

Report To The Mastery Examination Committee: Educators Question Whether SBAC Is The Best Mastery Examination For Connecticut's Students

Connecticut Education Association
January 11, 2016

Introduction

What follows is an overview of surveys, documents, and research concerning the 'Smarter Balanced' (SBAC) test.

Section I explores a survey of Connecticut teachers regarding the administration of and student reaction to the SBAC test; responses are grouped according to their relevance to the statutory charges of the Mastery Examination Committee. Also included are pertinent excerpts from surveys of teachers in Oregon and Washington, which parallel and support the findings in the Connecticut survey.

Section II addresses the issue of SBAC validity and reliability from the perspective of those in Connecticut, California, Vermont, and Maine.

Section III examines the issue of SBAC fairness and equity in general, and specifically as to Connecticut and students in high poverty districts.

I. Surveys of Educator Observations Regarding SBAC

A. The Connecticut Education Association consulted with Professor Steven Stemler of Wesleyan University on a survey of CEA teachers (1,666 participants) in May and June of 2015:

Professor Stemler estimates the margin for sampling error at +/- 2.3%. The results address some of the major charges to the Mastery Examination Committee. The charges to the Mastery Examination Committee (from Public Act 15-238) include the examination of the following:

1) The impact of the statewide mastery examination on teaching, students and student learning time.

- 90% of participating teachers agree with the statement, 'SBAC preparation takes away significant time and resources from teaching and learning in my classroom.'
- 86% agree that SBAC has a negative effect on the social and emotional wellbeing of children in their classroom.
- 97% do not agree with the statement, 'overall, SBAC has proven to be beneficial toward improving student learning in my classroom.'
- 54% reported that more of their students became distraught (e.g., crying or other manifestations of stress) during SBAC compared to other standardized tests (e.g., CMT).
- 90% agreed that 'the time it took to complete the SBAC test caused student frustration and apathy.'

2) The administration of the statewide mastery examination on computers or other devices.

- 71% agree that students ‘exhibited widely disparate and inequitable computer skills when taking the SBAC test.’
- 77% reported that their students have lost significant access to computers/technology throughout this school year because the SBAC test administration and preparation has limited their access to the computer lab(s) in their school.
- 57% reported technical problems not related to lack of technical support, including login problems and system crashes.

3) Whether the statewide mastery examination is an appropriate student assessment.

- 73% of participating teachers do not agree that the computerized test administration format is developmentally appropriate for their students.
- 85% of elementary school teachers reported that the computerized testing format was developmentally inappropriate for their students.
- 97% of participating teachers do not agree that SBAC is a useful indicator of school effectiveness.
- 85% of participating teachers agreed with the statement, ‘I generally view the SBAC as an obstacle for my students to overcome.’
- 43% reported that ‘significant portions of the test covered content that is not taught at my students’ grade level.’

4) Offers accommodations for students with disabilities and students who are English language learners.

- 64% of participating teachers did not agree that the SBAC's accommodations for students with disabilities worked well.

5) Responds to students’ needs.

- In a comparison of poor and affluent school districts, teachers said that the format of the SBAC test placed high-poverty district students at a greater disadvantage, with one-third more students adversely impacted by inequitable computer skills and developmentally inappropriate tests compared to students in affluent districts.
- 40% of teachers in Alliance Districts reported that “the majority of their students clearly gave up on the test and clicked through,” whereas only 15% of teachers in affluent school districts reported the same.

6) Informs teachers of student progress.

- 3% agreed with the statement, ‘SBAC results provided me with information about my students that I did not already know.’
- 92% did not agree that the proposed timeline for SBAC to deliver results would allow them to use the information in a meaningful way to assist students.

