

Running head: PREDICTED VALIDITY OF TAKS SCORES

Predicted Validity of TAKS Scores

Kaye Rogers

University of North Texas

September 28, 2007

Review of Literature

The No Child Left Behind Act (NCLB), signed into law in 2002, created a system for nation-wide accountability with serious sanctions for underperforming schools and school districts. NCLB strengthens a movement toward accountability in education that has been gathering momentum for nearly a decade. Now in its fifth year, the sanctions imposed by this law are still implemented, and the discussions regarding the intended and unintended consequences are regularly seen in the media. High-stakes testing in the United States is more widespread than ever before, and our nation relies on ability and achievement testing more than any other nation for making important decisions about individuals and schools (Dillon, 2005). We live in an era where there is strong support for public policies that use test results to compel changes in the behavior of students, teachers, and school administrators (Steeves, 2002). The tests are seen by some as the perfect policy mechanism because they are both effectors and detectors: they are intended to effect and cause change in the system, and then detect whether changes in the system actually occur. NCLB sanctifies this approach to school change.

FairTest (2002) recently noted that the United States already tests more children more often than any other nation.

The vision of a nation-wide, standards-based accountability system appears to be drawn from the content standards discussions at the state and national levels. As various groups worked to reach consensus on what is important for all students to know and be able to do if they are to be successful in the 21st century, others worked to design systems to hold everyone, in the system-from policymakers to educators to students-accountable for meeting those standards (Herman, 2004).

One piece needed to move from content standards to accountability was a form of measurement. Standards-based accountability (SBA) in education entails the development of content and performance standards, the administration of assessments aligned with the standards and the imposition of consequences tied to assessment results (Brewer et. al., 2007).

Accountability requires standards-based assessments to ensure students are learning what is expected. NCLB took the accountability movement even farther, focusing not only on “what” is expected (the content standards), but also on “how well” it should be learned (performance standards) (Linn, Baker & Herman, 2002).

Standards-based assessments are supposed to provide valuable information to schools and policymakers by measuring the status and progress of student learning. The test results are intended to support important insights on the nature, strengths and weaknesses of student progress relative to the standards, and educators are expected to use this feedback to make appropriate curriculum and instruction decisions as they redirect their efforts toward improving the learning of the student (Herman, 2004). Herman states that policy makers try to strengthen the accountability aspects of the system by establishing specific goals for school performance and attaching incentives and sanctions to achieving, not achieving or surpassing these goals. Tracing the role of high stakes testing, public education for the past 20 years shows that the system of accountability has not always possessed specific and stringent goals, but the attached sanctions and incentives define in many ways where the standards and accountability movement is now placed significant consequences and urgency to obtain goals set in place. (Sloan & Kelley, 2003).

While many non-educators may be unaware of the specific content and performance standards in the public schools in their states, these performance standards become more evident

to stakeholders as large-scale assessments to measure them are developed and implemented. Originating from the state and/or local level, the assessments make explicit what content is to be learned, and as performance levels and minimum passing scores are established, they set parameters for how well students and schools must perform to meet the standards. The assessments, therefore, become a primary vehicle for communicating what the standards really mean, and they provide strong signals to teachers and schools about what they should be teaching and what students should be learning (Taylor, Shepard, Kinner, & Rosenthal, 2003). . A unique feature of standards-based assessment is the intention not only to signal to teachers what to teach but also, with the multiple types and forms of assessment, to provide clues as to how to teach it. (Resnick & Resnick, 1985).

There are a number of qualified individuals who passionately speaking out against high-stakes testing; Donald Campbell is one such individual who uses a well-established social science law to support arguments. Campbell's (1975) law has two parts: the first which is concerned with the validity of the indicators used, and the other concerned with the organizations and the people that work with the indicators when they take on exceptional value. Campbell states, that the more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor. Madaus (2001) points out that Campbell has given the social sciences a version of the Heisenberg uncertainty principle. That principle, concerned with measuring the position and velocity of objects, informed physicists that if they measure one of these conditions they could not accurately measure the other at the same time. Madaus' version of the uncertainty principle with regard to Campbell's Law states that if high-stakes tests are used to assess students, teachers or schools, the corruptions and distortions that inevitably appear

