1/2), 83 - 105.

This paper examines the policies of some states incorporating high stakes testing programs. Of particular focus are the effects of high stakes testing patterns on teachers in certain states—how they perceive high stakes; how they prepare their students for testing; as well as how valid they believe high stakes testing to be. The paper also incorporates perspectives from administrators and parents. The authors believe there are serious problems with the current environment of high stakes, and they suggest policies that might minimize the negative consequences of testing without getting rid of high stakes testing all together.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: Michigan; MEAP; negative consequences; Golden Apple awards

Passman, R. (2000). *Pressure cooker: Experiences with student-centered*

teaching and learning in high-stakes assessment environments. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Dallas, TX. (ERIC Document Reproduction Service No. ED440146)

This is a case study of the impact of high-stakes assessment policies on the pedagogy of a fifth grade teacher in a large, mid-western school district. Passman argues that high-stakes testing conflicts with good teaching and has "a chilling effect on the implementation of student-centered, constructivist practices in the classroom." He asks the question, "Who is benefiting?"

- Category: Teachers and Teaching
- Keywords: Student-centered, constructivist teaching; assessment; pedagogy; fifth grade

Pedroza, A. (1998). Bordering on success: Mexican American students and high stakes testing. This paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA. (ERIC Document Reproduction Service No. ED420713)

This study sought to analyze and assess the effects of high-stakes testing on students in a rural district along the Texas-Mexican border. The study focused on "patterns of change" among these students since high-stakes were implemented—including student achievement, student retention, placement of students in special education programs, and also the district's response to promoting greater achievement within their schools. The author concluded that students living along the Mexican-American border may not be benefiting from standardization policies—policies that assume all American students are the same.

- Category: Student Achievement/ Special Populations
- Keywords: Texas; TAAS; Texas-Mexican border (district); promotion; retention; administrative response

Perkins-Gough, D. (2000). Accountability and high-stakes testing: Are we asking the right questions? ERS Spectrum, 18(4), 4 – 11.

This article presents an overview of the questions about high stakes testing currently being debated and proposes that understanding the issues of the debate is only a first step toward finding solutions. Perkins-Gough asserts that the next step is to ask additional questions with a focus on finding common ground between proponents and critics of high stakes testing. The article concludes with an appeal for proponents and critics to move beyond the rhetoric that masks the complexity of the issue. The author proposes that school leaders can help reframe the discussion in ways that put children's welfare first and engage both critics and proponents in finding common ground and common solutions.

- Category: Curriculum and Instruction; Teachers and Teaching

- Keywords: NAEP; achievement gap; NRC; Essential Academic Learning Requirements (EALR)

Phelps, R. P. (1996). Are U.S. students the most heavily tested on earth? *Educational Measurement: Issues and Practice*, 15(3), 19 - 27.

Both national data sources (the General Accounting Office and Organization for Economic Cooperation and Development) and international sources (the International Association for the Evaluation of Educational Achievement, the International Assessment of Educational Progress, and the Second International Mathematics and Science Study) provide mixed answers to questions regarding the character and extent of system-wide testing in the U. S. and thirteen, primarily European countries. Findings indicate that U.S. tests tend to be in a multiple-choice format, of shorter duration, and placed at key transitions in the students' careers. U. S. students tend to take a greater number of individual administrations of short, norm-referenced system-wide tests with no or low stakes attached to them as compared to students in all thirteen other countries. U. S. students also take more classroom tests in mathematics and science than their international counterparts, but less in reading. Data from the aforementioned sources suggest that one could argue the point to the extent of testing in the U. S. from either side.

- Category: Test Development and Administration
- Keywords: SIMMS; system-wide testing; international comparisons

Phelps, R. P. (1998). The demand for standardized student testing. *Educational Measurement: Issues and Practice*, 17(3), 5 - 23.

This article summarizes public responses on the test related sections of over 70 polls and surveys comprising over 100 questions and items. Phelps found the public is strongly in favor of high-stakes testing and wants more of them as well as more teaching to the test.

