Issues and Recommendations for a Resolution of the General Assembly

Regarding Validity and Reliability of the

Smarter Balanced Assessments

Scheduled for Missouri

in Spring 2015

Mary R. Byrne, Ed.D.

January 21, 2015

(updated: 3/2/15, 3/14/15)

E-mail: mary.byrne53@att.net

**EXECUTIVE SUMMARY**

**Introduction**

Tests that are valid and reliable are the only legally defensible tests that can be incorporated into any evaluation plan of students, teachers, and districts. Missouri’s State Board of Education and Department of Elementary and Secondary Education (DESE) are responsible for ensuring that statewide assessments administered in Missouri are valid and reliable, yet, they committed Missouri as a governing member of the Smarter Balanced Assessment Consortium (SBAC) in April 2010,[[1]](#footnote-1) before the test was developed, and even before the project manager for developing the pioneering assessment system was named..[[2]](#footnote-2) In October 2013, DESE contracted with McGraw-Hill to administer tests aligned to the Common Core State Standards[[3]](#footnote-3) despite restrictions “. . . that no funds shall be used to implement the Common Core Standards” as per HB 002 Section 2.050. McGraw-Hill has also contracted with SBAC to produce the SBAC test items.[[4]](#footnote-4) The SBAC summative assessments dates are Mar. 30-May 22, 2015. According to DESE staff, student assessments will be scored, but not used in teacher evaluation or district accreditation. An SBAC memo dated September 12, 2014[[5]](#footnote-5) indicates evidence of adequate validity and reliability is not available to support administration of the SBAC in spring 2015 and interpretation of scores.

**Statement of the Problem**

No independent empirical evidence of technical adequacy[[6]](#footnote-6) is available to establish external validity and reliability of the SBAC computer-adaptive assessment system. Test scores derived from tests that have no demonstrated technical adequacy are useless data points, and plans to collect evidence of validity and reliability provide no assurance that technical adequacy will be established. Therefore, the legislature has no assurance children and school personnel will not be subject to harm if the SBAC assessments continue to be implemented.

HB 1490 Section 160.526.2 states, “. . . the general assembly may, *within the next sixty legislative days,* [emphasis added] veto such implementation, modification, or revision [of assessments] by concurrent resolution adopted by majority vote of both the senate and the House of Representatives.” Concerns about the adoption, development, and implementation of the Smarter Balanced Assessments and contracts undertaken by DESE to administer them, should provoke legislative leadership to consider introducing and passing a resolution to veto their implementation in Missouri.

**Concerns**

A review of formal and informal resources available to the public and cited throughout this paper suggests:

1. Missouri agents of SBAC agreement did not exercise due diligence before committing Missouri to the development and administration of the SBAC;

2. Report of validity and reliability to legislative leadership does not inform that evidence of SBAC external validity and reliabilityis lacking[[7]](#footnote-7);

3. Publicly available formal and informal SBAC publications do not include evidence of reliability and validity [[8]](#footnote-8),[[9]](#footnote-9)**,**[[10]](#footnote-10),[[11]](#footnote-11);

4. External reviews of SBAC: “work in progress”/ “bizarre”/vulnerable to hacking [[12]](#footnote-12),[[13]](#footnote-13),[[14]](#footnote-14),[[15]](#footnote-15);

5. Performance Test Items are in SBAC Assessments, but restricted in HB 1490;

6. DESE flouted spending restrictions of HB 002 to administer SBAC in 2015[[16]](#footnote-16),[[17]](#footnote-17),[[18]](#footnote-18);

7. Judge Daniel Green, Circuit Court of Cole County, MO ruled SBAC is an illegal state compact on February 24, 2015;[[19]](#footnote-19)

8. U.S. Department of Education (DoE) Funding of SBAC violates the intent of federal laws[[20]](#footnote-20); and

9. Purpose of SBAC inconsistent with purpose of public education in Missouri.

Absence of technical adequacy voids claims linked to SBAC scores. The submission of *a plan* to establish external validity and reliability does not guarantee validity and reliability is assured in the future. Students, teachers and school districts are vulnerable to harm associated with an indefensible test, derived from controversial standards, promoted with unsubstantiated claims, and administered by nascent computer-adaptive technology; and therefore, the state is vulnerable to litigation stemming from its implementation and questionable uses. [[21]](#footnote-21), [[22]](#footnote-22), [[23]](#footnote-23),[[24]](#footnote-24)

**Recommendations**

The following recommendations for a resolution to reverse and investigate decisions made by the commissioner of education and state board of education are made on the grounds that, they did not exercise due diligence in the adoption of membership in SBAC prior to reviewing evidence of assessment validity and reliability and on their insistence on continuing its administration, without assurances of its technical adequacy:

(1) Legislative leadership should, within the number of legislative days allowed by HB 1490, support a resolution in both the House of Representatives and the Senate to stop implementation of SBAC assessments in Missouri and withdraw Missouri from SBAC as per Missouri’s original memorandum of understanding for the purpose of developing assessments fully under the control of Missouri;

(2) The resolution should request an independent review of all SBAC test items for technical adequacy and all test items of the ACT that will be used to determine “college readiness” in grade 11 by a testing expert having experience in development and evaluation of nationally used, standardized tests and is not associated with the SBAC or PARCC consortia ACT, state or federal departments of education, or testing companies that contract with any of the aforementioned entities, for the purpose of detecting fraud;

(3) The resolution should request a thorough investigation of appropriate uses of statewide standardized tests in Missouri (for example, whether student growth models relying on standardized test results, as required by the U.S. DoE’s No Child Left Behind Waiver Renewal application) for the purpose of challenging the conditions of the ESEA waiver offered by the U.S. DoE;

(4) The resolution should request an immediate audit of DESE and State Board of Education for the purpose of determining the legality of continued administration of SBAC in Missouri schools for purposes stated in Missouri’s original ESEA waiver application and prospective renewal application.

**Introduction**

Tests that are valid and reliable are the only legally defensible tests that can be incorporated into any evaluation plan of students, teachers, and districts. Responsible test developers and publishers must be able to demonstrate that it is possible to use the sample of behaviors measured by a test to make valid inferences about an examinee's ability to perform tasks that represent the larger domain of interest. Responsible consumers require access and review of evidence of validity and reliability before purchasing and administering a test. Missouri’s HB 1490 vests this responsibility in the State Board of Education.

HB 1490 also obliges commissioner of education is to report to submit a reliability and validity report of tests used in the Missouri Assessment Plan (MAP)[[25]](#footnote-25). For a discussion about the various uses of statewide standardized tests in Missouri, see Appendix A. Legislators receive this information because they are responsible for passing legislation to protect the public from harm, especially harm caused by activity of the state. Thus, legislative leadership must be assured that legal, ethical, and professional standards are met to implement DESE’s assessment plan. This is the case especially if the state board of education and local school boards must ensure schools comply with federal and state laws and regulation, and make high stakes decisions or claims about student academic performance (such as college or career readiness) based on results of those tests.

In October 2013, Missouri’s DESE contracted with McGraw-Hill to administer tests aligned to the Common Core State Standards in the 2014-2015 school years. McGraw-Hill is also contracted with the Smarter Balanced Assessment Consortium (SBAC) to produce the SBAC test items. A September 12, 2014 memo from SBAC to K-12 Leads[[26]](#footnote-26) indicates evidence of validity and reliability is not available to support claims of the SBAC in 2014-1015. The memo states,

Test reliability will initially be modeled through simulations using the item pool after item review, which is due to be completed December 31, 2014. Operational test reliability will be reported in the technical manual following the first operational administration in spring 2015. . . .

. . . Because this type of evidence continues to be gathered through the operational administration of the assessments, this table mostly reflects future plans for external validity research. (p. 2)

In other words, DESE is contracted to pay McGraw-Hill for the administration of tests in spring 2015 with test items developed by McGraw-Hill, having no evidence of validity and reliability to defend interpretations of student test scores produced from its administration. In short, Missouri is paying for a product that was never vetted before millions of tax dollars were invested, and is scheduled to be administered in Missouri schools in spring 2015 without proper evidence of technical adequacy.

**Background: Federal and State Laws Requiring Use of Valid and Reliable Assessments**

Evidence of adequate validity and reliability is important, first, to prevent harm to children whose education is determined by test results; and, second, to comply with federal and state laws and policies requiring assurance of validity and reliability. The following is a description of such laws and policies.

Individuals with Disabilities Education Act (IDEA). IDEA mandates the use of valid and reliable tests for the identification of students with special learning needs and justification for providing special education services. Under IDEA, evaluation procedures for students with disabilities must meet specific legal requirements to insure that students are being given an equitable and individualized learning program. Specifically,

(iii) [Assessments] are used for the purposes for which the assessments or measures are valid and reliable;

Testing materials must be accessible to optimize student performance and understanding of what's being evaluated. If a student has a visual impairment, the tests must be in larger print or Braille if the student is blind. On the other hand if the student is [an English Language Learner] ELL and has limited English skills, then the tests must be written in their language of origin to provide a valid assessment of skills and deficiencies.[[27]](#footnote-27)

No Child Left Behind Act (NCLB). NCLB requires that state assessment systems, including alternate assessments, “*be valid for the purposes* for which the assessment system is used; be consistent with relevant*, nationally recognized professional and technical standards, and be supported by evidence…of adequate technical quality* [emphasis added] for each purpose”[[28]](#footnote-28)

RsMO, HB 1490 Section 160.526.2. HB 1490 (having language identical to the SB 380 (1993)) states,

The state board of education shall [,] by contract enlist the assistance of such national experts . . . to receive reports, advice and counsel *on a regular basis pertaining to the validity and reliability of the statewide assessment system. The reports from such experts shall be received by the . . . state board of education*.