compromise the construct validity of the test. As the stakes associated with the test go up, so does the uncertainty about the meaning of a score on the test. Basically, in high-stakes testing environments, the greater the pressure to do well on the tests the more likely the score obtained by the students or schools is uninterpretable. With NCLB's emphasis on the subjects of math, science, English/reading and social studies, there are voiced concerns about the de-emphasis of instruction in other areas like art, music, physical education and technology (Chapman, 2005). Standardized-test scores and other variables used for judging the performance of teachers and school district performance have placed pressures on teachers, students, administration and districts to narrow instructional emphasis on the high stakes content standards, cheating and corruption (Magnuson, 2000).

Life altering decisions are made on the basis of high-stakes tests, such as promotion to a higher grade or retention in grade. Tests can determine who will receive a high-school diploma and who will not. Tests scores can determine if a school will be reconstituted with job losses for teachers and administrators when scores do not improve or cash bonuses when scores do improve. Therefore, test-givers should be certain that the construct measured by those who take tests with serious consequences attached to them is the construct that was intended (Madaus & Clarke, 2001). Too much uncertainty about the meaning of a score on a test can be psychometrically, morally and sometimes legally inappropriate. It also violates the standards that professionals and their professional organizations have agreed to use when constructing and administering tests (Kreitzer, Madaus, & Haney, 1989). There is a larger body of research, however, that suggests such exit tests are related to an increase in the numbers of students dropping out, particularly for students already at risk (Catterall, 1989). In one of the most recent large-scale studies on this issue, Haney (2000) studied the impact of the Texas Assessment of

Academic Skills (TAAS) on school completion in Texas and found evidence to suggest that the exit exam was associated with an increase in dropout rates, especially among African Americans and Hispanics. High-stakes testing, then, may increase the numbers of students leaving high school without a diploma: a minimum certification necessary in today's labor market (Madaus & Clarke, 2001).

Background on Standards and Accountability Measures

There was little concern about tying high stakes outcomes to testing prior to the 1970s. The federal government and the states used large-scale tests to monitor the status of the educational system and to provide information that might be helpful to teachers and large groups of students. The National Assessment of Educational Progress (NAEP), the only large scale federally commissioned achievement test, was designed solely with a monitoring role in mind (Stecher, 2002).

State Accountability Measures

Beginning with the minimum competency testing movement in the 1970s, policymakers began to use test results in new ways—specifically as the basis for decisions about individual performance (Herman, 2004). Tests grew more common in the 1980s, and the rationale for large-scale testing expanded from judging the performance of groups of students to influencing teaching practice (Popham, 1987). In 1983, The National Commission on Education released *A Nation at Risk*, which called for an end to the minimum competency testing movement and the commencement of a high stakes testing movement that would raise the nation's standards of achievement in an unprecedented manner. Although history has questioned the accuracy of the report, it did argue persuasively that schools in the United States were underperforming in comparison to other countries and that the United States was in jeopardy of losing its standing in

the world. Citing declines in national and international student test scores, deterioration in school quality, a diluted and unfocused curriculum, and setbacks in other indicators of U.S. superiority, the document created a nation-wide panic around the perceived weakening condition of the American education system (Amrein & Berliner, 2002).

Regardless of questions about the report's scholarly credibility, *A Nation at Risk* had its effects. The National Commission on Education called for more rigorous standards and accountability mechanisms to lift the United States out of its purported educational recession. The Commission recommended that states institute high standards to homogenize and improve curricula and conduct rigorous assessments to hold schools accountable for meeting those standards. In some ways, *A Nation at Risk* was the initial catalyst for standards-based reform in public education (Amrein & Berliner, 2002).

Researchers began to investigate teachers' reactions to external assessment in light of the changes that occurred in the uses of large-scale testing in the 1980s and 1990s. One criticism of high-stakes testing during the 1980s was that the emphasis on minimal competency levels for students resulted in schools teaching directly to these minimal competencies rather than the broader curriculum (Sloane & Kelly, 2003). In the mid-1990s, states began to implement statewide, test-based accountability systems, prompting renewed interest in the effects of testing on the practice of teaching. Large-scale studies of test validity and the effects of testing were conducted in a number of states that implemented such accountability systems.