- Category: Historical/Political/Legal Contexts
- Keywords: Public opinion; surveys; polls

Phelps, R.P. (1999). Why testing experts hate testing? *Fordham report*, 3(1). (ERIC Document Reproduction Service No. ED429089)

In response to educators and education "experts" who oppose high-stakes testing, Phelps examines four "case studies:" the National Assessment of Educational Progress (NAEP), the Texas Assessment of Academic Skills (TAAS), North Carolina's assessment program, and the SAT. Eight specific arguments against testing are rebutted including test score inflation, curriculum narrowing, lower-order thinking in instruction and content, declining achievement, bias against women and minorities, excessive cost, international trends toward less testing, and the opposition of teachers and others who care about children. Phelps claims the debate regarding testing is fundamentally

skewed, since pro-testing advocates are legally and ethically restricted from voicing their views. The author concludes that testing experts must reexamine their beliefs about teaching and learning and contribute their knowledge and efforts to ensure that testing is done well.

- Category: Curriculum and Instruction; Historical/Political/Legal Contexts
- Keywords: NAEP; TAAS; SAT; North Carolina; bias; lower-order thinking; curriculum

Phillips, S.E. (1993). Legal implications of high-stakes assessment: What states should know. Oak Brook, IL: North Central Regional educational Laboratory.

The book provides information regarding legal issues that may arise as a result of high stakes testing. This book provides state and national education policy-makers general guidelines for developing legally defensible assessment programs. These guidelines include four major areas of concern in statewide assessment: testing to award diplomas, potential bias against historically disadvantaged groups, testing accommodations for disabled persons, and performance assessment issues. The author recommends that policymakers familiarize themselves with legal and measurement issues surrounding their assessment programs in order to prepare for legal challenges to those programs.

- Category: Historical/Political/Legal Contexts
- Keywords: Legal issues; measurement; Title VII; EEOC; test validity; test reliability; ADA; IDEA

Phillips, S.E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education*, 7(2), 93 – 120.

It has become common practice to provide testing accommodations for persons with physical disabilities. Since the passage of the Americans with Disabilities Act of 1990, however, there has been an increase in requests for accommodations for people with mental and learning disabilities as well. This article looks at the validity and measurement problems that arise from granting accommodations for mental disabilities. At issue is the question of whether scores with accommodations for students with neurological and psychological problems are comparable to scores attained without accommodations. Several recommendations are provided for developing legally defensible accommodation policies.

- Category: Student Achievement/ Special Populations; Historical/Legal/Political Contexts
- Keywords: Students with Disabilities; testing accommodations; score comparison

Pipho, C. (2000). The real sting of high-stakes failure. *Education Digest*, 66(3), 18 – 22.

Current school report card strategies are compared to minimum competency testing in the late 1970s and early 1980s. Arguments for and against high-stakes testing provide the context for understanding testing issues such as cheating, and financial and legal consequences of high-stakes testing. The Texas Assessment of Academic Skills (TAAS), Massachusetts Comprehensive Assessment System (MCAS), and the Stanford Achievement Test, ninth edition (SAT-9) are used as examples.

- Category: Historical/Political/Legal Contexts
- Keywords: TAAS; MCAS; SAT-9; educational reform; cheating; legal issues

Pitoniak, M. J., & Royer, J. M. (2001). Testing accommodations for examinees with disabilities: A review of psychometric, legal, and social policy issues. *Review of Educational Research, 71(1), 53 – 104.*

The authors trace the history of legislation that relates to the testing of individuals with disabilities. They describe the types of accommodations provided in large-scale examinations as well as some of the legal concerns surrounding those accommodations. Accommodations alter the meaning of test scores and raise psychometric concerns around issues such as validity, reliability, equating, test items, and test constructs. The authors provide a detailed overview of the research on these psychometric issues as well as ideas for future research.

- Category: Student Achievement/Special Populations; Historical/Political/Legal Contexts
- Keywords: IDEA; individuals with disabilities; test accommodations; psychometrics; validity; reliability

Popham, W. J. (2001). *The truth about testing: An educator's call to action.* Alexandria, VA: Association for Supervision and Curriculum Development.

