

Fall 10-23-2009

# High Stakes Testing Literature Review and Critique

Youness Elbousty

*Lynn Public Schools, elbousty@yahoo.com*

Follow this and additional works at: [http://digitalcommons.uconn.edu/nera\\_2009](http://digitalcommons.uconn.edu/nera_2009)



Part of the [Education Commons](#)

---

## Recommended Citation

Elbousty, Youness, "High Stakes Testing Literature Review and Critique" (2009). *NERA Conference Proceedings 2009*. 16.  
[http://digitalcommons.uconn.edu/nera\\_2009/16](http://digitalcommons.uconn.edu/nera_2009/16)

Running head: REVIEW OF THE LITERATURE ON HIGH-STAKES TESTING

Review of the Literature on High-Stakes Testing

Youness Elbousty

Lynn Public Schools (MA)

Paper presented at the Annual Meeting of the Northeastern Educational Research Association, Rocky Hill, CT, October 21-23, 2009.

Address correspondence to: Youness Elbousty, Lynn Public Schools, 235 O'Callaghan Way, Lynn, MA 01905,

e-mail: [elboustyy@lynnschools.org](mailto:elboustyy@lynnschools.org), or [elbousty@yahoo.com](mailto:elbousty@yahoo.com);

### Abstract

Standardized testing has been long established in most of the schools in United States. States have attached "high stakes" to tests as a response to the federal law NCLB. Under this law, schools had to develop or alter their assessments which are administered to gauge school progress. While many agree that high-stakes testing has an impact on students; studies have been conducted to vet whether such impact has propitious or harmful outcomes. In this paper, I review and critique the literature on high stakes testing coupled with a close scrutiny of the research methods utilized in the articles under review.

Standardized testing has been long established in most of the schools in United States. In the 1970s and 1980s, many schools adopted a competency test which students had to pass to earn their high school diplomas. States have attached increasingly higher stakes to these tests as a response to the federal law No Child Left Behind. Under this law, schools had to develop or alter the assessments they administer to gauge school progress. As a result of these tests, schools can be either rewarded or levied harsh sanctions. While many agree that high-stakes testing has an impact on students, numerous studies have been conducted to vet whether such impact has had propitious or harmful outcomes. In this paper, I review the literature on high stakes testing by reporting the claims in the existing literature coupled with a close scrutiny of the research methods utilized in the articles under review.

Amrein and Berliner (2002a) conducted an analysis of 18 states to examine the effect of high stakes testing on students' learning. Since these tests can be manipulated by the states, these two researchers found it imperative to measure the growth in learning by examining four other standardized tests: ACT, SAT, NAEP, and AP, to find conclusive evidence about whether high stakes testing increases students' learning. The uncertainty principle, the concept that precise simultaneous measurement of some complementary variables is impossible, was used to interpret the data. Amrein and Berliner came to the conclusion that there was no clear evidence of students' learning even if the students' scores of the previously mentioned tests go up. Amrein and Berliner state that even if:

we assume that the ACT, SAT, NAEP, and AP tests are reasonable measures of the domains that a state's high-stakes testing program is intended to affect then we have little evidence at the present time that such programs work. Although states may demonstrate

increases in scores on their high-stakes tests, transfer of learning is not a typical outcome of their high-stakes testing policy. (p.52)

One might understand the reasoning behind these claims, even if they find the claims questionable. The curriculum frameworks developed by the states provide the base of what an educated, knowledgeable student is supposed to know and be able to do in that state. The curriculum frameworks are supposed to be tied to workplace literacy and help students be employable after graduation. The frameworks can be quite broad and all-encompassing, covering a breath of material considerably wider than assessments like the SAT, ACT or AP exams. The high stakes tests seem better equipped for evaluating student preparedness for higher learning and for job readiness. The other assessments in question have similar aspects, but they serve a different purpose, and thus their test results have different indicators. The national assessments, in particular the SAT, examine a narrow set of skills. State developed curriculum frameworks assess a much broader set of skills, as well as subject specific knowledge. Further, scores going up on these standardized tests present no difference from scores going up on state high stakes tests. Another limitation to this study is its reliance on archival time series to examine the effects of high stakes testing.