With the introduction of formal accountability systems in the 1990s, policymakers embraced a new, more potent vision for the role of assessment. They envisioned tests, often in combination with standards, as a mechanism to influence changes in practice, hoping to use them to exert a strong positive effect on schooling (Stecher, 2002).

States across the country are using testing programs with added dramatic incentives for students. A growing number of states are adopting policies that require students to meet a performance standard in order to be promoted to the next grade or to be granted a high school diploma (Herman, 2004). The system is intended to promote a continuous improvement model aimed at enabling all children to reach the standards: establish and monitor goals and benchmarks, assess progress, then use results on goal attainment to improve performance (Herman, 2004).

The majority of states in the United States has implemented or is in the process of implementing school accountability systems as a central component of efforts to improve student achievement (Borko, 2005).

The increase in states' use of academic standards and tests as educational policy tools has occurred rapidly. The number of states with mandated testing programs grew from twenty-nine in 1980 to forty-six in 1992. By 2001, forty-eight states were in the process of developing or had already implemented standards-based assessments, and thirty-three had accountability systems that hold students, teachers, school administrators, and/or district administrators responsible for student performance (Stecher & Barron, 1999).

Academic standards and tests have become even more central to state educational reform efforts with the passage of the No Child Left Behind (NCLB) Act of 2001 (Borko, 2005). McDonnell (1994) explained state policy makers intend for the assessment system and the policies linked to its use to shape not just student outcomes, but also what and how students are taught.

The most recent research on large-scale assessment has focused on the quality of scores and their appropriateness for accountability purposes (Linn, 2003). Much less attention has been

paid to the effects of standards-based assessments on school and classroom practices (Borko, 2005). Leading researchers and advocates for standards based reform emphasized the contrast between the new reform calling for high standards and previous reform efforts aimed at minimum competencies and basic skills. Politicians were no longer tolerant with the idea that profound changes in the educational system would require concerted effort over a long period of time (Taylor, 2003).

Newer descriptions of what standards-based reform is all about emphasize “accountability” as the necessary lever to create incentives for teachers and schools to attend to high standards and implement new curricula. Researcher Frederick Hess, a proponent of high stakes accountability, argued that for accountability to have a significant effect on educational quality, “educators must be rewarded or sanctioned on the basis of student performance” (Hess, 2002, p. 73). He recognized that such a system runs counter to the traditional values of the American educational system, which relied on the good will of teachers and the intrinsic motivation of students, insisting only a coercive accountability system is effective in transforming the quality of public schooling.

Federal Accountability Measures

Also with NCLB comes annual testing in the “core” areas of reading, mathematics and science. The law states that by the year 2005 – 2006 testing is required in reading and math in grades three through eight and once in high school. NCLB substantially increases the testing requirements for states and sets demanding accountability standards for schools, districts, and states. It does this by requiring each state to set measurable adequate yearly progress (AYP) objectives for all students taken as a whole and for subgroups of students as well. These subgroups are defined as students from economically disadvantaged families, students from

state-identified major racial or ethnic groups, students with limited English proficiency, and students with disabilities.

Each state department of education must certify that its definition of Adequate Yearly Progress presents, "the same high standards of academic achievement to all public elementary school and secondary school students in the state; is statistically valid and reliable; [and] results in continuous and substantial academic improvement for all students" (NCLB, 2001, Sec. 1111 (a)(2)(C)). NCLB requires each state to have consistent standards for all the students in its public schools. After establishing the standards, states must also identify, develop, and disseminate high quality, effective curricula aligned to those standards. NCLB also mandates that all states require public schools to administer statewide assessments in grades three through eight and in high school. The tests must be rigorous, demanding, academic, secular, neutral, and non-ideological (Chapman, 2005). To make Adequate Yearly Progress (AYP), schools must ensure at least 95% of all students and of the students in each subgroup are tested. Perhaps in recognition of the varied assessment measures from state to state, NCLB requires states to participate in NAEP tests in reading and mathematics at grades four and eight. NAEP tests are required to help identify states in which standards or state assessments in these subjects are not sufficiently rigorous (Porter, 2002).