It is Popham's contention that the educational reform strategies being implemented nationally, specifically high stakes testing reform, are having deleterious effects on students. In this book he examines the current misuses of high stakes testing and also offers possible solutions and methods for improving tests and testing programs. He also looks at ways to improve instruction and learning within the context of testing—that is, he looks at the prospect of creating tests that are "instructionally illuminating" for both educators and students.

- Category: Student Achievement/ Special Populations; Test Development and Administration
- Keywords: Assessment/measurement; classroom assessment; accountability; "instructionally illuminating"

Potter, D.C., & Wall, M.E. (1992). *Higher standards for grade promotion and graduation: Unintended effects of reform.* (ERIC Document Reproduction

Service No. ED348750)

This paper looks at the unintended effects of reform when it comes to grade promotion and graduation. A detailed background and history is given about South Carolina's Education Improvement Act (EIA) programs—including incentives and consequences associated with these programs. The data analyzed in this study shows modest increases in student achievement since the onset of educational reform, though the data also shows an increase in harmful effects on students from certain demographic groups. Potter and Wall stress the need for more studies that examine the deleterious effects of testing reform.

- Category: Student Achievement/ Special Populations
- Keywords: Education Improvement Act (EIA) of 1984; Cognitive Skills Assessment Battery (CSAB); effects of testing reform acts

Rafferty, E. A., & Treff, A. V. (1994). School-by-school test score comparisons: Statistical issues and pitfalls. *ERS Spectrum*, 12(2), 16 – 19.

This article addresses some of the decisions made by institutions in designing school profiles based on test scores. The authors outline how these scores may not fairly reflect disparate populations and may be rewarding questionable educational practices.

- Category: Historical/Political/Legal Contexts
- Keywords: Accountability; school report cards; cut off scores; test scores

Sacks, P. (2000). Predictable losers in testing schemes. *The School Administrator*, 57(11), 6, 8 - 9.

Sacks believes that current high stakes reform is nothing more than a costly, educational experiment. He contends that those most deleteriously affected by high stakes are poor and minority students, while privileged students excel. Sacks maintains that the astronomical price tag attached to high stakes reform is producing nothing more than misleading, quick-fix results.

- Category: Student Achievement/ Special Populations
- Keywords: Texas; TAAS; Massachusetts; MCAS; poor and minority students; cost; inequality

Sacks, P. (2000). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus Books.

Sacks discusses the entrenchment of standardized testing in our society. He addresses the stigmas attached to standardized testing programs—labeling some students as capable, while others are labeled incapable—based solely on scoring patterns. Included is a brief history of the anti-testing movement, as well as the price tag

accompanying exams. Sacks points to evidence that supports the claims of anti-testing proponents including: standards doing little to predict actual success; standards being tightly intertwined with socioeconomic class (something he calls the “Volvo effect”); and standards not promoting meaningful educational reform.

- Category: Historical/Political/Legal Contexts; Student Achievement/Special Populations
- Keywords: Social class; anti-testing movement; spending/cost

Schleisman, J. (1999, October). *An in-depth investigation of one school district's responses to an externally mandated, high-stakes testing program in Minnesota*. Paper presented at the Annual Meeting of the University Council for Educational Administration, Minneapolis, MN. (ERIC Document Reproduction Service No. ED440465)

Schleisman’s case study investigates the effects of an externally mandated, high stakes testing program in Minnesota—examining policy changes and practice changes at both the school and district level. She gives a detailed background of testing trends, and discusses educational policies and legislation in the context of Minnesota.

- Category: Historical/Political/Legal Contexts; Curriculum and Instruction
- Keywords: Minnesota; MBST; accountability; qualitative study; externally-mandated testing program

Shepard, L.A. (1989, April). *Inflated test score gains: Is it old norms or teaching the test? effects of testing project*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED334204)

Shepard, in this paper, explores the effects test-curriculum alignment and teaching to the test, have on test scores. Included in this paper are portions of interviews —questions asked of testing directors regarding teaching to the test and the processes through which standardized tests are chosen. Shepard’s interview questions also focus on test preparation tactics, the security of the exams, and instances of cheating. Evidence from her study concludes that curriculum alignment and teaching specifically to tests are deleterious to both instruction and learning.