. . . Within six months prior to implementation of or modification or revision to the statewide assessment system, the commissioner of education shall inform the president pro tempore of the senate and the speaker of the house of representatives about the *procedures to implement, modify, or revise the statewide assessment system, including a report related to the reliability and validity of the assessment instruments*, and the *general assembly may, within the next sixty legislative days, veto such implementation*, [emphasis added] modification, or revision by concurrent resolution adopted by majority vote of both the senate and the house of representatives.

The U.S. Department of Education’s non-regulatory guidance for academic standards and assessments (2003)[[29]](#footnote-29) defines validity and reliability as follows:

Validity

Validity is the *extent to which an assessment measures what it is supposed to measure and the extent to which inferences and actions made on the basis of test scores are accurate and appropriate* [emphasis added]. For example, if a student performs well on a reading test, how confident are we that the student is a good reader? A valid standards-based assessment is aligned with the intended learning and knowledge to be measured and *provides an accurate and reliable* [emphasis added] measurement of student achievement relative to the standard.

Reliability

Reliability is the degree to which the results of an assessment are dependable and consistently measure student knowledge and skills. *Reliability is an indication of the consistency of scores across different tasks or items* [emphasis added] that measure the same thing, across raters, and over time. . . .

The April 9, 2010 Federal Register Race to the Top Fund Assessment Program; Notice

Inviting Applications for New Awards for Fiscal Year (FY) 2010[[30]](#footnote-30) requires assessment consortia awarded federal grant money to publish results of test evaluations for validity and reliability in formal and informal mechanisms.

**Statement of the Problem**

RsMO HB 1490 Section 160.526.2 states, “. . . the general assembly may, *within the next sixty legislative days,* [emphasis added] veto such implementation, modification, or revision [of assessments] by concurrent resolution adopted by majority vote of both the senate and the house of representatives.” No independent empirical evidence of technical adequacy are available to establish validity and reliability of the SBAC computer-adaptive assessments aligned to the Common Core State Standards adopted by the State Board of Education in 2010; and therefore, the legislature has no assurance it can protect from harm those children and school personnel affected by use of the SBAC tests and student scores generated from them. Concerns about the adoption, development, and implementation of the SBAC assessments and legality of the agreements by DESE to administer them, should provoke legislative leadership to introduce and pass a resolution to veto their implementation in Missouri.

**Concerns**

**Missouri Agents of SBAC Agreement Did Not Exercise Due Diligence**

Governor Nixon, Commissioner of Education Nicastro, and members of the Missouri State Board of Education committed Missouri to participation in the Common Core State Standards Initiative and to SBAC as a governing member prior to the development of the test before any evidence of technical adequacy was available. That is, agreements were made on behalf of the people of Missouri despite the following:

* No validity and reliability report was available to support the purchase and administration of SBAC assessments as a measure of college and career readiness when the standards and assessment adoption process began;
* No validity and reliability report was available when Missouri was committed as a governing member of SBAC;
* No data are available to support the claim that implementation of the common core standards and administration of tests aligned to them will result in every student being college and career ready.

Further, no claim about college and career readiness can be validated until an entire cohort of students are educated from K though 12 in a school system implementing the Common Core State Standards and assessments aligned to them. Speaking before the NY Senate Education Committee, Michal Cohen, president of ACHIEVE, Inc., the organization that coordinated the development of the standards and is project manager of the PARCC consortium, stated that the full effects of the Common Core Initiative won’t be seen until an entire cohort of students, from kindergarten through high school graduation, has been effectively exposed to Common Core teaching. According to Sarah Reckhow, an expert in philanthropy and education policy at Michigan State University,

Usually, there’s a pilot test — something is tried on a small scale, outside researchers see if it works, and then it’s promoted on a broader scale . . . That didn’t happen with the Common Core. Instead, they aligned the research with the advocacy. . . . At the end of the day, it’s going to be the states and local districts that pay for this.[[31]](#footnote-31)

The invitation to the Race to the Top Fund Assessment Program requires assessment consortia awarded federal grants to publish reliability and validity data of consortia-developed assessments.[[32]](#footnote-32) The grant announcement was published almost a year after Governor Nixon’s agreement with the National Governors Association to develop and adopt a common set of standards and tests aligned to the standards[[33]](#footnote-33) therefore, he did not exercise due diligence in committing Missouri to a consortium developed test..

Persons involved in making the commitment could not assure Missourians that the SBAC computer-adaptive test met industry standards for technical adequacy, and that its use was legally defensible for the purposes of student testing and teacher evaluation as stipulated in Missouri’s ESEA (No Child Left Behind) Waiver.[[34]](#footnote-34) Their inability to review the technical adequacy of assessments prior to the commitment to adopt them and subsequent requirement of school districts to administer them without this data, leads to the conclusion that they did not, and are not exercising due diligence on behalf of the people of Missouri.

**Report of Validity and Reliability to Legislative Leadership Does Not Inform That Evidence of SBAC External Validity and Reliability Is Lacking**

In July 2014, Governor Nixon signed HB 1490. Section 160.526.2 states,

The state board of education shall [,] by contract enlist the assistance of such national experts . . . to receive reports, advice and counsel *on a regular basis pertaining to the validity and reliability of the statewide assessment system. The reports from such experts shall be received by the . . . state board of education*. [emphasis added]

. . . Within six months prior to implementation of or modification or revision to the statewide assessment system, the commissioner of education shall inform the president pro tempore of the senate and the speaker of the house of representatives about the *procedures to implement, modify, or revise the statewide assessment system, including a report related to the reliability and validity of the assessment instruments*, and the *general assembly may, within the next sixty legislative days, veto such implementation, modification, or revision by concurrent resolution adopted by majority vote of both the senate and the house of representatives*. [emphasis added]

As per language in the statute, which was a reiteration of statutory language in SB 380 passed in 1993, the state board should have been receiving information about the validity and reliability of the developing SBAC assessments since its piloting in 2013.

During an invitational webinar hosted by McGraw-Hill about the next generation of state assessment programs (October 28, 2010) , [[35]](#footnote-35) McGraw-Hill’s Rich Patz and Distinguished Professor Emeritus of Education Research at the University of Colorado Bob Linn[[36]](#footnote-36) discussed the challenges anticipated in establishing the validity and reliability of the new tests, especially with respect to performance test items, artificial intelligence scoring and computer adaptive testing. They affirmed, however, that old and enduring expectations for psychometrics of validity evidence and reliability standards are still operative. A September 12, 2014 memo from SBAC to K-12 Leads indicates evidence of validity and reliability is not available to support administration of the SBAC in 2015.[[37]](#footnote-37) On page two, the memo states,

Test *reliability will initially be modeled through simulations using the item pool after item review*, which is due to be completed December 31, 2014. Operational test reliability will be reported in the technical manual *following the first operational administration in spring 2015*. . . .[emphasis added]

. . . Because this type of evidence continues to be gathered through the operational administration of the assessments, *this table mostly reflects future plans* [emphasis added] for external validity research.

To date, test reliability and external validity information is not available despite confirmation by Commissioner Nicastro and DESE to members of the legislature that a purpose of the SBAC pilot test and 2014field test was to ensure the validity and reliability of test items. In her April 3, 2014 letter to the members of the General Assembly, [[38]](#footnote-38) Commissioner Nicastro wrote, “Information from the Field Test will be used to *ensure that test questions are valid and reliable for all students* [emphasis added] and are without bias. . . . In 2014-2015, Missouri’s test vendor CTB/McGraw-Hill will administer the updated assessments.” DESE’s May12, 2014 Legislative Q&A [[39]](#footnote-39) includes, “**What was the purpose of these pilot tests?** The purpose of the pilot was to test the assessment system and the different item types to *ensure validity and reliability*.” [emphasis added] (p. 7)

The discrepancy between promise and delivery in the spring 2014 communication to Missouri’s general assembly and the September 2014 SBAC memo is concerning. A full validity study should have been completed by now, before administration of the SBAC; yet, the memo leaves results of a full validity study unavailable. Actual reliability may be “refined” and re-computed annually after each administration; but the *expected reliability should have been available based on the field testing results*. No explanation is given in the September SBAC memo for a need for simulations to report reliability. (For a description of the process for developing valid and reliable test items for large scale tests, see Appendix B).

Additional concern raised by the SBAC memo is the statement, "*Development of a comprehensive set of content specifications that cast the Common Core content standards in terms of evidence statements about what students will be expected to know and do*," (Test Development Phase, bullet 2, p. 1) . Similar content appears in SBAC’s General Specifications (2012),

CCSS were not specifically developed for assessment and contain a great deal of rationale and information about instruction. Therefore, following a practice many states have used in the past, Smarter Balanced distilled from the CCSS a set of content specifications expressly created to guide assessment development. (p. 9)[[40]](#footnote-40)

The statement indicates that problem s of establishing validity and reliability may be attributable to the Common Core State Standards themselves, because they were actually insufficient for guiding the development of test items. Development of a comprehensive set of content specifications would not be necessary if the language of the standards was as clear and concise as proponents claim. Rather, the statement suggests that writing test items for SBAC requires rewriting and expanding the standards themselves. *This also means that the real standards students will be measured against are not from CCSS, but rather result from closed discussions from the SBAC development team*. Questions that must be addressed are:

* What external group is validating that effort?
* Does any state education agency have a real say in test design or content?
* How are teachers supposed to access the real standards the tests are designed to assess?