The centerpiece of the performance standard aspects of NCLB is the concept of AYP, which is a measurable target for improved test scores in reading, mathematics, and science. Scores must be tracked in relation to students' race, ethnicity, gender, socio-economic status, status in special education and limited English proficiency. AYP data must be summarized for each subject, grade level, school, and subgroup of students. The data in these reports flow upward from the district to the state level, then to the federal level. Federal officials report to

Congress on degrees of compliance with NCLB (Chapman, 2005), and although the critical differences make it ill advised, comparisons between states occur often in the popular media. Popho (2000) summarized samples of newspaper reports of statewide assessment results in Massachusetts, Texas, California, and Colorado. Reporters' accounts included accusations in Boston of teachers cheating to improve their students' test scores, and the California Department of Education ranking its school districts, which led to debates on opportunities in the affluent versus socio-economically disadvantaged systems throughout the state.

While the use of high-stakes testing is becoming more common, the overall picture of state testing programs remains quite varied. The research conducted on the implementation and impact of state testing systems reflects this interstate and intrastate variability. State studies have been largely unsystematic and have involved testing programs with different stake levels for students and schools or different testing formats such as multiple-choice or performance-based. Research efforts have also typically examined the role of test preparation and the relationship between the state test and the content standards (Pedulla, et. al, 2003; National Board on Educational Testing and Public Policy (NBETTP), 2003).

A growing body of evidence suggests that high-stakes testing can be a driving force behind fundamental change within schools (Koretz, Linn, Dunbar, & Shepard, 1991; Hoffman et. al., 2001; Taylor et al., 2003). However, there is a difference of opinion as to whether this change carries a positive or negative impact. For example, while some feel that the granting of rewards or the threat of sanctions is essential for promoting quality teaching and encouraging higher student achievement, others have found that high-stakes testing limits the scope of classroom instruction and student learning in undesirable ways (Stecher, Barron, Chun, & Ross, 2000; Wright, 2002).

Although there is differing opinion among scholars regarding the impact of high-stakes testing on student achievement, there are virtually no studies that have investigated the predictive validity of high-stakes exams, especially exit-level exams, on career performance beyond high school. Therefore, the purpose of this study is to determine if exit-level exams have predictive validity among graduates regarding college acceptance rates, career choice, and performance in their career choice beyond high school.

Method

Participants

The participants in this study included 449 graduates from a large urban high school in Texas. Among the participants, 50.4% were identified as female, while 49.6% were identified as male. Regarding ethnicity, 93 % were Anglo, 1% were African-American, 4.6% Hispanic and 6.9 % Other ethnicities. Among the participants, 27.9% were accepted at a Tier I university, < 1% were accepted at a Tier 2 university, 13.3% were accepted at a Tier 3 university, 13.3% were accepted at a Tier 4 university, 43% of students were accepted to private universities, Tier 5 universities or junior colleges, 5.4% entered the workforce or entered the military immediately following graduation. In Texas, the exit-level exam is the Texas Assessment of Academic Skills (TAKS) exam that is administered to measure how well students have progressed academically. The TAKS is aligned with the state-mandated curriculum, the Texas Essential Knowledge and Skills (TEKS). Each student must pass the exit-level TAKS to receive a Texas high school diploma.