It should also be noted that certain states provide financial inducements to schools to get students to participate in AP subjects; this practice may influence testing outcomes as pool of students taking this voluntary test increase, the average score on the test decreases. Passing rates and average scores on assessments like AP exams can be manipulated by a state if the state provides financial inducements to schools to get students to participate in AP subjects. An example of such inducements can be found in the State of Florida, which provides an extra .25 full time equivalent student funding to a school for every AP class a student takes. As statistics

on student enrollment in AP classes seem to indicate, the fiscal inducement to get students into AP exams seemed to have had an impact in Florida, since Florida lead the country with the highest percentage of students taking AP classes, yet has the lowest pass rate on the AP exams. The low pass rate may be attributable to the state policy of paying schools extra money to get students into AP classes, and nothing to do with the impact of Florida's high stakes testing program. (OPPAGA, 2006)

As a follow up to their previous study, Amrein and Berliner (2002b) investigated whether score on ACT, SAT, AP, and NEAP increased as a result of high stakes and high school graduation tests' implementation. They studied the data from 28 states, and they concluded that there was no improvement in these previously mentioned four standardized tests with the introduction of high stakes testing.

The abovementioned assessments have differential stakes for students, with each assessment engendering a different reaction from students. The assessments used by Amrein and Berliner have differential stakes for students. Students may have to pass state mandated assessments in order to graduate: if students don't pass, they don't graduate. Thus, these state mandated assessments have high stakes for students. However, NAEP scores have no stakes for students. The scores are not tied to individual students, nor are such scores part of a student's record. The ACT and SAT scores may have high stakes to students, especially if students apply to competitive institutions of higher education with low rates of admittance. However, if students apply to non-competitive institutions, the SAT or ACT scores, while counting for something in the college admission process, may not count for much and may not rise to the level of high stakes. SAT and ACT scores at the statewide level also depend upon the percentage of students in the state taking this voluntary exam. If an ever increasing percentage of students take the SAT,

the mean score of the state may remain the same, but the socio-economic components of those taking the test may change slightly. If there is a slight tendency for lower socioeconomic status (SES) students to take the SAT, and SAT scores remain stable, the state can be congratulated for those stable SAT scores because the expectation would be students with lower socioeconomic status (SES) would predict to lower SAT scores.

These two previous studies have generated controversy in the educational field, as many subsequent studies have been conducted to vet Armein and Berliner's claims. In a two year study, researchers Martin Carnoy and Susanna Loeb (2003) conducted the same study as Amrein and Berliner, but they, however, came to the conclusion that NEAP mathematics scores in states with high stakes are dramatically higher than the states without them. Further, they argue that the more rigorous the accountability measures in the all 50 states, the greater the gains in the NEAP mathematics scores. One might argue that the Amrein and Berliner study is more comprehensive in scope, as they looked at four standardized tests. Thus, it might be assumed to engender more accurate results than Carnoy and Loeb report. However, the conflicting studies must raise the question about the real results of standardized testing and the real impact of these rigorous testing methods.

The methodologies used in these studies are worth questioning. All these researchers studied similar sets of data, yet they arrived at different conclusions. Berliner stated in one of his interviews that:

Different methods yield different results. All this should do is get more researchers involved so that we get more data. It wouldn't surprise me if we find high-stakes testing has positive results in some states and negative results in others (as cited in Viadero, 2003).

Hoffman and Nottis (2008) conducted a study to investigate middle school students' perception of high stakes tests. In this study, students were asked to answer a questionnaire and draft a letter to the school principal, wherein they were allowed to openly share their thoughts about these tests. A student stated, in his terse reply, "It's a useless and worthless test and the only good purpose, I think, what it should be used for would be to start a fire, to light up other tests, in order to incinerate them and lift them from the face of the Earth in a gigantic bonfire" (Hoffman, Nottis, 2008, p.218 ). Limitations of this study include small sample and anecdotal evidence based on a survey.