The September 12, 2014 SBAC memo to K-12 Leads is dated 18 days before Commissioner Nicastro’s September 30, 201 4, letter to Missouri’s legislative leadership as per HB 1490.[[41]](#footnote-41) However, Nicastro’s letter did not clearly state the problems encountered with gathering SBAC’s validity and reliability data. Rather, pages 1-2 of the letter made no mention that the SBAC assessments would be administered in spring 2015 as a component of the Missouri Assessment Plan despite its lack of external validity and any reliability data. Pages 3-5 of the letter discusses the importance of establishing validity of an assessment as endorsed by several assessment organizations, and the outline of a *plan* to establish validity; however, it does not provide *evidence of* the validity of the SBAC assessments.

The following is a compilation of comments from four reviewers of the letter and appendices, all with doctorates related to education and professional background in education assessment:

* First, the nature of the sample [districts and students included in the pilot studies] is problematic. In the Missouri Assessment Program Update 2014-2015 file, it states that school districts had the opportunity to participate in the pilot tests. It is possible that those that volunteered to participate were different in some regard than those districts that did not volunteer. If so, then the representativeness of the sample is suspect and the external validity (generalizability) of the results is suspect as well. [The admission that pilot test data were generated from convenience samples rather than representative samples of students is confirmed in the SBAC memo found in Appendix J of Commissioner Nicastro’s letter.]
* Appendix D (sample assessment items) includes a large fraction of bad items. This bodes ill for whatever assessment it represents, as these are only a handful of items that were supposed to be carefully reviewed. If those are full of problems, imagine a regular test.
* Appendix E is a quite good plan of what needs to be validated and how. The rest of the appendices, though, do not offer clear evidence that the plan was fully or correctly executed. Possible difference between paper-and-pencil and computer was mentioned as an action item in E, but I saw no report on the analysis of the two [types of] administrations. [That is, there is no discussion whether the two versions of the SBAC assessments will yield comparable scores that would allow for local school, local district, or statewide aggregation; the technical considerations for generating comparable scores particularly for variable form computer-adaptive tests and fixed form paper-and-pencil tests are substantial, and indeed the state-of-the-art for providing such comparisons is in its infancy. The “phase-in over 3 years” recommendation is shaky, much more so for computer-adaptive tests than for fixed form computer-administered tests and counterpart fixed-form paper-and-pencil tests.[[42]](#footnote-42)]
* Appendix H "small scale trials" indicate the reliability and validity of various actions there seem barely marginal.
* Appendix I pilot test analysis seems generally useless -- the items were not real items, so whatever the findings are they are likely irrelevant.
* Appendix J -2013 Pilot Test Report file (page 12) refers to this problem of representativeness of the sample as well as to several other problems with using the pilot test data to draw conclusions. The use of a convenience sample is all they have in this case. [Convenience sampling is nothing more than preliminary pilot data. The use of data gathered from this type of sampling for high stakes decision-making (such as student placement or teacher evaluations) is not defensible].
* Appendix K field test report is a meaningless summary. For a serious consideration a full technical of the field test must be attached.
* Appendix L summative alignment study is useless. For a serious consideration one needs to see the full study, not some comments about how the study is planned to be done during some workgroup meeting.
* Appendix M test validation worksheet is helpful in the sense that it shows what SBAC is planning on having ... maybe. But until all those reports are in place*, we have no idea whether they were completed, or what were the results of the analyses.*
* Appendix P grade level blueprints is actually interesting, in that it *clearly shows that the mathematics assessment cannot be considered reliable or having content validity*. Having at most 15 items (15 points, some items may be worth more than 1 point) to address the broad scope of the priority clusters at various grades seems clearly insufficient. There is no reasonable way that 15 items can adequately sample the content of grade-level mathematics with meaningful content validity or sufficient reliability. Further, the mathematics assessment in total has too few items, even including items addressing supporting clusters and process skills.
* Appendix R where it is clearly visible that *most standards will get a single test item to assess them, if any at all. This is grossly inadequate*. [emphasis added]
* Consequently, the memo to Dempsey and Jones is an empty document not providing the legislature with sufficient information to draw any conclusion.
* The Missouri Assessment Program update is an administrative blueprint of state expected actions that does not address the validity and quality of the assessment at all.
* The other concerns regarding the methods used to establish reliability and validity: In the Missouri Assessment Program Update 2014-2015 file (page 17), both split halves or parallel forms and inter-rater reliability methods are mentioned. However, actual reliability coefficients (that is, tables of decimals, interpreted to describe how well multiple sets of observed scores correlate, and are necessary to evaluate the quality of an assessment tool), are not listed anywhere.
* In addition, no validity coefficients are listed anywhere. For example, convergent and discriminant validity coefficients could be obtained using Campbell & Fiske's Multi-trait method (this method is referenced in the file). Also, some criterion-related validity results should be reported. No evidence is reported that test results are predictive of anything, including college or workforce readiness. Minimally, test results should relate to current classroom performance or grade point average (GPA), but this important validity evidence is not found in this report.
* I suspect that SBAC is working hard on getting [validity and reliability] in place but (a) they are probably not ready yet and MO legislature need them now, (b) some results may not look so good and SBAC will try to hide them from outsiders (like your legislature), and (c) some reports were probably dropped due to oversight or failure to perform the work.
* In short, no information in the letter or appendices allows the legislature to make an informed decision regarding the administration of SBAC for protecting the public and ensuring Missouri’s compliance with federal and state laws. Establishing validity is critical and the legislature should use the lack of timely evidence as a reason to act.

A Missouri teacher who is a reading specialist conducted informal analyses on several reading passages posted as SBAC practice items using online measurement tools.[[43]](#footnote-43) She found that “. . . **the 6th grade ELA practice test is measuring our 6th graders at a level used for 8th grade measurement on a nationally normed and validated assessment [the National Assessment for Education Progress]. Use of inappropriate levels of difficulty in reading passages is not consistent with the definition of criterion-referenced assessment. It is an indicator of an assessment constructed with test items that are not valid for the intended purpose of measuring what a student instructed at a specific grade level should know and be able to do.**

In addition to indications of problems with test item validity, comparison of test scores across several years is rendered invalid if the test is not administered according to standard procedure from year to year. February 5, 2015 memo from DESE’s Assistant Commissioner Sharon Helwig states,

In order to ensure a smooth administration of the new assessment system, Missouri will administer computerized “fixed form” assessments, much like the EOC model, for the spring 2015 summative assessments for grades 3-8. . . . Test items will be Smarter Balanced items and will be designed according to the Smarter Balanced test blueprint. . . . Missouri remains a part of the Smarter Balanced Assessment Consortium. The change will have no impact on future implementation of a computer adaptive test. Fixed forms will be used for one year, and Missouri will complete implementation of computer adaptive testing in the 2015-16 school year.[[44]](#footnote-44)

In essence, administration of a fixed form, renders any data gathered from the 2015 administration of the SBAC items useless because the procedure “fixed form” is not consistent with the pilot and field test standardization and cut score procedures described in previous SBAC reports published since the inception of the SBAC development; it will have no bearing on the computer adaptive test planned for administration in the 2015-2016 school year; and the above problems in the SBAC item design will not be remediated by the fixed form delivery.

**Formal and Informal SBAC Publications Do Not Include Evidence of Reliability and Validity**

The April 9, 2010 Federal Register “Race to the Top Fund Assessment Program; Notice Inviting Applications for New Awards for Fiscal Year (FY) 2010”[[45]](#footnote-45) requires assessment consortia awarded federal grant money to publish results of test evaluations for validity and reliability in formal and informal mechanisms. The latest SBAC annual report to the U.S. DoE[[46]](#footnote-46) (July 2014) states,

The Department acknowledges the difficult work building a next-generation assessment system to measure whether students have the knowledge and skills necessary to succeed in college and the workforce. *Smarter Balanced continued to experience challenges in Year 3 around adherence to established timelines and making sure the items and tasks developed met the consortium’s quality criteria*. *The consortium should continue to evaluate whether its quality control processes are sufficient and provide close oversight over the development of future items and tasks to ensure that established timelines and quality criteria are being met*. (p. 27)

The content of the above excerpt concurs with the observation that the documents submitted to the legislative leadership do not include a clear and cogent report confirming SBAC’s validity and reliability. In addition, the October 2014 SBAC publication, *Smarter Balanced Tests of the Test: Successful, Field Test Provides Clear Path Forward* [[47]](#footnote-47) is very revealing about the lack of evidence regarding test item validity ty and reliability. Note the following excerpts with emphasis added:

As this report will describe, however, *states’ readiness as of spring 2014 varied significantly across and, presumably, within states*, as did the readiness of adults to administer them. . . . (pp. 1-2)

The Field Test served a variety of purposes. *Its primary purpose was to “test” the computer-delivered questions and embedded tools to ensure that they function properly,* *are clear, and meets criteria for inclusion in the spring 2015 state summative assessments* required under the No Child Left behind Act. (p. 2)

*The Field Test was also an important trial run of the entire assessment delivery system*, the readiness of state, district, and school personnel to administer the tests, and the readiness of students to take them. (p. 2)

*This report is based on a review of the survey results from 13 of the 22 members.* (p. 2)

*This report is not based on a scientific sample, nor is it a scientific analysis of consistently worded survey questions across states; rather, this review reflects the author’s findings regarding major themes across the results from these 13 states*. (p.2)

In sum, the SBAC field test report provides no evidence of a well-designed methodology to control for interfering variables for the development of validity and reliability evidence supporting the quality of the test items. Rather, the report summarizes the author’s interpretation of responses to a non-scientific survey of participant perceptions of the online delivery experience. Assessments cannot be "equitable" without independent empirical evidence of technical adequacy. Again, this report adds support to the contention that the information submitted to legislative leadership is not useful for confirming technical adequacy of the SBAC.