References

- Abrams, I. M., Madaus, G. F. (2003). The lessons of high stakes testing. *Educational Leadership*, 61(32), 31-35.
- Amrein, A.L., & Berliner, D.C. (2003). The effects of high stakes testing on student motivation and learning. *Educational leadership*, 60(8), 32-38.
- Borko, H. (2005, February). *The impact of state accountability on classroom practices*. Washington, DC: Workshop on incentives and test-based accountability for The National Academics Center for Education Board on Testing and Assessment.
- Brewer, D., Augustine, C., Zellman, G., Ryan, G., Goldman, C., Stanz, C., Constant, L. (2007). Education for a New Era: Design and Implementation of K-12 Education Reform in Quatar, Rand Corp., Santa Monica, CA
- Campbell, D.T. (1975). Assessing the impact of planned social change. In G. Lyons (Ed.), *Social research and public policies: The Dartmouth/OECD Conference*. (Chapter 1, pp. 3-45). Hanover, NH: Dartmouth College, The Public Affairs Center.
- Chapman, Laura H. (2005). No Child Left Behind in Art? *Art Education*, 58(1), 6-16.
- Catterall, J.S. (1989). Standards and school dropouts: A national study of tests required for graduation. *American Journal of Education*, 98(1), 1-34.
- Dillon, S. (2005). Facing state protests: U.S. offers more flexibility on school rules. *New York Times.Com* Retrieved January 18, 2007, from <http://www.nytimes.com/2005/04/08/education/08child.html>
- FairTest: The National Center for Fair & Open Testing. 2002. Will more testing improve schools? Fact sheet. Cambridge, Mass.: FairTest. Available at: <http://www.fairtest.org/facts/Will%20More%20Testing%20Improve%20Schools.html>.

Herman, J.L. (2004). The effects of testing on instruction. In S. H. Fuhrman & R.F. Elmore (Eds.), *Redesigning accountability systems for education* (pp.141-161). New York: Teachers College Press.

Hess, F. M. (2002). Reform, resistance...retreat? The predictable politics of accountability in Virginia. In D. Ravitch (Ed.), *Brookings paper on education policy 2002* (pp. 69-122). Washington DC: Brookings Institution.

Koretz, D., Linn, R., Dunbar, S., & Shepard, L. (1991). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests*. Chicago, IL: Annual Meeting of the American Educational Research Association.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*. 32(7), 3-13.

Linn, R. L., Baker, E.L., & Herman, J.L. (2002). *The CRESST Line*, 1-3 (Newsletter of the National Center for Research on Evaluation, Standards, and Student Testing [CRESST}, University of California, Los Angeles)

Madaus, G. & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from one hundred years of test data. In G. Orfield & M. L. Kornhaver (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Foundation Press.

McDonnell, L. M. (1994). Assessment policy as persuasion and regulation. *American Journal of Education*, 102, 394-420.

Pedulla, J. J. (2003). State mandated testing: What do teachers think? *Educational Leadership*, 61(3), 42-46.

- Pipho, C. (2000). The sting of high-stakes testing and accountability. *Phi Delta Kappan*, P. 645.
- Popham, W.J (1987). The merits of measurement driven instruction. *Phi Delta Kappan*, 68, 697-682.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.
- Resnick, D. P. & Resnick, L. B. (1985). Standards, curriculum and performance: a historical and comparative perspective. *Educational Researcher*, 14(4), 5-20.
- Sloan, F. C. & Kelley A. E. (2003, Winter). Issues in high stakes testing programs. *Theory Into Practice*, 42(1), 12-17.
- Stecher, B. M. (2002). Consequences of Large-scale, high-stakes testing on school and classroom practice. In L.S. Hamilton, B.M. Stecher, & S.P. Klein (Eds). *Making sense of test-based accountability in education* (pp. 79-100). Santa Monica, CA: RAND.
- Stecher, B. M. (1999). *Quadrennial milepost accountability testing in Kentucky* (CSE Tech. Rep. No 505). Los Angeles, CA: University of California, National Center for Research on Evaluation , Standards and Student Testing.
- Stecher, B., Barron, S. L., Chun, T., & Ross, K. (2000). *The effects of the Washington State education reform on schools and classroom*. (CSE Tech. Rep. No. 525). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards and Testing.
- Steeves, Kathleen Anderson "Are we measuring student success with high-stakes testing?". Educational Forum, The. Spring 2002.
- http://findarticles.com/p/articles/mi_qa4013/is_200204/ai_n9078276

Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). *A survey of teachers' perspectives on high stakes testing in Colorado: What gets taught, what gets lost* (CSE Technical Report 588). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

The National Council on Measurement in Education. (NCME) and the American Psychological Association (APA) joined with the American Educational Research Association (AERA) to define the standards for test construction and use. These are available at AERA.net