

The Lingering Influence of High-Stakes Testing Policies on the Postsecondary Prospects of Latino Students

Author(s): *David R. Garcia*

Affiliation: Mary Lou Fulton Teachers College, Arizona State University

2011

Abstract

This manuscript presents a unique perspective on the long-term effects of accountability policies by examining the relationship between high school graduation tests and postsecondary readiness using a set of hypotheses based on two psychometric concepts: construct validity (the transference of knowledge on tests to other non-test settings) and consequential validity (curriculum narrowing). The findings illustrate the lingering influence of curriculum narrowing to graduation tests that have little transference to postsecondary requirements. As high schools prepare students to pass graduation tests they may diminish students' postsecondary prospects. The influence of curriculum narrowing was strongest among predominantly Latino high schools, even those that met accountability requirements.

Introduction

Federal initiatives such as *Race to the Top* and the Common Core Standards are intended to raise academic expectations and prepare students for "success in college and careers" (Common Core Standards Initiative, 2010). These ambitious goals hinge upon elevating student academic achievement in the K-12 system where student achievement is largely defined by standardized test scores. What are the long-term consequences, however, when high-stakes standardized tests are not related to postsecondary outcomes?

This manuscript presents a unique perspective on the long-term effects of high-stakes testing by using two psychometric concepts (construct validity and consequential validity) to posit a set of hypotheses about the relationship between high stakes tests and postsecondary readiness. The findings and discussion sections focus on the implications of the study for predominantly Latino high schools.

Using Construct and Consequential Validity to Create Hypothesized Relationships between High-Stakes Graduation Tests and Postsecondary Readiness. Construct validity pertains to the extent to which high-stakes tests are associated with external indicators of academic achievement. "Establishing construct validity requires a network of empirical research that shows a high correlation of the measure with other indicators that theoretically should relate to it" (Smith & Fey 2000, p. 337). The underlying assumption of high-stakes graduation testing policies is that students have acquired knowledge that is transferrable to a domain other than the test itself, such as postsecondary entrance requirements. To the extent that the skills and knowledge on graduation tests are transferable to other academic indicators then students who pass the graduation tests should also be ready for postsecondary education, regardless of the academic standing of the high school they attended.

As one considers the broader impact of high-stakes testing policies the focus shifts to a question of consequential validity (Messick, 1995). "Establishing consequential validity requires the accumulation of evidence and a rational argument that students who take the test are benefited and not harmed as a result of actions that are tied to test results" (Smith & Fey 2000, p. 337). Curriculum narrowing is a particularly damaging outcome of high-stakes testing because this practice is often associated with rigid teaching methods, "drill and kill" strategies, standardization, and limited course offerings (Diamond & Spillane, 2004; Mintrop, 2003; McNeil 2000). Traditionally low-performing schools have faced the most pressure from accountability policies and have been the most likely to engage in curriculum narrowing activities such as replacing substantive curriculum with test preparation activities (McNeil & Valenzuela, 2001).

The concepts of construct and consequential validity can be combined to set forth a series of hypotheses about the relationship between high stakes graduation tests and postsecondary readiness. If the graduation test has a high degree of construct validity, then one should expect a weak relationship between school level test scores and postsecondary readiness regardless of the degree of curriculum narrowing at the school level. This pattern is evident because in preparing students to pass the graduation test, all schools (even low-performing schools) engage in pedagogical activities that simultaneously prepare students for postsecondary education (see Figure 1).

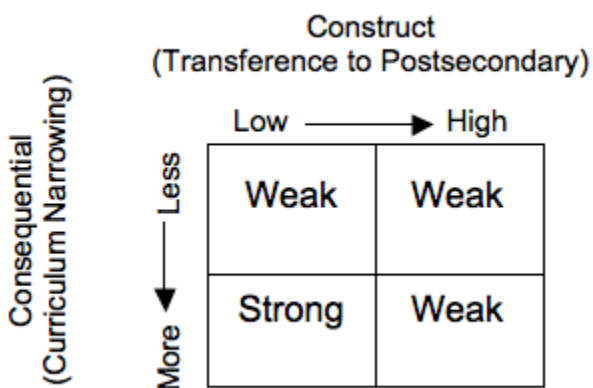


Figure 1. Hypothesized relationships between high-stakes graduation test results and postsecondary readiness.