**External Reviews of SBAC: “A Work in Progress”/ “Bizarre”/Vulnerable to Hacking**

Mr. Doug McRae,[[48]](#footnote-48) a former McGraw-Hill executive specializing in test development, has persistently questioned Smarter Balance’s methodology and timetable in testimony before the California state board of education and published articles. [[49]](#footnote-49),[[50]](#footnote-50)

McRae commented in 2013, “ . . . developing “next generation” assessments has been slow. The idea that we will have a fully functional college and/or career readiness assessment instrument by 2014-15 is fanciful thinking.”[[51]](#footnote-51) He participated in the online SBAC cut score setting activity in October 2014 and was skeptical about its technical adequacy. In a November 16, 2014 article [[52]](#footnote-52) McRae stated that any operational test needs more than qualified test questions to yield valid scores. It must also have valid scoring rules to generate meaningful scores for students, for teachers, for parents and for valid aggregate scores for schools, districts and important subgroups of students.

Setting cut scores on field test data gathered on a test that is not demonstrated to have adequate technical quality -- especially on a nascent “next generation” computer-adaptive test that is piloted simultaneously with test items -- to assess mastery of standards that have not been fully implemented in classrooms across the state is “bizarre.” More specifically, one state education department psychometrician described the process as follows:

It’s really bizarre to set cut scores based on field-test data. You can’t possibly project” accurately what proportions of students will score at the four levels of the test. He and other assessment experts said that field-test data are not good predictors of performance on the operational test because students are unfamiliar with the test, and often teachers have had less experience teaching the material that’s being tested. [[53]](#footnote-53)

McRae wrote,

*It is quite clear to me that the cut-score-setting exercises conducted by Smarter Balanced . . . will not produce final or valid cut scores for timely use with spring 2015 Smarter Balanced tests*. . . . [States that plan] to use the *cut scores recommended by the panels that met in October for disseminating millions of test scores in spring 2015 . . . are faced with the prospect that those scores will have to be “recalled” and replaced with true or valid scores just months after incorrect scores are disseminated*. [emphasis added] This is not a pretty picture for any large-scale statewide assessment program.

Further, McRae contends,

. . . *84 % of the questions were either multiple-choice or “check-the-box” questions that could be electronically scored, and these questions were very similar or identical to traditional “bubble” tests*. Only 16 percent of the questions were open-ended questions, which many observers say are needed to measure Common Core standards.

The online exercise used a set of test items with the questions arranged in sequence by order of difficulty, from easy questions to hard questions. The exercise asked the participant to identify the first item in the sequence that a Category 3 or B-minus student would have less than a 50 percent chance to answer correctly. I identified that item after reviewing about 25 percent of the items to be reviewed. *If a Category 3 or proficient cut score is set at only 25 percent of the available items or score points for a test that has primarily multiple-choice questions, clearly that cut score invites a strategy of randomly marking the answer sheet. The odds are that if a student uses a random marking strategy, he or she will get a proficient score quite often. This circumstance would result in many random (or invalid and unreliable) scores from the test, and reduce the overall credibility of the entire testing program.* [emphasis added]

In addition to problems of test content and scoring, the online administration of the SBAC creates problems in data security. Eva Baker (2012), Director of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) stated, “With tests administered by computers, there is a strong chance that hacking of computer systems will occur. . . . I have great confidence that students who seem to be able to get into banking systems will have little trouble with our State examination technology.”[[54]](#footnote-54)

Online student assessment data is particularly vulnerable to security breaches by computer hacking, as compared to paper and pencil tests. Computer hacking is the practice of modifying computer hardware and software to accomplish a goal outside of the creator’s original purpose.[[55]](#footnote-55) According to one source,

Hacking is a systematic, tiresome process in which the attacker attempts methodically to locate computer systems, identify their vulnerabilities, and then compromise those vulnerabilities to obtain access. . . . One of the most frequently cited motivations is that hacking is fun and is like solving a game or a puzzle.[[56]](#footnote-56)

Motivations of malicious attackers range from identity theft for personal financial gain[[57]](#footnote-57) to modifying personal grades[[58]](#footnote-58) or harming victims by changing their test scores. [[59]](#footnote-59) Unscrupulous hackers may sell student data to a third party.[[60]](#footnote-60), [[61]](#footnote-61) Although there is no statutorily recognized foundation for launching lawsuits over data breaches, cyber liability[[62]](#footnote-62) can have expensive repercussions for school districts not protected by breach insurance[[63]](#footnote-63) if the victim of a data breach can show he was harmed by the negligence to employ accepted standards of data security.[[64]](#footnote-64)

**Performance Test Items in SBAC Assessments Appear To Violate HB 1490**

HB 1490 Section 161.096.2 stipulates “Quantifiable student performance data shall only include performance on locally developed or locally approved assessments, including but not limited to formative assessments developed by classroom teachers.” In an apparent violation of this language in the statute, the Smarter Balanced Assessment Consortium General Items Specifications (2012) state, “The performance tasks [commonly seen as short answer or essay questions] will be administered on a matrix-sampling basis.” (p. 10)[[65]](#footnote-65) Performance tasks are also described as follows:

[Performance tasks]…will provide a measure of the student’s ability to integrate knowledge and skills across multiple [content] standards — a key component of college- and career readiness. Performance [tasks] will be used to better measure capacities such as depth of understanding, research skills, and complex analysis, which cannot be adequately assessed with [selected response] or constructed response items. (p. 31)

The prohibition of performance items in HB 1490 protects Missouri’s students given the experimental nature of the SBAC performance test items and the persistent tensions among test developers in its design. During an invitational webinar hosted by McGraw-Hill about the next generation of state assessment programs, Linn (October 28, 2010)[[66]](#footnote-66) discussed the history of performance test items in statewide testing dating from the 1990’s, and explained that they were abandoned because of the difficulty in establishing reliability as well as the cost of administration and scoring. He explained that six to twelve performance tasks may be needed to reach a desired level of reliability, and, “The needed number of tasks for obtaining reliability higher than is likely to be practical for state assessments.” Current descriptions about the design of SBAC performance tasks are available online. [[67]](#footnote-67) Examples of practice test scoring guides for evaluating practice SBAC performance tasks are also available for public review. [[68]](#footnote-68), [[69]](#footnote-69) The availability of practice items, however, is not assurance of valid, reliable and fair testing, especially given the lack of transparency about how many test items are needed to assure reasonable reliability and precautions needed scoring procedures.

Strategies to control the cost of scoring the performance items actually increases the potential for problems in establishing reliability. With respect to the number of items needed to assure reliability of student responses, it appears that the same difficulty in establishing reliability experienced in the 1990s have resurfaced. According to EdWeek, in fall of 2012, the SBAC consortium voted to sacrifice the number of performance items from an original design that included multiple, lengthy performance tasks, to include only one such task in each subject in order to reduce the time of testing. [[70]](#footnote-70) Commissioner Nicastro was quoted as stating, “There are many unanswered questions and a lot of anxiety about the tests,” she said. “The additional rigor and higher expectations of the common standards wouldn’t make it unreasonable to expect the tests to be a little bit longer. But still, *we have some folks concerned about testing*.”[emphasis added]

SBAC Executive Director, Joe Willhoft, indicated that SBAC redesigned the original test to enable reports on student performance in specific areas of math and English/language arts because, “*The U.S. Department of Education, in particular, pressed for that*.” [emphasis added] The consortium reduced the number of items despite concerns from the SBAC technical advisory committee about a pared-down test’s ability *to report meaningfully on student,* as opposed to classroom- or district-level, *performance*; that is, “ . . . a shorter version of the test is more limited in what it can validly say about an individual student’s performance.” In January 2013, Linn co-authored a critique of the consortia tests published by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), [[71]](#footnote-71) reporting that, “. . . CRESST did not conduct a close analysis of Smarter Balanced performance and item specifications *because of limitations of the evidence*.” [emphasis added] (p. 12)

Another source of problems in establishing reliability is in the scoring of the performance items. According to the 2015 DESE Assessment Manual[[72]](#footnote-72),

CTB is collaborating with the Data Recognition Corporation (DRC) and DESE to deliver Missouri’s Spring 2015 Grade-Level Assessments. Missouri educators will use DRC’s eDIRECT online platform for enrollment and test administrator processes and INSIGHT (DRC’s online delivery system) for test delivery. CTB will provide handscoring and reporting services. (p. 3)

A review of the manual indicates there is no section describing how the scorers will be selected or how the hand scoring procedures will be standardized to minimize scorer error or bias. The McGraw-Hill CBT website describes its availability of hand scoring services and states that, “Experts train and qualify professionals to score large-scale assessment programs.”[[73]](#footnote-73) A Kelly Services jobs posting indicates that CTB is partnering with Kelly Services to employ evaluators at CTB’s scoring locations (California, Indiana, and Florida) and provides information about the qualifications CTB is seeking for SBAC test scorers.[[74]](#footnote-74) The posting states:

* Evaluators assign scores to student responses for various assessment tests.
* Extensive training is provided by specialists.
* Work is performed at one of CTB’s scoring centers.
* Basic computer skills are required.
* Qualified applicants must have a bachelor's degree or higher.
* Applicants must have knowledge of standard English-language writing conventions and/or other content-specific knowledge such as science, mathematics, social studies, etc.