- Category: Curriculum and Instruction
- Keywords: Teaching to the test; curriculum alignment; spurious test scores

Shepard, L.A. (1991). *Will national tests improve student learning? Phi Delta Kappan, 73(3), 232 – 238.*

In this piece, Shepard examines the negative effects of standardized testing, the vision behind national examinations, and the problems that must be resolved if the goals of testing are to be met. Some of the negative aspects of testing she examines

include: pressure attached to high stakes yielding misleading data surrounding student achievement; curriculums being narrowed as a result of high stakes; and children with special needs being rejected by the high stakes culture. Shepard cautions that past research pertaining to the successes and failures of previous testing reform must be analyzed and considered if current testing policies are ever to be successful.

- Category: Curriculum and Instruction; Student Achievement/ Special Populations
- Keywords: Minimum Competency Testing (MCT); National Education Goals Panel (NEGP)

Shepard, L.A., & Dougherty, K. C. (1991). *Effects of high-stakes testing on instruction*. (ERIC Document Reproduction Service No. ED337468)

A brief overview is given of the evolution of standardized testing regarding student achievement—from standards being used solely as a tool for informing parents and monitoring trends to becoming one of the hottest debates surrounding educational reform. In this study teachers from two high-stakes districts were surveyed with questions concerning testing preparation and effects of testing on instruction. Some of the major findings of this study are: teachers feel pressure to improve test scores from administrations and the media; due to the emphasis on standardized tests, teachers focus on basic skills instruction; and four weeks of test preparation does not include the one to two weeks spent administering the exams. Extensive tables are included in this study to illustrate the findings.

- Category: Curriculum and Instruction; Test Development and Administration
- Keywords: Teaching to the test; test preparation; instructional effects; pressure to improve scores

Sloan, K. (2000, April). *Teacher agency and the TAAS: Maintaining the ability to "act otherwise."* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED441830)

Teachers are the fundamental focus of this paper, but also of interest is whether high stakes testing actually leads to better instruction and greater achievement. The study concentrates on four elementary schools, each with bilingual classrooms that demonstrate "exemplary" performance on the TAAS. How those schools have developed and tailored reading programs to conform to the TAAS is also looked considered. In addition, the author is specifically concerned with how the development of these programs has affected teachers. Sloan's data suggests that teachers and administrators demonstrate a "continuous flow of conduct" regarding the TAAS.

- Category: Teachers and Teaching; Student Achievement/ Special Populations
- Keywords: Texas; TAAS; teacher agency; impact/effect on teachers; bilingual classes/students

Smith, M. L. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona Student Assessment Program (ASAP): Equity and teacher capacity building* (CSE Tech. Rep. No. 425). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

This is the final report of a study on the conception, negotiation, and implementation and final demise of a measurement-driven reform effort in Arizona. This multi-site, multi-method study examines the assumptions, policies, ideologies and political values behind the conception of the Arizona Student Assessment Program and its impact on administrators, teachers, students, and schools. Several critiques are offered at local and state-wide levels, of why this measurement-driven reform effort failed to deliver its intended effects.

- Category: Historical/Political/Legal Contexts; Curriculum and Instruction
- Keywords: Arizona; ASAP; elementary; educational policy; measurement-driven reform

Smith, M.L., Heinecke, W., & Noble, A. J. (1997). *The politics of assessment: A case study of policy and political spectacle* (CSE Tech. Rep. No. 468). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

This is a report of a study on the demise of the Arizona Student Assessment Program (ASAP). This empirical study examined how the political culture in Arizona intersected with state-wide events in assessment policy. This report describes the components that made up what was to be ASAP, the historical and political underpinnings of its origin in 1992 to its eventual elimination in 1995. Interviews with and observations of policy makers, teacher surveys, focus groups with teachers and administrators, in-depth case studies of four elementary schools, and extensive reviews of the political and educational documents were used as data for this report. Seven general assertions are made about what can be learned from Arizona's change in assessment policy and how these changes were influenced by the political climate.

