Preparing Students for State Assessments: Cautions and Suggestions

Bruce Arnold
UCLA Curtis Center Mathematics and Teaching Conference
March 1, 2014
Playing the Game of High Stakes Testing

“My first few years of teaching, I did not fully realize how much was riding on state mandated testing. I did not understand the *game* of high stakes testing.

I only want to spend time on the material that my students will see on the test. If I have time later in the year, I may go back and cover that material, but I want to make sure I can get everything covered that they are going to be tested on first.

Testing blueprints tell you what percentage of questions is going to be asked over a particular topic. If that concept only represents one question on the test, then is it worth the four days of class it is going to take to get all of your students to be able to answer it?

Most districts invest money to help students pass the state test because of how much is riding on them.

In the month leading up to the tests, we answer hundreds of questions just like they see on the actual test.

I believe that you can increase any student’s scores just by teaching them test taking strategies.”

-ht­p://teaching.about.com/od/assess/a/Playing-The-Game-High-Stakes-Testing.htm
Overview

• High-stakes tests
• Campbell’s Law
• Campbell’s Law and Education
• NRC Report “Incentives and Test-Based Accountability in Education”
• Test preparation
High-Stakes Tests

“A high-stakes test is any test used to make important decisions about students, educators, schools, or districts, most commonly for accountability. In general, ‘high stakes’ means that test scores are used to determine punishments, accolades, advancement (for students), or compensation.”

-http://edglossary.org/high-stakes-testing/
High-Stakes vs. Low-Stakes

“High stakes are not a characteristic of the test itself, but rather of the consequences placed on the test.”
-en.wikipedia.org/wiki/High-stakes_testing

“What distinguishes a high-stakes test from a low-stakes test is not its form but its function.”
-edglossary.org/high-stakes-testing/

“Certain uses of achievement test results are termed “high stakes” if they carry serious consequences for students or for educators.”
Examples of High-Stakes Tests

- Driver's license tests
- Many job interviews
- Drug-screening tests
- High school exit examinations
- NCLB tests
- Ph.D. oral exams
- Professional licensing and certification examinations (such as bar exams, medical exams)
High-Stakes Tests Concerns

“(I)f high-stakes testing programs are implemented ... where tests lack sufficient reliability and validity for their intended purposes, there is potential for serious harm. Policy makers and the public may be misled by spurious test score increases unrelated to any fundamental educational improvement ... and curriculum and instruction may be severely distorted if high test scores per se, rather than learning, become the overriding goal of classroom instruction.”

Campbell’s Law

“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.”
Examples of Campbell’s Law

• Airline on-time statistics
• Diesel truck emissions
• TV sweeps weeks

Campbell’s Law Applied to Education

“Achievement tests may well be valuable indicators of general school achievement under conditions of normal teaching aimed at general competence. But when test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways.”

– Assessing the Impact of Planned Social Change, Campbell, 1976
“Incentives and Test-Based Accountability in Education” (NRC, 2011)

“The charge called for the committee to examine research related to the use of incentives and to synthesize its implications for the use of test-based incentives in education... The use of test-based incentives in education has been growing for several decades. However, it was in the first decade of the 21st century—which saw the enactment of NCLB, the maturation of the state movement for using high school exit exams, and the strong interest in using newly-available student test data to tie teacher pay to value-added analyses of their students’ test results—that the use of test-based incentives truly took hold of the education policy world. At the same time, there has been a transformation in the rigor of the methods used to analyze educational data. The combination of policy experimentation and new research methods has produced the set of studies that are reviewed in this report.”

Accountability Systems

Federal and state governments have devised systems (e.g., AYP, API) that hold students, teachers, principals, districts, and states accountable for how much students learn.

– Consequences attached to large-scale tests (e.g., STAR, CAHSEE) are usually a key component.

– Consequences include HS graduation, promotion, teacher pay, school organization, funding.

– Such tests become high-stakes for students, parents, teachers, principals, and/or administrators.

-Incentives and Test-Based Accountability in Education, 2011
Incentive Theory

“People are pulled toward behaviors that offer positive incentives and pushed away from behaviors associated with negative incentives.”