On another CTB website, the salary offered in exchange for “professional services” is $11. 05 per hour for work assigned on a project basis; with most projects lasting from several days to several weeks; and, running from approximately March through June.[[75]](#footnote-75)

The arrangement may have appeal from a cost-saving perspective, but, New York City’s Department of Education cancelled a contract for hand scoring with CTB/McGraw Hill in 2013 after lengthy delays and problems at the scoring center.[[76]](#footnote-76) NYC reverted back to teacher scoring conducted at NYC central locations. [[77]](#footnote-77) Schools in Indiana experienced similar problems with CTB, causing invalidation of student tests.[[78]](#footnote-78)

The “professionals” to whom CTB refers need not be certified teachers; in fact, they need not have academic or work experience associated with education. Under this scoring system, part-time scorers having no teaching certification in any of the content areas are more qualified to determine how well Missouri’s students are performing than their teachers. As per the ESEA waiver, Missouri teachers will be evaluated using a modeling technique that incorporates student scores from their responses to the SBAC without knowledge as to who is conducting the scoring or how it is done. Therefore, their performance evaluation will be substantively affected by part-time employees in out-of-state locations who are not professional peers, have no knowledge of the teacher or the students’ classroom work, but, only have access to students’ observed test responses to tests that have no demonstrated validity or reliability. District MSIP scores will also have the hand scorer’s handiwork embedded in them.

Administration of the SBAC with performance items appears to violate Missouri law. HB 1490’s restrictions on performance test items are consistent with lessons learned about performance testing in the state of Kentucky. Richard Innes, an education analyst at Kentucky’s Bluegrass Institute for Public Policy Solutions (BIPPS) reported,

Writing Portfolios and Mathematics Performance Events both failed and have been removed from the accountability system as student product scored events. The very poor track record for this sort of material in assessment is why I have very little hope for PARCC or SBAC, both of which say they are using somewhat similar approaches, though SBAC is more aggressively doing so.

Kentucky’s experience with such assessment items should be a red flag about their efficacy, not used as a way to praise such assessment approaches.[[79]](#footnote-79)

In his new report, “Selling ‘Performance’ Assessments with Inaccurate Pictures from Kentucky,” [[80]](#footnote-80) Innes counters claims increasingly made by some national researchers that the Bluegrass State had success with past attempts at incorporating performance-testing ideas into its assessment policies.

Parsi and Darling-Hammond [authors of about “Performance Assessments: How State Policy Can Advance Assessments for 21st Century Learning”] speak positively about the implementation of performance assessments from Kentucky’s 1990’s Kentucky Instructional Results Information System (KIRIS) testing program, . . .

But there’s more to the story. The reality of what happened with performance items for assessments in Kentucky over the past two-plus decades is not a success story.

Innes notes that policymakers need to know that a number of “performance” assessment ideas – some of which were expensive to implement, like “Writing Portfolios,” “Math Portfolios” and “Performance Events” – actually worked very poorly as assessment tools in Kentucky and several essentially failed completely.

. . . Policymakers nationwide deserve a complete picture of what happened in Kentucky so they can carefully consider if recent advances might overcome the problems Kentucky faced – and was never able to overcome – with performance-assessment items.[[81]](#footnote-81)

**DESE Flouts Spending Restrictions of HB 002 in 2013 to Administer SBAC in 2015**

During the 2013 legislative session, the general assembly prohibited DESE from entering into contracts to implement common core when it passed and the governor signed HB 002. In October 2013, however, DESE entered into contract with McGraw-Hill to administer online assessments aligned to common core standards in 2014-2015. [[82]](#footnote-82) The State Board of Education approved an 18 million dollar increase from the previous testing budget. [[83]](#footnote-83) McGraw-Hill also develops the test items for SBAC.[[84]](#footnote-84)

**Federal Judge Ruled SBAC is an Illegal State Compact**

On February 24, 2015, Judge Daniel R. Green, Circuit Court of Cole County, Missouri ruled in Sauer v Nixon that the governor and the commissioner of education, and state board of education committed Missouri taxpayers to “an unlawful interstate compact to which the U.S. Congress has never consented.”[[85]](#footnote-85) In response to the ruling, DESE released this statement on February 25, 2015,

This lawsuit deals with our membership in the Smarter Balanced Assessment Consortium of states that are working together to create common tests. It does not impact the use of the tests themselves. We do not expect this to impact the administration of statewide tests this spring.

The Department’s General Counsel is reviewing the ruling and considering our legal options.[[86]](#footnote-86)

Though Judge Green’s decision should be sufficient to terminate Missouri’s membership in SBAC, DESE continues its plan to administer SBAC testing,[[87]](#footnote-87) indicating no concern that a test, developed by an illegal entity “whose existence *and operation* [emphasis added] violate the Compact Clause of the U.S. Constitution,” may be inappropriate on any level. An explanation for this apparent disregard for the rule may be found in the Missouri memorandum of agreement with SBAC-UCLA and an SBAC note “Freely-Available Policy Rationale Regarding Non-Member Access.” As a member of SBAC prior to Sauer v Nixon, Missouri was entitled to a “Vendor Specification Package,” that is “the set of requirements, analyses, specifications,

and other materials that SB provides to Member for the purpose of facilitating Member’s

use of one or more vendors for the implementation, operation, and delivery of the

Assessment System.” After the court’s decision, Missouri was still entitled to access to the tests as per the SBAC note mentioned above which states,

. . . the federal government intended that the materials developed using the Race to the Top federal grant would be free to any state. Further, some argue that a state’s prior membership fees should provide continued access to Consortium materials developed in part with those member fees. This note responds to this question and the underlying premise.

The Department expects that any state or entity wishing to administer the consortium’s assessment system (either the summative assessments or the entire assessment system) without modification after the grant period would be permitted to do so, provided they comply with the consortium’s requirements for administering the assessments. [[88]](#footnote-88)

Given this statement by SBAC, and because McGraw-Hill is Missouri’s vendor to develop items and administer the 2015 MAP assessments, and McGraw-Hill is contracted with SBAC to develop test items in English and mathematics as stated above while simultaneously contracted to develop SBAC tests and to Missouri, DESE defends its plan to continue administration of the 2015 statewide assessments as planned in Commissioner Nicastro’s letter to legislative leadership in September 2014. The fact that the Sauer v Nixon found that the operation of SBAC is illegal demonstrates DESE’s failure to grasp the comprehensive implication of the case.

**U.S. DoE Funding of SBAC Violates the Intent of Federal Laws.**

According to the SBAC website, “The Consortium’s projects are funded through a four-year, $175 million grant from the U.S. DoE, comprising 99 percent of activity resources, with the remaining support provided through generous contributions of charitable foundations.”[[89]](#footnote-89) The U.S. DoE’s funding of the SBAC violates the intent of Congress to prohibit federal intrusion in education as stated in the following laws:

Department of Education Organization Act (1979) Sec. 103 (a) *It is the intention of the Congress . . . to protect the rights of State and local governments* and public and private educational institutions in the areas of educational policies and administration of *programs and to strengthen and improve the control of such governments and institutions over their own educational programs and policies . . . (b) no provision of a program administered by the Secretary [of Education] or any such officer to exercise any direction*, supervision, or control over the curriculum, program of instruction, . . . [emphasis added];

General Education Provisions Act (20 USC § 1232a*) No provision of any applicable program shall be construed to authorize any department, agency, officer, or employee of the United States to exercise any direction, supervision, or control over the curriculum*, program of instruction, . . . [emphasis added];

No Child Left Behind Act (NCLB) Section 9529 “Prohibition On Federally Sponsored Testing (a). Notwithstanding any other provision of Federal law and except as provided in subsection (b), no funds provided under this Act to the Secretary or to the recipient of any award may be used to develop, pilot test, field test, implement, administer, or distribute any federally sponsored *national test* in reading, mathematics, or any other subject, unless specifically and explicitly authorized by law” [emphasis added].

Though the U. S. DoE may argue that, technically, it did not violate the laws because it did not fund a *single* national test, but two consortia to develop tests aligned to common standards; it most certainly did violate the laws by *exercising direction* in the development or selection of curriculum and instruction in alignment to The Common Core State Standards through funding the tests of those consortia as per the Race To The Top Grants to consortia.[[90]](#footnote-90)

**Purpose of SBAC Inconsistent with Purpose of Public Education in Missouri**

As stated by SBAC, “Test content alignment is at the core of content validity and consequential validity (Martone and Sireci, 2009).” Content alignment is a concern at several levels. A test is valid for a particular purpose and definition. A test can be valid for one purpose, but invalid for others -- the better the fit or alignment, the better greater the validity of the test. The purpose of a statewide test of student achievement in education (an important construct), therefore, is most valid if it matches the state’s purpose of education. In Missouri, the purpose of publically funded education is embedded in its constitution. Article IX, Section 1(a) states that the purpose of public education is as follows:

Free Public Schools—-Age Limit

A general diffusion of knowledge and intelligence being essential to the preservation of the rights and liberties of the people, the general assembly shall establish and maintain free public schools *for the gratuitous instruction of all persons* [emphasis added] in this state within ages not in excess of twenty-one years as prescribed by law.