- Category: Historical/Political/Legal Contexts; Curriculum and Instruction
- Keywords: Arizona; ASAP; policy; elementary education

Smith, M. L., & Fey, P. (2000). *Validity and accountability in high-stakes testing*. *Journal of Teacher Education*, 51(5), 334 - 344.

The authors of this article contend that in the current high stakes testing culture, accountability and validity are being placed in oppositional camps. Smith and Fey give very detailed definitions of each term. They also explain how accountability and validity under current educational reform plans affect teachers, assessment policies, and students. High stakes reform, according to them, produces instructional and assessment techniques that especially harm students who do not come from privileged

backgrounds.

- Category: Test Development and Administration
- Keywords: ASAP; TAAS; Validity; accountability

Smith, M.L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, 10(4), 7 - 11.

This article explores the effects of external testing programs on elementary schools' curriculum and instruction. An outline, based on extensive case studies in two elementary schools in one district in Arizona, Smith and Rottenberg, on some of the unintended consequences high stakes testing has on teaching and learning is presented.

- Category: Curriculum and Instruction; Teachers and Teaching
- Keywords: Arizona; ITBS; teaching to the test; case study

Stake, R. (1998, July 21). Some comments on assessment in U.S. Education. *Education Policy Analysis Archives*, 6 (14). <http://epaa.asu.edu/epaa/v6n14.html>

This article discusses the role assessment has played in reforming American education. Assessment has taken many forms but it is now synonymous with standardized testing. The author explains that assessment-driven reforms have done little to improve our educational system but they continue to be seen as the means to better teaching and learning. Rather than add more tests, he recommends that better descriptive studies on the consequences of assessment on teaching and learning be conducted.

- Category: Curriculum and Instruction
- Keywords: Educational reform; assessment

Stoskepf, A. (1999). The forgotten history of eugenics. *Rethinking schools*, 13(3). http://www.rethinkingschools.org/Archives/13_03/eugenic.htm#5b

Stoskepf gives an historical overview of eugenics in the 20th century. He contends that though eugenic laws in America were repealed, eugenic ideology is still manifest in the public education sector, specifically in education reform—as evidenced by the first IQ tests, certain textbook material, and even college course offerings. Stoskepf traces, through historical analysis, specific ways eugenics has appeared (and been deemed highly valid and reliable) in educational environments. He also warns that the current high stakes testing practices might yield consequences similar to early educational (eugenic) reform.

- Category: Historical/Political/Legal Contexts
- Keywords: Eugenics; historical educational reform; current high stakes

Taylor, K., & Walton, S. (1997). Co-opting standardized tests in the service of learning. *Phi Delta Kappan*, 79(1), 66 - 70.

Although there is a growing trend toward performance-based assessments, norm-referenced, multiple-choice tests are still commonplace and are likely to remain so in the future. These test scores are used to make important decisions about students, but students who are accustomed to constructivist approaches to learning have little opportunity to experience conditions common to standardized, multiple-choice tests. The authors review their classroom-based study of ways to prepare students for standardized tests, while maintaining the integrity of the students, the school's curriculum, and the test. The study began with the premise that the norm-referenced, multiple-choice test is a particular type of literacy format, which students can learn to negotiate. Seventy-nine fourth and fifth grade students participated in a series of interactive workshops for a period of several weeks. The effectiveness of the workshop was assessed through student self-reports, performance on the standardized tests, and teacher notes. Mixed reactions of the community reflected the charged atmosphere surrounding standardized testing.

- Category: Curriculum and Instruction
- Keywords: Performance-based assessment; norm-referenced tests; constructivist methods; ethical considerations; literacy; community

Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan*, 82(5), 358 - 63.

High-stakes testing reform is viewed as a medium for satisfying political agendas in this piece—agendas that ultimately retract from learning and teaching experiences. Thompson compares and contrasts test-driven reform (high-stakes reform) to authentic, standards-based reform, and argues that no single exam should be used to evaluate a child's entire education.

- Category: Curriculum and Instruction; Test Development and Administration
- Keywords: Authentic reform; test-driven reform

Thurlow, M.L., & Johnson, D.R. (2000). High-stakes testing of students with disabilities. *Journal of Teacher Education*, 51(4), 305 - 314.