–Essentials of Psychology, Bernstein, 2010

“Incentives only become powerful if the individual places importance on the reward.”

High-Stakes Assessments and Incentives

“The effect of the incentive is likely to differ, encouraging greater performance for those who are able to reach the target with some extra effort but discouraging performance for those who believe they are unlikely to reach the target at all.”

-Incentives and Test-Based Accountability in Education, 2011
High-Stakes Assessments and Incentives (cont.)

“Incentives will often lead people to find ways to increase measured performance that do not also improve the desired outcomes.”

-Incentives and Test-Based Accountability in Education, 2011
Alignment of Performance Measures to Desired Outcomes

“(Performance) measures need to be chosen so that behavior that increases the measures also increases the desired outcomes.”

-Incentives and Test-Based Accountability in Education, 2011
Failure to Measure Desired Outcomes

“(T)ests fall short as a complete measure of desired educational outcomes. Most obviously, the typical tests of academic subjects that are used in test-based accountability provide direct measures of performance only in the tested subjects and grade levels. In addition, less tangible characteristics—such as curiosity, persistence, collaboration, or socialization—are not tested. Nor are subsequent achievements, such as success in work, civic, or personal life, which are examples of the long-term outcomes that education aims to improve.”

-Incentives and Test-Based Accountability in Education, 2011
Tests as Estimates from a Subset of a Domain

“(Tests) will measure only a subset of the domain being tested. Some material in the domain will be reflected in the test and other material in the domain will not.”

-Incentives and Test-Based Accountability in Education, 2011
Score Inflation

“If teachers move from covering the full range of content to focusing specifically on the portion of the curriculum on the test, it is possible for test scores to increase while learning in the untested portions stays the same or even declines.”

-Incentives and Test-Based Accountability in Education, 2011
Test Validity

“Standardized tests may be compromised by teachers coaching their students because of the high stakes consequences of the results. In either case, the instructional process is subverted when assessment results become the goal of instruction.”

-Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as Indicators for the Assessment of Educational Outcomes, Sanders & Horn, 1995
Test Validity (cont.)

• Reported gains on high-stakes tests may badly overstate real gains in student performance.
• Test-specific gains that don’t translate to other measures, e.g., NAEP, are suspicious for test inflation.
• “Sawtooth” pattern observed when implementing a new test.

Use Non-Incentivized Performance Measures for Evaluation

“It is usually necessary to find different performance measures – that are not being used in the incentive system – to use when evaluating how the incentive system is working....

To avoid having results determined by the score inflation that occurs in the high-stakes test attached to incentives, researchers should use low-stakes tests that do not mimic the high-stakes tests to evaluate how test-based incentives affect achievement.”

-Incentives and Test-Based Accountability in Education, 2011
Low-stakes Assessments

“Tests best foster learning when they have informational significance in a non-controlling way—that is, by providing individuals with specific feedback that points the way to becoming more effective or more competent, but without pressure or control.”

-Incentives and Test-Based Accountability in Education, 2011
Minimizing Score Inflation and Improving Test Validity

1. Broaden tests
2. Choose appropriate indicators
3. Use multiple measures
4. Add information about student performance

-Incentives and Test-Based Accountability in Education, 2011
Smarter Balanced (and PARCC) Trying to Broaden Tests

“Smarter Balanced assessments will go beyond multiple-choice questions to include extended response and technology enhanced items, as well as performance tasks that allow students to demonstrate critical-thinking and problem-solving skills.”


“In math, students will have to solve complex problems, show their work, and demonstrate how they solved the problem.”