The Missouri constitution is a compact with the citizens of this state guaranteeing that the public schools were for gratuitous instruction – instruction without cause or justification. The purpose of education for an enlightened citizenry, educated in the liberal arts – the education of the elite -is evident in an inscription in the dome of the state capitol. Visible from the third floor of the rotunda is the proverb,

“AS THE GOVERNMENT GIVES FORCE TO PUBLIC OPINION IT IS ESSENTIAL THAT PUBLIC OPINION SHOULD BE ENLIGHTENED IN PROPORTION”

Missourians of qualifying age are guaranteed an opportunity to receive an education that equips them for self-direction, self-government, and pursuit of happiness. Just as the Missouri constitution alludes to the purpose of public education as the preservation of personal independence; similarly, President Jefferson, author of America’s Declaration of Independence described the purpose of public education as a means of seeding the government with the best possible leaders regardless of the socio-economic status of families. He wrote in his Virginia Plan,

. . . *our plan which prescribes the selection of the youths of genius from among the classes of the poor, we hope to avail the state of those talents which nature has sown as liberally among the poor as the rich, but which perish without use if not sought for and cultivated*. [emphasis added] . . . The people . . . guardians of their own liberty.[[91]](#footnote-91)

Public education is an assurance to Americans of preserving America; not an assurance to businesses that government will shield them against risk and costs associated with workforce development. Evidence of an effort to sway legislators away from the constitutional underpinnings of publicly funded education after the governor and state board of education approved adoption of the Common Core State Standards and tests aligned to the standards is found in a report from the Educated Citizenry 2020 Committee to the General Assembly, published December 2010 . In its report, *Educated Citizenry 20/20*[[92]](#footnote-92), the committee determined that the purpose of public education was for, “. . . *all levels of the education community and the business sector to ensure that Missouri’s schools and institutions are meeting the demands of the workforce* [emphasis added] both in quality and area of preparation” (p. 4) -- a statement incompatible with the Missouri constitution.

The committee’s report is important because perceptions about the purpose of education impact the theory of action underlying test design. A test designed to evaluate student knowledge and skills needed to protect the rights and liberties of the people should focus on reading and math, and also history, government, and civics. According to SBAC, “The theory of action articulated by the Consortium illustrates the vision for an assessment system that will lead to inferences that ensure that all students are well-prepared for college and careers after high school.” (p. 14). [[93]](#footnote-93) Thus, the SBAC assessments are better aligned to the purpose of public education expressed in the Senate Education Committee report than to the purpose expressed in the Missouri constitution. As such, the SBAC is inadequate for evaluating the constitutional purpose of public education.

**Conclusion**

Review of a September 30, 2014 letter and appendices from the commissioner of education to Missouri’s legislative leadership does not appear to include evidence of test validity and reliability that would support confidence in the technical adequacy of the SBAC assessments scheduled for implementation in spring 2015. Absence of validity and reliability of the Smarter Balanced Assessments voids claims that might be made about student “college and career readiness” and school personnel evaluations linked to SBAC scores. Students, teachers and school districts are vulnerable to problems associated with an indefensible test, [[94]](#footnote-94)derived from controversial standards, promoted with unsubstantiated claims, and administered by nascent computer-adaptive technology; and therefore, the state is vulnerable to litigation stemming from its implementation and questionable uses. [[95]](#footnote-95), [[96]](#footnote-96), [[97]](#footnote-97),[[98]](#footnote-98)

**Recommendations**

The following recommendations for a resolution to veto implementation of the SBAC in Missouri and investigate decisions made by the commissioner of education and state board of education to implement the SBAC:

(1) Legislative leadership should, within the number of legislative days allowed by HB 1490, support a resolution in both the House of Representatives and the Senate to stop implementation of the Smarter Balanced Assessments in Missouri and withdraw Missouri from SBAC as per Missouri’s original memorandum of understanding with SBAC;

(2) The resolution should also request an independent review of technical adequacy by a testing expert having experience in development and evaluation of nationally used, standardized tests, and is not associated with the SBAC or PARCC consortia, state or federal departments of education, or testing companies that contract with any of the aforementioned entities for the purpose of detecting fraud;

(3) The resolution should request a thorough investigation of appropriate uses of statewide standardized tests in Missouri (for example, whether student growth models relying on standardized test results, as required by the U.S. DoE’s No Child Left Behind Waiver Renewal application, are appropriate) for the purpose of challenging the conditions of the ESEA waiver offered by the U.S. DoE;

(4) The resolution should request an immediate audit of DESE and State Board of Education for the purpose of determining the legality of continued administration of SBAC in Missouri schools for purposes stated in Missouri’s original ESEA waiver application and prospective renewal application.

Appendix A

**Uses of Statewide Standardized Tests in Missouri &**

**Importance of Technical Adequacy Information for Test Selection**

In Missouri, standardized assessments of student achievement are incorporated into a state board approved assessment plan that (rightly or wrongly) uses student achievement test scores for at least five purposes. Since the 1990s, purposes include (1) tracking of student achievement, and may be used in identifying students for referral to a special education evaluation as per IDEA; (2) informing development of curriculum and instruction to ensure equitable access to education opportunities for all children as per NCLB; and (3) monitoring performances of school districts; (4) teacher evaluations; and (5) district accreditation.

The first three uses of student scores on standardized tests are traditionally associated with the advent of the No Child Left Behind Act. A new purpose for the use of student test scores is required to meet conditions of Missouri’s ESEA (No Child Left Behind) Waiver. That is, teacher evaluations are now tied to student test scores in a statistical process known as Value Added Modeling (VAM). Despite the fact that Missourians soundly rejected Amendment 3 which incorporated VAM into its proposal, VAM continues to be required by the U.S. DoE as a condition of relief from the ill-conceived condition of the NCLB that 100% of K-12 students score as *proficient* in math and English language arts by 2014.

Also, as a result of recent education reform movement associated with the 2008 federal regulations regarding Annual Yearly Progress and policies of the current administration, the Missouri School Improvement Plan (MSIP5) includes student growth scores derived from common core aligned statewide assessments as a component of the protocol used for district accreditation. Currently, HB 1490 gives teachers and school districts a reprieve from evaluation using test scores administered in 2015.

Selection of a statewide achievement test, such as SBAC, must be legally defensible. That is, it must meet professional standards of validity and reliability, for all of the above purposes and uses. The SBAC website states that SBAC is a next generation assessment that includes the design and development of new item types, and new ways of scoring student responses, within the context of, developing technologies that will be used to deliver, score and present results from assessments. [[99]](#footnote-99)The explosion of nascent test item design and test delivery system is cause for concern that the time constraints do not allow for adequate development of either. For example, SBAC’s (2012) General Items Specifications document[[100]](#footnote-100) states:

Evidence-centered assessment design (ECD) is an approach to creating educational assessments in terms of *evidentiary arguments built upon intended constructs* [that is, concepts or ideas], with explicit attention paid to the *potential influence of unintended constructs* [emphasis added] (Mislevy, Steinberg, & Almond, 2003). (p. 6)

SBAC’s Year Three Report (2014)[[101]](#footnote-101)reveals,

Smarter Balanced continued to experience challenges in Year 3 around adherence to established timelines and making sure the items and tasks developed met the consortium’s quality criteria. The consortium should continue to evaluate whether its quality control processes are sufficient and provide close oversight over the development of future items and tasks to ensure that established timelines and quality criteria are being met. In Year 4, as Smarter Balanced develops additional items and tasks, it will need to be attentive to areas where it will need to improve the overall performance of the item pool. (p. 27)

Simply put, interpretations of what students know and are able to do are basically, educated guesses -- inferred from responses to a sample of test items. The items are designed from a logical chain of reasoning, careful control of interfering factors, and good-faith efforts at representing a universe of knowledge and skills as fully as possible. The trustworthiness of what can truthfully be said about what students know and are able to do from test results is directly linked to the integrity of the underlying assumptions and concepts from which test items are made and the quality of the tests. That integrity of the assessment is defended by evidence of the overall validity of the test with respect to the purposes for which it is used and by demonstrating acceptable levels of reliability and validity of test items.

Constructing goods assessments is time consuming and tedious. The short timeline and multiple purposes of the SBAC pilot and field tests are not indicative of the care and patience needed to develop a trustworthy, legally defensible assessment, capable of providing valid scores to serve all of the purposes described above.

Appendix B

**Steps To Generate Evidence That Test Items Are Valid And Reliable**

Stanley Rabinowitz, Director of Assessment & Standards Development Services at the National Center on Evaluation, Standards and Student Testing (CRESST), and Director of WestEd/CRESST’s Assessment and Accountability Comprehensive Center, and Project Manager for SBAC, outlined 11 steps in the development of large scale assessments. He wrote,

Development or selection of large scale assessments entails considerations that are universal for assessments for all populations, and others that are more relevant for the ELL population *particularly when testing for high stakes purposes*, such as school and district accountability under NCLB. Thus, *it is important to identify for each major assessment development step, the specific actions that must be taken to ensure valid and fair testing for all participating students*, [emphasis added] including ELLs. [[102]](#footnote-102)

Illustration 1 depicts a projected timeline describing an overview of SBAC development from Kansas Commissioner of Education Diane DeBacker (presented to Phi Delta Kappa, October 18, 2011).[[103]](#footnote-103) Note that the master plan for the development of the tests and item specifications was organized before the release of the final draft of standards in June 2010 and the projected adoption of the standards by all states was even later. Illustration 2, on page 9 depicts a flow chart describing the development of test items aligned to targeted academic standards.