B. The Oregon Education Association surveyed its members (1,279 participants) in June 2015 regarding the SBAC test:

The results were similar to Connecticut. The OEA concluded, “the alarming pattern of inconsistencies, technical difficulties, and other underlying problems with the administration of the assessment demands action.” Specifically, the survey found that:

- 95% of participating teachers reported that preparation for and administration of SBAC significantly disrupted the student learning process.
- 95.9% characterized SBAC as disruptive or very disruptive on other programs and facilities such as the use of computer labs, library access, and classroom time in non-tested subjects.
- 68% reported computer equipment and software failures.
- 78% reported that many students did not have the keyboarding skills necessary to allow them to take the test efficiently.
- 74% reported that SBAC did not meet the needs of students who require student-specific accommodations.
- The English Language Arts SBAC test took 55% of the students between 6.25 and more-than-10 hours to complete (25.7% finished in 4.25 to 6 hours, 15.3% in 2.25 to 4 hours, and 1.5% in 0 to 2 hours).
- The Mathematics SBAC test took 36% of the students between 6.25 and more-than-10 hours to complete (31.8% finished in 4.25 to 6 hours, 27% in 2.25 to 4 hours, and 4.1% in 0 to 2 hours).
- SBAC testing took much longer than expected for many students, with approximately one-third of students taking between 12 and 20 hours to complete both the ELA and Mathematics tests.
- “The text-heavy and unclear instructions were barriers for all students, but particularly for students with disabilities and English Learners.”

C. The Washington (State) Education Association surveyed its members (5,409 participants) in June 2015 regarding the SBAC test:

The results were also similar to Connecticut:

- On a scale of one to four, with one being “very disruptive” and four being “not at all disruptive,” participating teachers rated the overall impact of SBAC administration on their buildings and facilities as 1.3.
- On a scale of one to four, with one being “more negative” and four being “more positive,” teachers rated the overall SBAC experience compared to the previous mastery exam as 1.7.
- 89% reported students experiencing technical issues; those most frequently cited were login problems, trouble with audio settings, students losing work, and failure of in-test tools (e.g., fractions, calculator).

- “The most frequently reported ELA student challenge was unclear instructions, followed by frustrating test layout (e.g., split screen, scrolling and difficulties navigating).”
- “The most frequently reported Mathematics student challenge was using the in-test tools followed by confusing question wording.”
- “Schools with poor technology infrastructure reported longer testing windows causing school-wide disruptions for several months.”
- “Many respondents felt the SBA test was more of a technology proficiency test than a test on ELA and Mathematics.”
- “Many respondents pointed out socioeconomic bias due to the SBA’s online administration. Less affluent communities and districts have less access to technology at home, and are more likely to have older equipment and fewer computers per student.”
- “English Language Learner teacher respondents rated their student’s keyboarding proficiency and their students’ familiarity with their devices significantly lower than all respondents.”
- “Students with accommodations were often cited as rushing through the test and/or giving up quickly.”
- “The mean completion time (for the SBAC ELA test) for students in grades 3-5 was approximately 2 hours more (6.1 hours) than the SBA estimate (4 hours).”
- “The mean amount of time (for the SBAC Mathematics test) for grades 3-5 was approximately 2 hours more (4.98 hours) than the SBA estimate (3 hours).”

II. Questions Regarding the Lack of Validity and Reliability of SBAC

A. In the CEA survey, Professor Steven Stemler concluded that concerns regarding SBAC’s computer interface and developmental inappropriateness (especially for elementary school students) threaten its validity:

“The first theme to emerge centers relates to the use of computers for the administration of the testing.... There was variability among students, of course, as 71% of teachers reported that students exhibited widely disparate and inequitable computer skills when taking the SBAC test. These problems appear to be exacerbated even further for elementary school students, where one-third (32.7%) of elementary school teachers stated that a majority of their students lacked computer skills sufficient to succeed on the test (as compared to the 9% of high school teachers expressing a similar view).

“When one compares the results by District Reference Group/SES, it is clear from that data that students in the lowest performing districts are being the most disadvantaged by their relative lack of computer skills. Specifically, 37% of teachers in the Alliance School Districts reported that the majority of their students lacked sufficient computer skills to succeed on the test. By comparison, only 6% of teachers in the highest performing DRGs made the same assertion. Furthermore, 73% of teachers overall felt that the computerized testing administration was not developmentally appropriate for their students....