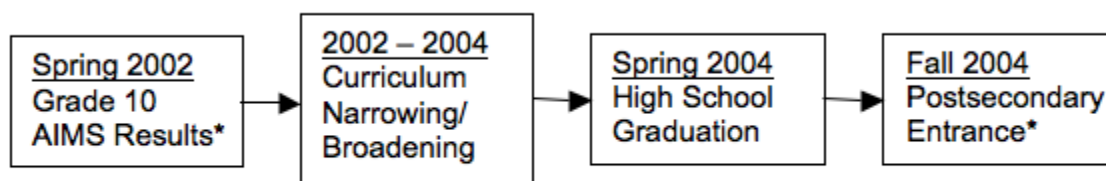
If the graduation test has low construct validity, then the relationship between curriculum narrowing and the percentage of students ready for postsecondary education strengthens. Schools with lower test scores will have lower rates of postsecondary readiness and vice versa. A stronger relationship arises because as low-performing schools focus on students passing the graduation test they are pressured to narrow the curriculum and place more instructional emphasis on test-related skills and knowledge that have little transference to preparing students for postsecondary education. Conversely, schools with

high graduation test scores do not experience the same pressure to narrow the curriculum and are able to broaden the curriculum to focus on content with greater transference to postsecondary education.

Methods

The primary methodological challenge to studying the impact of high school graduation tests at the postsecondary level is the ability to track students over time. Large-scale longitudinal tracking systems that span the P-16 sectors are rare and even with such systems, one must contend with a natural attrition problem. Those students most able to attend a postsecondary institution are those most likely to pass the graduation tests in the earliest administrations, leaving a multi-year gap between the final test administration and entrance into a postsecondary institution. My approach is to use a quasi-cohort tracking method that follows an aggregate group, such as a graduating class, over time.

Design. To create a quasi-longitudinal cohort for the Class of 2004, data were collected at two points in time when all students were available. The first data collection point was the 10th grade census administration of the Arizona Instrument to Measure Standards (AIMS) high school test in spring 2002 (see Figure 2). The second data collection point occurred in fall 2004 when the cohort Class of 2004 entered a postsecondary institution (state universities or community colleges).



* Data collection point

Figure 2. Timeline for the Class of 2004

Data. The study included two primary data sources. The first data source was 10th grade, school level results from the AIMS in reading and mathematics for the 2002 academic year. The second data source was postsecondary readiness results for the year 2004-2005 academic year derived from a report required by Arizona State Statute (Arizona Revised Statutes §15-1822).

The high schools under study are located in Maricopa County, which includes the Phoenix metropolitan area. In total, the study included 79 high schools that graduated a total of 16,208 students in the Class of 2004. Overall, 57% of the students in the Class of 2004 matriculated into one of the Arizona community colleges or universities.

Analytical groupings. The high schools were organized into two groups using the average percentage of Latino students (29%) across all county high schools as the divider to identify predominantly Latino high schools and predominantly White high schools. In addition, predominantly Latino and predominantly

White high schools with the same school accountability rating were compared to each other in order to hold the pressures of school accountability policies constant. Finally, the following scale was used to interpret the correlation coefficients: weak (0.0 to 0.4), moderate (0.41 to 0.60), strong (0.61 to 0.80) and very strong (0.81 to 1.0).

Findings

On the AIMS graduation test, predominantly White high schools outperformed predominantly Latino high schools. For example, in the sophomore year, 47% of students in predominantly White high schools passed the 10th grade AIMS mathematics test compared to 22% of students in predominantly Latino high schools (see Table 1). In addition, 74% of the predominantly White high schools were given Arizona's highest school performance ratings (*excelling* and *highly performing*) compared to only 9% (n=3) of predominantly Latino high schools.

A higher percentage of students from predominantly White high schools entered a postsecondary institution. When the comparison is narrowed to *performing* schools, however, the postsecondary entrance gap nearly disappears. On average, 50% of students from predominantly White *performing* high schools entered a postsecondary institution compared to 46% of students from predominantly Latino *performing* high schools ($t_{79} = 1.99, p > .10$).

Table 1
School Demographic and Academic Statistics, by High School Type

	Predominantly White	Predominantly Latino
High schools (n)	(47)	(32)
Student demographics		
Hispanic	0.13	0.52
White	0.80	0.35
Other minority	0.08	0.14
Grade 10 AIMS (meets/exceeds)		
Math	0.47	0.22
Reading	0.78	0.53
School accountability ratings		
Underperforming	0.00	0.03
Performing	0.26	0.88
Highly Performing	0.38	0.06
Excelling	0.36	0.03
Postsecondary entrance	0.61	0.47
Postsecondary readiness		
Math	0.46	0.25
English	0.79	0.59

What is the relationship between graduation test results and postsecondary readiness? The relationship between the percentage of students passing the 10th grade mathematics graduation test and postsecondary readiness two years later was strong and positive ($r = .67, p < .01$). The strength and directionality of the variables suggest that the remnants of curriculum narrowing can be tracked into students' postsecondary experiences. The results were similar for English.

What is the relationship between predominantly Latino and predominantly White high schools? The relationship between graduation tests and postsecondary readiness is weaker for predominantly White high schools than for predominantly Latino high schools. For predominantly White high schools, the percentage of students that passed the graduation test had a weak association with postsecondary readiness ($r = .40, p < .01$). For predominantly Latino high schools the association between graduation test results and postsecondary remained positive and strong ($r = .73, p < .01$). In other words, the potentially negative remnants of curriculum narrowing on postsecondary readiness were more detectable for predominantly Latino schools only. The English results exhibited a pattern similar to mathematics.