According to illustration 1, item development, review and piloting were all scheduled for completion in the 2012-2013 school years, at the same time the online administration of SBAC was also piloted. SBAC’s Race to the Top Year 3 report confirms this was the case on page 9, “Smarter Balanced gathered essential information from the pilot test about the accessibility features of the assessment system, the performance of individual test items, and the functionality of the test platform.“[[104]](#footnote-104) The national scale and multi-tasking within the pilot phase is inconsistent with Rabinowitz’s description of piloting test items on small groups, prior to large scale test piloting and formal field testing. College Board authors Luecht and Sireci (2011) put it this way,

Theoretically, if the test-bank or item pool is thoroughly checked before it is activated and if the computerized test delivery software and associated algorithms are fully tested and found to be robust under all potential problem scenarios, and if all data references for interactions between the examinees and the items are logged without error, additional QC may not be necessary. However, few if any CBT programs consistently meet these conditions on an ongoing basis and many QC/QA errors probably go undetected, altogether. (p. 15) [[105]](#footnote-105)

Field testing, conducted in spring 2014, culminated in a document that reported survey data, but did not report test quality data.[[106]](#footnote-106)

Illustration 1. Smarter Balanced Timeline



Note, each step of test and item development influences and constrains subsequent ones. That is, any omissions in prior stages will result in those same alignment gaps in the actual test. According to SBAC’s Race to the Top Year 3 report, development of quality test items was so problematic; the item development process has revised. The report notes,

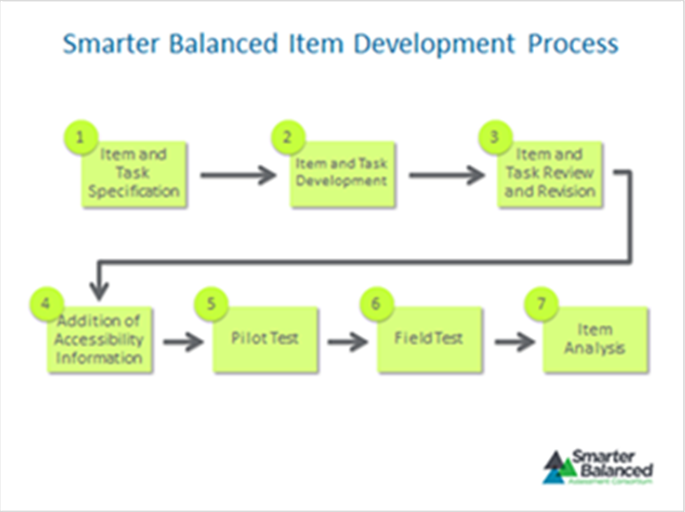
Several challenges were noted by Smarter Balanced during item development, particularly regarding the quality control measures. The Department’s Technical Review in spring 2013, which brought together national assessment experts to review the consortium’s progress, also highlighted the need for Smarter Balanced to review and revise its item development process prior to the next phase of item development. (p. 9)

SBAC described difficulties in test item development, but provided no data to justify continued administration of the SBAC for evaluating students. The concluding section of the report states,

Smarter Balanced continued to experience challenges in Year 3 around adherence to established timelines and making sure the items and tasks developed met the consortium’s quality criteria. The consortium should continue to evaluate whether its quality control processes are sufficient and provide close oversight over the development of future items and tasks to ensure that . . . quality criteria are being met. (p.27)

Stanley Rabinowitz, Director of Assessment & Standards Development Services at WestEd, serves as the lead SBAC Project Manager at West Ed.[[107]](#footnote-107)T able 1 outlines Rabinowitz’s 11 steps of test development[[108]](#footnote-108) and includes descriptions of each step as well as excerpts of commentary pertaining to SBAC as related to the step.

Illustration 2. Flow chart of test item development[[109]](#footnote-109)



**Rabinowitz’s 11 Steps Required for Large Scale Assessment Development and**

**Commentary regarding SBAC Uses and Development**

(NOTE: Italics in the table are added)

Assessment analysis is a process of *making claims based on reasoning from evidence*. That evidence is produced by student responses to test items. Psychometric data of reliability and other validity analyses establish the sufficiency, or quality, of the evidence for substantiating each claim.

|  |  |  |
| --- | --- | --- |
| Steps **1. Determine the purpose** of the test | | |
| Feature of Step | Explanation of Step | Evidence gathered about SBAC |
| Identify specific information about the prioritization of the targeted content**,**  Analysis of targeted content represents an important early warning strategy for supporting the content validity of the tests | (1) students. The theory of action . . . an assessment system that will lead to inferences that ensure that all students are *well-prepared for college and careers after high school*.(p. 14) <http://www.teapartymedia.net/20141214/Smarter%20Balanced%20Assessment%20Consortium.pdf>  *Claims made from test results must be aligned to the purpose*(s) of the test.  See claims Tables 2 and 3, next page. | *To achieve the goal that all students leave high school ready for college and career, Smarter Balanced is committed to ensuring that assessment* and *instruction embody the CCSS and that all students*, regardless of disability, language, or subgroup status, *have the opportunity to* learn this valued content and *to show what they know and can do*.  Marc Tucker (2013) stated, “being ready to be successful in the first year of a typical community college program is tantamount to being ready for both college and work.  [http://www.ncee.org/wp- content/uploads/2013/05/NCEE\_ExecutiveSummary\_May2013.pdf](http://www.ncee.org/wp-%20content/uploads/2013/05/NCEE_ExecutiveSummary_May2013.pdf)  Stotsky & Wurman (2010) wrote, “There has been a striking lack of public discussion about the definition of college readiness (e.g., for what kind of college, for what majors, for what kind of credit-bearing freshman courses) and whether workplace readiness is similar to college-readiness (e.g., in what kind of workplaces, in the non-academic knowledge and skills needed).” <http://stopccssinnys.com/uploads/Emperors_New_Clothes_National_Standards_Weak_College_and_Career_Pioneer__61.pdf> |

**Explicit Claims Derived From the Common Core English Language Arts/Literacy And Math Standards**

Explicit claims derived from the Common Core English language arts/literacy and math standards are in Tables 2 and 3 below:

Table 2.

|  |  |
| --- | --- |
| Claims for English language arts/literacy | |
| Reading | “Students can read closely and critically to comprehend a range of increasingly complex literary and informational texts |
| Writing | “Students can produce effective writing for a range of purposes and audiences” |
| Speaking/Listening | “Students can employ effective speaking and listening skills for a range of purposes and audiences” |
| Research/Inquiry | “Students can engage appropriately in collaborative and independent inquiry to investigate/research topics, pose questions, and gather and present information |
| Language Use | “Students can skillfully use and interpret written language across a range of literacy tasks. |

Table 3.

|  |  |
| --- | --- |
| Claims for Mathematics | |
| Concepts &  Procedures | “**Students can explain and apply mathematical concepts** and carry out mathematical procedures with precision and fluency.”\* |
| Problem Solving | “Students can frame and solve a range of complex problems in pure and applied mathematics.” |
| Communicating Reasoning | “**Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others.”\*** |
| Data Analysis and Modeling | “Students can analyze complex, real-world scenarios and can use mathematical models to interpret and solve problems. |