- http://www.parcconline.org/sites/parcc/files/PARCCFactSheetandFAQsBackgrounder_FINAL.pdf
CDE transitioning to new system of assessments to try to cover the full breadth and depth of the curriculum

- Graduation rate
- Bonus points for disadvantaged
- Credit for Special Education Certificate of Completion
- “a-g” requirements
- Community college courses
- AP and/or IB courses and exams
- Honors courses
- Algebra II completion
- Chemistry completion
- Industry certificate
- CTE pathway and courses
- ROCP pathway
- EAP exams
- ACT/SAT exams
- Visual and performing arts
- Work experience

-Report of Survey Results on the Inclusion of New Indicators in the High School Academic Performance Index (API), CDE, 2014
Inappropriate Test Prep

• Narrowing the curriculum to teach only those topics covered in assessments.
• Devoting instructional time based on the relative amount of test coverage.
• Teaching test-taking strategies like eliminating incorrect answer choices or substitution.
• Applying known test construction rules to select responses.
• Excessive focus on the test instead of the curriculum content (e.g., drilling students on released questions from prior tests).

-Incentives and Test-Based Accountability in Education, 2011
How to Prepare Students for High-Stakes Tests

“The best way to prepare students is to provide them with a robust curriculum of instructional units specifically articulated to the CCSS. Educators should discuss curricula with staff in grade levels above and below them to ensure that students are receiving a fluent learning transition from grade to grade.”

-Smarter Balanced Questions and Answers, CDE
How to Prepare Students for High-Stakes Tests (cont.)

• The best academic preparation for state assessments is good instruction.

• It is the standards and frameworks, not the tests, that guide instruction – prepare students by focusing on the curriculum.

• Prepare students with test-taking strategies designed to improve their performance on any test (not a specific test).

• Use released test questions to teach test-taking skills rather than to drill students on a narrow set of questions.

- Guidelines on Academic Preparation for State Assessments, CDE, 2009
Appropriate Test-Taking Strategies to Teach

- Using time efficiently
- Understanding directions
- Placing answer correctly on answer sheet
- Checking answers
- Exposing students to various test formats

-Guidelines on Academic Preparation for State Assessments, CDE, 2009
Appropriate Use of Released Test Questions

• Analyzing released items to understand why standards are assessed in a particular way.
• Analyzing released test questions to distinguish between correct and incorrect answers.

- Guidelines on Academic Preparation for State Assessments, CDE, 2009
We wanted to understand the logistics of SBAC administration in comparison to the STAR assessment; what was the same and what was different?

We wanted to (1) learn about SBAC administration, (2) see how our current technology worked, and (3) observe our students working with SBAC tasks.

Students had to write about their thinking and make justifications for their answers; it wasn’t only about finding the ‘right’ answer.

The test items were interesting for the students.

The students didn’t know how to carry out computer-based tasks in some situations. Four challenging situations: (1) tasks that involved moving between two or more screens, (2) operating spreadsheets and calculators, (3) manipulating virtual objects such as geometric constructions, and (4) editing electronic text.

- [http://www.scoe.org/pub/htdocs/blog-tech.html?id=62#.UwKA8v2vsII](http://www.scoe.org/pub/htdocs/blog-tech.html?id=62#.UwKA8v2vsII) (Sonoma County Office of Education)
SBAC Practice Tests Spring Pilot (2013)

• Administrators, teachers, and students need to see and experience the SBAC firsthand before 2014-15.
• Plan time to prepare and do trials with your students. You won’t do well if you wait until the last minute.
• Use the SBAC Help Line. They are knowledgeable and helpful.
• Have a master list with SSID numbers and student names.
• Do the practice assessment.
• During the testing period, there needs to be a dedicated person who will manage the SBAC; it is not reasonable to think the classroom teacher can do it alone.

-http://www.scoe.org/pub/htdocs/blog-tech.html?id=62#.UwKA8v2vslI (Sonoma County Office of Education)
Cautions and Suggestions

• Cautions
  – Incentives lead people to focus on test prep
  – Test prep narrows the taught curriculum
  – Test prep inflates test scores that don’t generalize to mastery of intended domain

• Suggestions
  – Provide students with robust curriculum
  – Articulate curriculum above and below
  – Use low-stakes tests that add information about student performance
  – Don’t teach narrowly to the high-stakes test
  – Share appropriate test-taking strategies
Conclusion

“(Ignoring score inflation) leads to an illusion of progress and to erroneous judgments about the relative performance of schools. More important, it cheats the students who deserve better and more effective schooling.”