“The fact that the medium of test administration may be interfering with students’ ability to demonstrate their knowledge of the content of the test is a major threat to the validity of the test

results. For many years, the field of assessment has struggled with the fact that word problems in mathematics do not simply test mathematical knowledge, but rather are often strongly associated with reading comprehension as well. Consequently, such items are rarely a pure measure of the construct of interest (mathematical knowledge). It appears that we may find ourselves in a similar circumstance in the context of modern technology. *If computer literacy skills are interfering with students' ability to demonstrate what they know and can do in the domains of interest (in this case, Mathematics and English Language Arts), then the new tests are not a valid measure of their intended constructs.*"

B. Similar questions regarding SBAC validity and reliability have been raised in other states, including California. Some of those concerns were summarized by Dr. Roxana Marachi, Associate Professor in the Department of K-8 Teacher Education at San Jose State University, in a letter to the California State Board of Education for the July 2015 State Board of Education meeting:

- "How is standardization to be assumed when students are taking tests on different technological tools with vastly varying screen interfaces? Depending on the technology used (desktops, laptops, chromebooks, and/or ipads), students would need *different* skills in typing, touch screen navigation, and familiarity with the tool."
- "How are standardization and fairness to be assumed when students are responding to *different* sets of questions based on how they answer (or guess) on the adaptive sections of the assessments?"
- "How is fairness to be assumed when large proportions of students do not have access at home to the technology tools that they are being tested on in schools? Furthermore, how can fairness be assumed when some school districts do not have the same technology resources as others for test administration?"
- "How/why would assessments that had already been flagged with so many serious design flaws and user interface problems continue to be administered to millions of children without changes and improvements to the interface?"
- "How can test security be assumed when tests are being administered across a span of over three months and when login features allow for some students to view a problem, log off, go home (potentially research and develop an answer) and then come back and log in and take the same section? (This process was reported from a test proctor who observed the login, viewing and re-login process)."

C. In March 2015, the Vermont State Board of Education unanimously voted to suspend the use of SBAC scores for the 2014-2015 school year for the purpose of annual school evaluation determinations because of questions regarding SBAC validity:

Stephan Morse, chairman of the Vermont State Board of Education stated:

- "Until students' education has been guided by the new standards and schools have practiced administering and interpreting SBAC, the results will not support reliable and valid inferences about student performance and should not be used as the basis for any consequential purpose. *Unless empirical studies confirm a sound relationship between performance on the SBAC and critical and valued life outcomes ("college and career-*

ready"), test results should not be used to make consequential judgments about schools and students."

D. In 2015, Maine rejected the SBAC test because of questions about its validity:

Legislation endorsed by the education committee by a unanimous vote, directed Maine to withdraw from the Smarter Balanced consortium and instead solicit proposals for new tests created with more input from educators and the public.

- Mount Desert Elementary School Principal Scott McFarland told the school system board that, having seen some of his student's Smarter Balanced test results, he thinks the assessment is seriously flawed. "*I've seen enough of it, seen enough glitches to know that it's invalid data,*" he said. Julie Meltzer, Director of Curriculum said local school personnel spent "days and days" preparing for implementation of the Smarter Balanced assessment. Mount Desert Superintendent Howard Colter said that for the state to pull the plug on Smarter Balanced after just one year is "just stunning," given the amount of "energy and time and money and training that went into it." Meltzer said that everyone involved with the local schools had done their best to get ready for the new assessment and to try to make it work. "Our teachers stepped up; our technology people stepped up; our administrators stepped up; our families stepped up; and that was really impressive," she said. From "Test fatally flawed, school officials say," *Mount Desert Islander*, June 5, 2015.
- Maine Gov. Paul LePage signed a bill in 2015 that requires the state to leave Smarter Balanced, the testing consortium that provided the common-core aligned assessments state students took this past spring.