This article addresses the intended and unintended consequences of high stakes testing on students with disabilities. The authors take a detailed look at considerations that must be addressed for students with disabilities to participate in the current assessment system, including purposes, accommodations, and alternative assessment. Some intended consequences that they point to include: using test score information to modify curriculum, to modify pedagogy, and to address the strengths

and weaknesses of educators. Unintended consequences range from increased referrals to special education services to lowered expectations and narrowed curriculums. The advantages and disadvantages of diploma options are also examined in this piece.

- Category: Student Achievement/ Special Populations
- Keywords: Individualized Education Plans (IEPs); Individuals With Disabilities Education Act Amendments of 1997 (IDEA 97); diploma options

U. S. Department of Education Office for Civil Rights. (2001). *The use of tests as part of high-stakes decision-making for students: A resource guide for educators and policy-makers*. Washington, DC: Authors. <http://www.ed.gov/offices/OCR/testing/index1.html>

This resource guide is specifically aimed at providing educators and policy-makers with the necessary knowledge and tools to make accurate and fair high-stakes decisions about students based on test scores. The guide covers professionally sound test measurement principles as well as federal constitutional, statutory, and regulatory nondiscrimination principles. Both are intended to provide a framework for establishing nondiscriminatory testing principles especially when these are used to make placement, promotion and graduation decisions about students. Also included are glossaries of legal and test measurement terms; testing accommodations currently used by states; and a compendium that describes current federal nondiscrimination statutes and regulations relevant to testing issues.

- Category: Historical/Political/Legal Contexts; Student Achievement/ Special Populations; Test Development and Administration
- Keywords: Measurement principles; nondiscriminatory legal principles; test accommodations; promotion; graduation

Valencia, R. R., & Guadarrama, I. N. (1996). *High-stakes testing and its impact on racial and ethnic minority students*. In L. A. Suzuki, P. J. Meller, & J. G. Ponterotto (Eds.), *Multicultural assessment: Clinical, psychological, and educational applications* (pp. 561 – 610). San Francisco, CA: Jossey-Bass.

This chapter in the Handbook of Multicultural Assessment describes the high stakes testing campaign and its impact on racial and ethnic minority students. The authors approach the subject by beginning with background on the history of high stakes testing followed by a discussion of the controversies associated with high stakes testing and concluding with a discussion of alternative assessments to high-stakes testing, with special attention to the implications for assessment of minority children. The effects specific to minority students include: the disproportionate failure rate among minority students on state-mandated competency tests, the disparate, negative impact of teacher competency tests on minority students who seek teaching certificates, and the extreme negative impact of "school-based competency tests" of "take-over" (loss of accreditation, school closure, and subsequent receiver-ship by the state).

- Category: Student achievement/Special populations; Historical/Political/Legal contexts
- Keywords: Minority students; alternative assessments; CAT

Wheelock, A., Bebell, D. J., & Haney, W. (2000a, November 2). What can student drawings tell us about high-stakes testing in Massachusetts? *Teachers College Record*. <http://www.tcrecord.org/Content.asp?ContentID=10634>

This study explored students' reactions and opinions to taking the Massachusetts Comprehensive Assessment System (MCAS) through an examination of students' drawings of themselves taking the tests. 411 drawings gathered from 4th, 8th, and 10th grade students were used. Student depictions confirm the authors' beliefs that students do not respond in a uniform way to the expectations inherent in high-stakes assessment policies and practices. Several drawings and interpretations are described.

- Category: Student Achievement/ Special Populations
- Keywords: Massachusetts; MCAS; student motivation; student drawings

Wheelock, A., Bebell, D. J., & Haney, W. (2000b, November 2). Student self-portraits as test-takers: Variations, contextual differences, and assumptions about motivation. *Teachers College Record*. <http://www.tcrecord.org/Content.asp?ContentID=10636>

This study explored the educational and social factors that might explain the variety of students' responses to taking the Massachusetts Comprehensive Assessment System (MCAS). Several developmental differences in student responses are offered. The assumption that high stakes tests are necessary to motivate students is criticized. A variety of student attitudes and beliefs are offered. This report hopes to stimulate discussion of the assumptions surrounding high stakes testing policies and school-based practices.