Pope (2001) anecdotally narrates the schooling journey of five high schools students to surface the intrinsic ills in our schools. Her work did not deal with high stakes testing, yet it investigated testing in schools, and revealed the insurmountable pressure put on students. She sets forth that the overall reliability on testing leads to academic dishonesty acts, such as cheating and plagiarism. This claim is very weak, as it lacks any empirical evidence and it is merely anecdotal.

Kohn (2004) argues that high stakes testing engenders positive outcomes. He posits that: students and teachers need high-stakes testing to know what is important to learn and to teach; teachers need to be held accountable through high-stake tests to motivate to teach them better, particularly to push the laziest ones to work hard; students work harder and learn more when they have to take high stakes testis (as cited in Bonner III, 2007). Unfortunately, this assertion appears to be based on observations and reflections, and it lacks empirical evidence.

Here are some of the assumptions made by those who support high stakes testing. They argue that,

- Students and teachers need high-stakes tests to know what is important to learn and to teach;
- Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the laziest ones to work harder;
- Students work harder and learn more when they have to take high-stakes tests;
- Students will be motivated to do their best and score well on high-stakes tests: scoring well on the test will lead to feelings of success, while doing poorly on such tests will lead to increased effort to learn. (as cited in Amrein and Berliner, 2002a). All these assertions call for empirical evidence.

This literature review has presented studies that are in favor and against high stakes tests, yet it has been challenging to find studies replete with highly rigorous empirical evidence to conclusively assert whether these tests increase or decrease students' learning. As Goertz, a co-director of the Center for Policy Research in Education, states, "I don't think we'll ever have the definitive answer that high-stakes accountability, per se, is good or bad" ( as cited in Viadero, 2003). Goertz seems to indicate a lack of hope perpetuated by the continued emergence of studies with flawed designs lacking in empiricism, ones which generate inconclusive results open to creative interpretation from opposing viewpoints.

## References

- Amrein, A.L., & Berliner, D.C. (2002a, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). Retrieved March 7, 2009, from <http://epaa.asu.edu/epaa/v10n18/>
- Amrein, A.L., & Berliner, D.C. (2002b, December ). The impact of high-stakes tests on student academic performance, Tempe, AZ : *Arizona State University Education Policy Research Unit (EPRU)*. Retrieved March 7, 2009, from <http://www.asu.edu/educ/eps/EPRU/documents/EPST-0211-126-EPRU.pdf>
- Bonner III, C.E.( 2007). From coercive to spiritual: what style of leadership is prevalent in k–12 public schools? Retrieved March 9, 2009, from <http://idea.library.drexel.edu>
- Carnoy, M.,& Loeb, S. ( 2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation & Policy Analysis*, 24 (4), 305-331. Retrieved March 3, 2009, from <http://www-personal.umich.edu>
- Hoffman, L., & Nottis K. (2008). Middle school students' perceptions of effective motivation and preparation factors for high-stakes tests. *National Association of Secondary School Principals*, 92 (3), 209-223. Retrieved February 18, 2009, from <http://onlinersagepub.com>
- Kohn, A. (2004). Many children left behind. In Meier, D & Wood, G (Ed.), *Nclb and the effort to privatize public education (pp.79-97)*. Boston: Beacon Press.

Office of Program Policy Analysis & Government Accountability. (2006). *Acceleration programs provide benefits but the costs are relatively expensive*. Retrieved March 5, 2009, from <http://www.oppaga.state.fl.us>

Pope, D. (2001). *“Doing school” how we are creating a generation of stressed out, materialistic, and miseducated students*. New Haven: Yale University Press.

Viadero, D. (2003). Study finds higher gains in states with high-stakes tests. *Education Week*, 22, April 16: 10. Retrieved March 11, 2009, from <http://www.northwestern.edu/ipr/events>