Table 2

Correlations between High School Test Scores and Postsecondary Readiness by Type of High School

	Predominantly White	Predominantly Latino
High schools (n)	(47)	(32)
All Schools		
Math	0.40	0.73
English	0.56	0.87
Performing Schools	(12)	(28)
Math	0.58	0.76
English	0.45	0.75
Excelling/Highly Performing Schools	(35)	(3)
Math	0.47	--
English	0.20	--

What is the relationship holding the pressure of school accountability constant? The primary comparison of interest is among *performing* schools that experienced a similar degree of public pressure to score well on the graduation tests. Among *performing* schools, the relationship between the graduation tests and postsecondary readiness continued to differ between predominantly Latino and White high schools. The predominantly Latino *performing* high schools maintained strong relationships (mathematics $r = .76, p < .01$; English $r = .75, p < .01$). Among predominantly White *performing* high schools, conversely, the relationships were moderate in mathematics ($r = .58, p < .05$) and English ($r = .45, p < .05$). Interestingly, the relationships were weakest (as hypothesized) among the *highly performing* and *excelling* predominantly White high schools where one would expect less curriculum narrowing (mathematics: $r = .47, p < .05$; English: $r = .20, p > .20$).

Discussion

In order to put the findings in proper context, it is important to remember that the student subpopulation tracked in this study represents the most academically prepared students that followed the traditional path from high school to a postsecondary institution. Despite preparing for and presumably passing the same graduation test and meeting the same graduation requirements, a higher percentage of students from high schools with higher percentages of students passing AIMS were prepared for postsecondary education than students from high schools with lower test scores. This pattern is likely an artifact of Arizona's high-stakes test, which has a low degree of construct validity (AIMS Task Force, 2009) and curriculum narrowing as a result of accountability pressures.

The consistently strong relationships, however, held for predominantly Latino high schools only. Thus, as state policies pressure low-performing Latino high schools to focus on the high-stakes test, these mandates also force schools to educate students to counterproductive ends, at least as it applies to student postsecondary prospects.

The most striking finding is that among *performing* high schools, the differences between predominantly Latino and predominantly White high schools remained, even though one would not expect *performing* Latino schools to engage in curriculum narrowing. One possible explanation is that in the absence of other meaningful academic feedback, predominantly Latino high schools rely upon the test scores as a legitimate indicator of student readiness for postsecondary education.

Conclusion

As part of the larger education policy discourse, this study makes two noteworthy contributions. First, encouraging postsecondary completion while at the same time skirting how graduation tests divert students from success at the postsecondary level is a counterproductive policy orientation. Incongruent policies encourage policy churn (Hess, 1999) and perpetuate educational inequality. Second, the study is an argument to broaden the public discussion of which indicators count as a measure of academic performance in accountability systems. Many have called for multiple measures to expand the criteria that are used to make decisions about schools and students (Valenzuela, 2002; American Educational Research Association, 2000). The unenthusiastic response may be due to fact that the recommended measures (e.g. grades, portfolios, site visits) hold little credibility among policymakers. Postsecondary readiness indicators fit the popular notion of improving a state's educational system in order to acquire economic advantages. Simultaneously, when high schools are expected to educate students to succeed beyond the secondary context, they are encouraged to broaden rather than narrow the academic curriculum.

References

AIMS Task Force (2009). *Report and Recommendations of the AIMS Task Force*. Phoenix, AZ.

American Educational Research Association. (2000). AERA position statement on high-stakes testing in Pre-K -- 12 education. Retrieved from <http://aera.net/policyandprograms/?id=378>.

Arizona Revised Statutes. Vol. §15-1822. *Report; academic performance of high school graduates*.

Booher-Jennings, J. (2005). Below the bubble: "Education triage" and the Texas accountability system. *American Education Research Journal*, 42(2), 231-268.

-

Common Core Standards Initiative. (2010). *Common Core State Standards*. Retrieved from <http://www.corestandards.org>.

Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record* 106(6) 1145-1176.

Hess, F. M. (1999). *Spinning Wheels: The Politics of Urban School Reform*. Washington, DC: Brookings Institution.

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

McNeil, L., & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in a public education*. New York: The Century Foundation Press.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 10(1), 5-8.

Mintrop, H. (2003). The limits of sanctions in low-performing schools: A study of Maryland and Kentucky schools on probation. *Education Policy Analysis Archives*, 11(3). Retrieved from <http://epaa.asu.edu/epaa/v11n3.html>.

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

Valenzuela, A. (2002). High-stakes testing and U.S.-Mexican youth in Texas: The case for multiple compensatory criteria in assessment. *Harvard Journal of Latino Policy* 14, 97-116.