E. A majority of states have withdrawn from SBAC and PARCC due to validity and reliability concerns:

- "PARCC began as a cooperative between 26 states, but now only six and the District of Columbia will use the test. Smarter Balanced began with 31 states—some states joined both groups—and now counts 15." From: "Massachusetts's Rejection of Common Core Test Signals Shift in U.S.," *New York Times*, November 21, 2015.
- In 2015, Missouri Gov. Jay Nixon signed a budget that ended funding for Smarter Balanced and required the state to seek a new exam. Wisconsin has already issued a request-for-proposal to replace Smarter Balanced in 2016.

F. Connecticut schools officials voice concern about test results and SBAC:

- Fairfield School Superintendent David Title was skeptical of the 2015 SBAC results: "While his students scored typically higher than the state average there was an unusually huge gap in scores between the two high schools, with Ludlowe scoring much higher than Warde. At Ludlowe, 86.8 percent students passed language arts compared to 67.8 percent passing at Warde. In math, 64.5 percent passed at Ludlowe while 34.6 percent passed at Warde. 'To me this calls into question the validity of the scores at Warde,' Title said, suggesting that a group of Warde students did not take the test seriously, dragging down the results." From: "Lower scores on the Common Core," *Connecticut Post*, August 28, 2015.

- Connecticut Association of Public School Superintendents (CAPSS) Executive Director Joseph J. Cirsuolo:

“Right now, the results of state tests are used too often to determine whether school districts, schools, educators and children are successful. This is not only a misuse of state test results but also a perpetuation of the myth that every child learns at basically the same pace. The present expectation for state test results, for example, stipulates that every child in the fifth grade will have mastered particular knowledge and skills when we have known for decades that different people learn at different rates of speed when learning different things at different times in their lives.

“CAPSS, therefore, continues to advocate for a transformation in the use of state test results that would make them the validators of already collected evidence of success instead of the determiners of that success.” From: “Superintendents’ group seeks radical overhaul of state testing,” *Connecticut Mirror*, October 28, 2015.

- Madison School Superintendent Thomas Scarice:

“The educational community, the very field that finds its existence in the care and welfare of children, particularly where ‘the story’ matters most, continues to kneel and bow at the altar of big data, any data that can be captured, with flagrant disregard for its importance or meaning.... Sadly, too many teachers have been trapped in mindless data exercises that irresponsibly neglect the story behind the numbers, turning children into faceless numbers...hence dehumanizing the sacred process of fostering the growth and development of our children....

“What will the SBAC data mean? Nothing. Absolutely nothing at all. Numbers in isolation, lacking story and context. But, worse yet, numbers based on a specious assessment that will in time suffer and die from a credibility crisis. Sadly, some communities will be asked to explain why this one indicator supersedes volumes of contextual data that form a completely different narrative and tell another story entirely....

“Data is meaningless without context, without a story. Consequential data based on a spurious assessment is dangerous.” From “Superintendent: What will new Common Core test scores mean? Pretty much nothing,” *Washington Post*, August 6, 2015.

G. Importance of testing validity:

- W. James Popham, well-known expert in educational assessment and professor emeritus at the University of California Graduate School of Education and Information Studies:

“Pumice is a porous form of volcanic rock. To use pumice as the foundation for a major building would be an instance of engineering idiocy. And yet, in education, we seem to have something going on that’s quite analogous, and it is every bit as foolish. We have created a system of educational accountability whose conceptual cornerstone is the performance of students on annual accountability tests. If students’ scores on those tests are high, it is thought educators are doing a satisfactory job. If students’ score on those tests are low, it is thought educators are flopping. But, of course, a pivotal assumption in this enterprise is that the accountability tests being used are, in truth, capable of differentiating between appropriately taught and inappropriately taught students.

“Educational accountability’s test-based house of cards, however, should collapse in a heap if the tests being employed to evaluate educational quality are unable to do this job

properly. Accordingly, wouldn't it seem sensible given the enormous amount of money being spent on educational accountability and the enormous amount of instructional harm potentially caused by unsound decisions based on unsuitable tests, for considerable attention to have been given to the evaluative adequacy of the nation's educational accountability tests? But astonishingly, such scrutiny of our accountability tests has simply not transpired. What if these cornerstone accountability tests are merely pumice pretending to be granite?" From: *Everything School Leaders Need To Know About Assessment*, W. James Popham, Corwin Press, 2010, pg. 88.