- Category: Student Achievement/ Special Populations
- Keywords: MCAS; student motivation; student drawings

Whitford, B. L., & Jones, K. (2000). Kentucky lesson: How high stakes school accountability undermines a performance-based curriculum vision. In B. L. Whitford and K. Jones (Eds.), *Accountability, assessment, and teacher commitment: Lessons from Kentucky's reform efforts* (pp. 9 – 24). Albany, NY: State University of New York Press.

The Kentucky Education Reform Act (KERA) implemented in 1990 is often cited as a model that other states can learn from. These authors ask: "What lessons can be learned from Kentucky's experiences?" The authors contend it is too early to determine if KIRIS has improved student learning. There are, however, certain

unintended effects that require further research. KIRIS has jeopardized good teacher-student relationships; limited instructional practices; decreased teacher morals; and undermined the performance-based assessment approaches it was intended to promote.

- Category: Teachers and Teaching; Curriculum and Instruction
- Keywords: Kentucky; KERA; KIRIS; instructional strategies; performance-based assessment; accountability

Wideen, M. F., O'Shea, T., Pye, I. & Ivany, G. (1997) High-stakes testing and the teaching of science. *Canadian Journal of Education*, 22(4), 428 - 44.

This study examined the relationship between high-stakes testing and the teaching of science in 10 districts in British Columbia, Canada in regards to policy decisions at the district and school levels, teaching strategies in 8th, 10th, and 12th grades and the impact of high-stakes tests on curriculum and instruction. Data were obtained through classroom observations of 56 teachers in grades 8, 10, and 12 as well as through interviews of teachers, principals, district staff, and ministry officials. The authors concluded that high-stakes testing was leading teachers away from "exemplary" science teaching strategies by narrowing the range of instructional practices offered in the classroom.

- Category: Curriculum and Instruction; Teachers and Teaching
- Keywords: Authentic; active learning; constructivism; secondary level education

Wongbundhit, Y. (1996). Administration of standardized competency tests: Does the testing environment make a difference? *ERS Spectrum*, 14(2), 3 – 8.

This article outlines optimal conditions for helping students perform well on standardized tests. Wongbundhit describes the administration of tests in Dade County schools where the tests were given on Saturdays rather than during the school week in order to avoid overcrowding and distractions. The article presents a comparison of the school week administration in 1992 and the Saturday administration in 1993 in order to evaluate any differences in student performance, participation rates and test material security due to test administration factors. The study found that the Saturday administration provided optimal testing conditions with the effects of a positive impact on student performance; including in all subgroups. The study also found that the Saturday administration did not adversely affect the participation rates of students nor affect the security of test materials.

- Category: Test Development and Administration
- Keywords: HSCT; student performance; test participation rate; test material security; minority groups

Yeh, S. S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher*, 30(9), 12 – 17.

Yeh argues that since test designers consider tests to lead instructional strategies, then it follows that developing test questions that foster critical thinking skills could push teaching in such a direction. He proposes conceptualizing critical thinking as 'careful argumentation' and suggests strategies for designing test items that push instruction in ways that would involve critical reading and discussion as preparation.

- Category: Test Development and Administration; Curriculum and Instruction
- Keywords: Critical thinking; test item construction; teaching to the test

Yen, W. M., & Ferrara, S. (1997) "The Maryland school performance assessment program: Performance assessment with psychometric quality suitable for high stakes usage". *Educational and Psychological Measurement*, 57(1), 60 – 84.

Implemented in 1991, the Maryland School Performance Assessment Program (MSPAP) is a performance-based testing program covering reading, writing, language usage, mathematics, science, and social studies. This study was conducted to ascertain whether the MSPAP has the psychometric characteristics necessary for use in high stakes decision-making. An outline of the scoring process, test difficulty, score validity, construct and consequential validity is provided as evidence that it does.

- Category: Test Development and Administration
- Keywords: Psychometrics; scaling; equating; score accuracy; validity