III. Questions Regarding The Fairness and Equity of SBAC

A. Testing and the Achievement Gap, generally:

- “The achievement gap has deep roots — deep in out-of-school experiences and deep in the structures of schools. Inequality is like an unwanted guest who comes early in these children's lives and stays late. Policies and practices that are likely to narrow gaps in achievement need to be broad and comprehensive if they are to check inequality at the outset of a child's academic career and create the conditions in which every child can flower, achieve, and attain in school and in life....

“Unrealistic expectations for schools may be used to provide excuses for public policy, and thus ignore policies that might prevent learning gaps from opening. Schools are where we institutionalize learning; they are also where we tend to institutionalize blame.” From: “Parsing The Achievement Gap II,” Paul E. Barton, Richard J. Corley, ETS Policy Information Center Report, 2009.

- “Most importantly, test-based accountability is failing on its most important mandate—eliminating the achievement gap between different groups of students. While racial gaps have narrowed slightly since 2001, they remain stubbornly large. The gaps in math and reading for African American and Latino students shrank far more dramatically before No Child Left Behind—when policies focused on equalizing funding and school integration, rather than on test scores. In the 1970s and '80s, the achievement gap between black and white 13-year-olds was cut roughly in half nationwide.” From: “Sorry, I'm Not Taking This test,” *Mother Jones Magazine*, September/October 2015.
- “What NCLB has demonstrated, 15 years in, is something policymakers already knew—that standardized test scores are strongly correlated with a student's family income.” From: “The Lessons of No Child Left Behind,” *National Journal*, April 24, 2015.
- According to the Center for American Progress, urban high school students, who tend to be minorities, spend 266 percent more time on district-mandated exams than suburban students. “Black and Latino families want world class public schools for our children, just as white and affluent families do. We want quality and stability. We want a varied and rich curriculum in our schools. We don't want them closed or privatized.

“We want to spend our days learning, creating and debating, not preparing for test after test.... High stakes standardized tests have been proven to harm Black and Brown children, adults, schools and communities. Curriculum is narrowed. Their results purport to show that our children are failures.” From: The Journey for Justice Alliance, a coalition of 40 organizations that represent parents and students of color in 23 states;

letter to Congressional leadership in May 2015, urging Congress to move away from a reliance on tests.

- “Rather than contribute to achieving the twin goals of educational excellence and equity, our study suggests NCLB’s focus on standardized testing and accountability harms the learning and teaching process. It narrows curricula and distorts teaching especially for low-performing students. Perhaps the most serious consequence of NCLB’s theory of reform is that it ignores and obscures other school-based sources of inequality that include race and SES-linked access to teacher quality, continued segregation of schools along lines of race and class, ability grouping and curricular tracking, and resource inequality among the schools.” From: Roslyn Mickelson, Jason Giersch, Elizabeth Stearns, and Stephanie Moller, “How (and Why) NCLB Failed to Close the Achievement Gap: Evidence from North Carolina, 1998-2004,” *ECI Interdisciplinary Journal for Legal and Social Policy*: Vol. 3: Iss. 1, Article 1, (2013).

B. SBAC fairness and equity issues in Connecticut:

1) In the CEA Teacher Survey, responses were distinguishable between District Reference Groups (DRGs) to explore the degree to which teacher perceptions of SBAC differ across school communities. Two of the DRGs that were compared were the highest (A- Darien, Easton, New Caanan, Redding, Ridgefield, Weston, Westport, etc.), and the lowest two DRGs combined (H/I- Bridgeport, Danbury, Hamden, Hartford, New Haven, etc.). Compared to DRG A, teachers in DRGs H/I (Alliance Districts) were:

- Significantly more likely to report that a majority of their students did not exhibit computer skills sufficient to succeed on the SBAC test (37% in DRG H/I said the majority of their students lacked the computer skills needed to succeed, compared to only 6% in DRG A).
- Significantly more likely to report the computerized test administration is not developmentally appropriate (only 18% of teachers in DRG H/I said SBAC was developmentally appropriate, compared to 43% in DRG A).
- Significantly more likely to report that their students became distraught during SBAC as compared to other standardized tests (61% in DRG H/I compared to 36% in DRG A).
- Significantly more likely to report that their students ‘clearly gave up’ on the test by clicking through several test questions without taking the time to read the questions (40% in DRG H/I compared to 15% in DRG A).
- Significantly more likely to report that SBAC had a negative effect on the social and emotional well-being of the children in their classroom (91% in DRG H/I compared to 79% in DRG A).
- Significantly more likely to report that SBAC is not an appropriate measure of the Common Core State Standards (only 10% of teachers in DRG H/I said SBAC was an appropriate measure compared to 24% in DRG A).
- Significantly more likely to report viewing SBAC as an obstacle for their students to overcome (87% of teachers in DRG H/I reported SBAC as an obstacle compared to 79% in DRG A).

2) In Connecticut, SBAC scores are strongly correlated to family income and whether students need special accommodations (special needs, ELL). In the middle section of the table below (“Statewide results by free or reduced price meal eligibility status”), the three income/meal status categories (free, reduced, and not eligible) show dramatic progression in achievement as family income increases.

| Subgroup | English Language Arts Percent at Level 3 & 4: Meets or Exceeds the Achievement Level | Mathematics Percent at Level 3 & 4: Meets or Exceeds the Achievement Level |
|----------------------------|---|---|
| High Needs* | 30.6% | 16.4% |
| Students with Disabilities | 14.6% | 8.2% |
| English Learners | 10.1% | 7.0% |

*A high need student is someone who is eligible for free/reduced price meals or is an English learner or a student with a disability

| Statewide results by free or reduced priced meal eligibility status | English Language Arts Percent at Level 3 & 4: Meets or Exceeds the Achievement Level | Mathematics Percent at Level 3 & 4: Meets or Exceeds the Achievement Level |
|--|---|---|
| Free | 30.0% | 15.4% |
| Reduced | 45.6% | 26.4% |
| Not Eligible | 69.2% | 52.3% |
| All Students | 55.4% | 39.1% |

| Statewide Results by Race/Ethnicity | English Language Arts Percent at Level 3 & 4: Meets or Exceeds the Achievement Level | Mathematics Percent at Level 3 & 4: Meets or Exceeds the Achievement Level |
|--|---|---|
| American Indian or Alaska Native | 44.4% | 24.9% |
| Asian | 75.7% | 66.0% |
| Black or African American | 30.3% | 13.9% |
| Hispanic/Latino of any race | 32.8% | 17.3% |
| Native Hawaiian or Other Pacific Islander | 55.8% | 37.5% |
| Two or More Races | 57.4% | 39.8% |
| White | 67.3% | 50.3% |
| All Students | 55.4% | 39.1% |

Conclusion

Serious questions as to the validity, reliability and fairness of the SBAC test have been raised in Connecticut and other states. Even before the recent changes in federal law that ended many of the misguided mandates of the No Child Left Behind legislation, states had moved decisively away from the federally sponsored SBAC and PARCC tests. That trend is likely to continue or pick up speed with the greater flexibility that is now allowed at the state level.

SBAC has been shown to be problematic for all students, but especially younger students, those from low-income families, those who need special accommodations, and students without computer access at home. The poorly-designed, computer-based format of SBAC creates a technology gap that hurts students in high poverty districts and fails to provide a clear picture of the achievement gap.

There is a better way to proceed. Connecticut can design a test for the benefit of all students, not testing consortiums. It can avoid relying on a test that punishes students on the basis of income and lack of access to digital technology. It can respect the need for resources in high poverty schools and high poverty communities. Connecticut can avoid the impossible scenario of using one high-stakes test for multiple and conflicting purposes that result in distorted and harmful outcomes for students.