\*Criticisms of these mathematics targets/claims is that the constructed written responses to complex math questions which must be read embeds language as an extraneous factor in math testing. Therefore, the resultant score is *not a math score*, but, a reading and writing score. Populations most affected by the verbal component of these constructed responses to math questions are males, students for whom English is a second language, students with reading or writing disabilities, and students whose home language is not standard English which is frequently associated with lower socio economic status. Students, who receive a math score below proficient due to language barriers rather than math ability, will be erroneously recommended for remediation in math – robbing students of time for instruction in their actual area of need.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | (2) Teachers need to learn how to serve the instructional needs of their students | Mastery of standards should ensure that the student is able to participate fully in the instructional experience necessary to achieve proficiency in challenging content across all subject areas. Test scores are reviewed to identify areas of instruction needing improvement.  National Academies letter to Sec. Arne Duncan (2009) “Tests that mimic the structure of large-scale, high-stakes, summative tests, which lightly  sample broad domains of content taught over an extended period of time, are unlikely to provide the kind of fine-grained, diagnostic information that teachers need to guide their day-to-day instructional decisions. In addition, an attempt to use such tests to guide instruction encourages a narrow focus on the skills used in a particular test—“teaching to the test”—that can severely restrict instruction. Some topics and types of performance are more difficult to assess with large scale, high-stakes, summative tests, including the kind of extended reasoning and problem solving tasks that show that a student is able to apply concepts from a domain in a meaningful way. The use of high-stakes tests already leads to concerns about narrowing the curriculum towards the knowledge and skills that are easy to assess on such tests; (p. 11) |
|  | |
|  | | (3) District and state officials need to determine the effect of their assessment and accountability policies and provide appropriate support at the classroom level  (3) District and state officials need to determine the effect of their assessment and accountability policies and provide appropriate support at the classroom level (CONTINUED) | National Academies letter to Sec. Arne Duncan (2009) states, “When scores are averaged at the classroom, school, district, or state level, some of these sources of measurement error (e.g., guessing or momentary distractions) may average out, but other sources of *error become much more salient*. Average scores for groups of students are affected by exclusion and accommodation policies (e.g., for students with disabilities or English learners**),** *retest policies for absentees, the timing of testing over the course of the school year, and by performance incentives that influence test takers’ effort and motivation. Pupil grade retention policies may influence year-to-year changes in average scores for grade-level cohorts*.” p. 3  BOTA’s previous work on potential approaches to developing  common standards and assessments (Uncommon Measures, 1999; Embedding Questions, 1999)concluded that this aspiration is very difficult to accomplish in practice. The fundamental problem relates to dissimilarities across states in their standards, instruction and curriculum, and uses of test scores, as well as the assessments themselves: these dissimilarities ultimately make assessment results incomparable.  If states indeed adopt fully common standards and develop common assessments, these concerns would be reduced, but even seemingly small deviations from fully common standards and assessments will introduce important limitations in the degree of comparability of outcomes.  For instance, to be fully comparable, the assessments would need to be administered under standardized conditions across the country. This means that time limits, the length of testing sessions, the instructions given to test takers, test security provisions, and the use of manipulatives and calculators would need to be the same everywhere. The test administration would need to be scheduled at an appropriate time during the school year such that all the students who will take the test have been exposed to and have an opportunity to learn the content covered on the test. The stakes attached to the assessment results would also need to be constant across the country to ensure that students are equally motivated to perform well. Including state specific  content on the assessments—that is, content that differs across states—introduces the issue of context effects. These state-specific items would need to be placed on the assessment in a location where they would not influence how students perform on the common items.. |
|  | | (4) Missouri’s ESEA waiver requires that teacher evaluations included student test scores as a component of the employment decision process beginning in 2015. | National Academies (2009) wrote in a letter to Sec. of Ed. Arne Duncan, “The considerable majority of experts at the workshop cautioned that although VAM approaches seem promising, particularly as an additional way to evaluate teachers, there is little scientific consensus about the many technical issues that have been raised about the techniques and their use. . . . “*As with any effort to isolate causal effects from observational data when random assignment is not feasible, there are reasons to question the ability of value-added methods to achieve the goal of determining the value added by a particular teacher, school, or educational program” (Linn, 2008, p. 3*). Teachers are not assigned randomly to schools, and students are not assigned randomly to teachers. Without a way to account for important unobservable differences across students, *VAM techniques fail to control fully for those differences and are therefore unable to provide objective comparisons between teachers who work with different populations*. As a result, value-added scores that are attributed to a teacher or principal may be affected by other factors, such as student motivation and parental support. . . . *It is not possible to*  *calculate valid value-added measures for teachers with students who have achievement levels that are too high or too low to be measured by the available tests.* (p. 9) . . .At present,*the best use of VAM techniques is in closely studied pilot projects* . . . (p. 10)  HB 1490 Sec. 161.855.4 states, “. . . any time the state board of education or the department of elementary and secondary education implements a new statewide assessment system, develops new academic performance standards, or makes changes to the Missouri School Improvement Program, *the first year of such statewide assessment system and performance indicators shall be utilized as a pilot year* for the purposes of calculating a district's annual performance report under the Missouri school improvement program. *The results of a statewide pilot shall not be used* to lower a public school district's accreditation *or for a teacher's evaluation*.” |
|  | |
|  | | (5) Missouri’s MSIP5 includes student growth scores derived from standardized testing <http://dese.mo.gov/sites/default/files/MSIP-5-comprehensive-guide.pdf> | HB 1490 Sec. 161.855.4 states, “. . . any time the state board of education or the department of elementary and secondary education implements a new statewide assessment system, develops new academic performance standards, or makes changes to the Missouri School Improvement Program**,** *the first year of such statewide assessment system and performance indicators shall be utilized as a pilot* year for the purposes of calculating a district's annual performance report under the Missouri school improvement program .*The results of a statewide pilot shall not be used to lower a public school district's accreditation* or for a teacher's evaluation.” |
| * format of the items and response modes,   •intended difficulty of the items  See items specification graphic on p. 21 | | * Selected Response - students to select one or more responses from a set of options. * Constructed Response - students to produce a text or numerical answer. These include short items that require 1-2 sentences or a number or equation, and longer items that require extended writing or mathematical explanation. Some items can be computer scored, and some require hand-scoring by trained raters. * *Performance Response* - activities that involve significant interaction of students with stimulus materials and/or engagement in problem solving. Each Performance Task asks questions that are based on a real-world situation to gauge a student’s ability to integrate knowledge and skills across multiple standards, and to better measure capacities such as depth of understanding, research and writing skills, and complex analysis, which cannot be thoroughly assessed with other item types. * Technology-Enhanced – take advantage of the computer-administered format to assess students’ skills and understanding of content by collecting non-traditional responses such as a drawn object or edited text | The Practice Tests provide a preview of the Smarter Balanced assessments. However, *users should note that they do not encompass the full range of content that students may encounter on the operational assessments* and therefore should not be used to guide instructional decisions. In addition, *the Practice Tests are presented in a “fixed form” (i.e., not computer-adaptive) and users will not receive reports or scores*. Although the Practice Tests follow a fixed-form model, the operational assessment system will be computer-adaptive. <http://dese.mo.gov/sites/default/files/asmt-sbsc-pt-fact-sheet.pdf>  Many [Performance Tasks] PTs will require up to 120 minutes in which to administer. Additional time might be necessary for prework or group work, as required by a particular task. (Items Specifications, p. 32)  HB 1490 Sec. 161.096.2 states, “*Quantifiable student performance data shall only include performance on locally developed or locally approved assessments, including but not limited to formative assessments developed by classroom teachers***.**  SBAC performance items cannot be evaluated if Missouri’s schools are to be in compliance with HB 1490.  Measured Progress/ETS Collaborative (2012) indicates that test items are designed to fit the restrictions of Artificial Intelligence (AI) scoring. From SBAC Items Specifications (p. 28): “The item types under current consideration within these Item Specifications *occupy various points along the continuum of difficulty for AI scoring methods*, consistent with the Consortium’s desire to ensure broad and deep representation of all targeted proficiencies and standards, and *to drive innovation beyond the current state of the art in AI scoring technology*. (<http://www.teapartymedia.net/20141214/Smarter%20Balanced%20Assessment%20Consortium.pdf> )  Item format and response mode were designed for computer administration and scoring; compromising clarity and ease of response by testees. Principal John Nelson (2014) at Fairground Middle School Nashua, New Hampshire wrote in an open letter, “The FMS staff collectively believe that the Smarter Balance Test is inappropriate for our students at this time and that the results from this test will not measure the academic achievement of our students; but will be a test of computer skills and students’ abilities to endure through a cumbersome task.. . . .” Teacher comments about test administration reported in his letter included “•This was more of a test on the computer skills than on the math concepts. If I was a student I would just pick out an answer and move on .•Too tedious—kids will get sick of it and just guess to move on. •Kids won’t even get past the computer directions. <http://www.nashuatelegraph.com/csp/cms/sites/Telegraph/dt.common.streams.StreamServer.cls?STREAMOID=qNlJwCxm_SW2Tj_4DQAkZWOEnW7zZSDoJSn%24GbbR4Jb4DnHN_Z%24NabiQP_5U6zLe4Aw%246wU9GSUcqtd9hs3TFeZCn0vq69IZViKeqDZhqNLziaXiKG0K_ms4C2keQo54&CONTENTTYPE=application%2Fpdf&CONTENTDISPOSITION=SBAC+feedback+FMS+teachers+.pdf> |
| * administration and scoring procedures   [an data security] | | SBAC Pilot testing by Scott Whiting | Nov 27, 2012  The Smarter Balanced Assessment Consortium (SBAC) will be conducting a pilot test in the spring of 2013. The pilot will be a computer-based administration in the content areas of English language arts/literacy (ELA) and mathematics. Items are aligned to the Common Core State Standards and will include selected response, constructed response, and performance tasks. <http://www.smcaa.org/news/NewsArticle/12-11-27/SBAC_Pilot_testing.aspx>  The assessment system will be computer-administered. Smarter Balanced anticipates a three-year phase-in period (until the 2017-2018 school year) during which school districts may administer a comparable  paper-and-pencil version of the summative assessment (p.7)  <http://www2.ed.gov/programs/racetothetop-assessment/reports/sbac-year-3.pdf> | Smarter Balanced Assessment Instruments. These are the tests that will actually be administered to students. For the computer-adaptive component*, the specific items administered to students are uniquely determined for each student based on the item-selection algorithm to be developed for the adaptive testing system. The performance tasks will be administered on a matrix-sampling basis. Thus, test alignment studies will not simply be studies of a limited number of fixed tests*. (p. 11)  [When new computer-adaptive assessments are phased in] there is always an issue of whether paper-and-pencil and computerized versions of an assessment can be implemented at the same time and be expected to yield comparable results. The issue of comparability is a very challenging one. *There are few or no operational programs that have successfully implemented parallel paper-and-pencil and computerized tests that give comparable scores. The general advice is to avoid trying to run parallel programs. This approach is expensive and the scores from the assessments are not sufficiently comparable to be used* interchangeably. (p. 9) <http://www.wested.org/resources/the-road-ahead-for-state-assessments/>  Eva Baker (2012), Director of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) stated, “With tests administered by computers, there is a strong chance that hacking of computer systems will occur. . . . I have great confidence that students who seem to be able to get into banking systems will have little trouble with our State examination technology.”  <http://www.teapartymedia.net/20141214/CRESST%202012.pdf> |
| 2. Identify and prioritize the standards on which to base test content | | | |
| Feature of Step | Explanation of Step | | Evidence gathered about SBAC |
| The Elementary and Secondary Education Act (aka No Child Left Behind) requires that state accountability assessments be aligned with state content standards.” . . .  The SBAC test is designed to reflect the expectations of content, rigor, and performance that make up the Common Core State Standards (Item Specification, p. 8) | The specifications define what knowledge, skills, and abilities are to be measured. These specifications also describe the evidence that will be collected when measuring these knowledge, skills, and abilities, and they identify the characteristics of the items and tasks that are to be designed in a way to accurately collect that evidence. | | . . . data on consortia *plans for assessing deeper learning . . .are based on the full domain of potential assessment targets* from which items and tasks will be selected for testing. *Given time constraints for test*  *administration, it will not be possible for a single test form to include all targets.* <http://www.hewlett.org/uploads/documents/On_the_Road_to_Assessing_DL-The_Status_of_SBAC_and_PARCC_Assessment_Consortia_CRESST_Jan_2013.pdf> |
| 3. Develop test specifications | | | |
| Feature of Step | Explanation of Step | | Concerns generated from published descriptions of SBAC |
| •Content specifications  •Item specification | Rabinowitz (nd) *Content assessments typically include knowledge of key vocabulary/terminology germane to the domain.*  Variety of SBAC Item Responses  -Selected Response  -Constructed Response  -Extended Response  -Performance Tasks  -Technology-Enabled  -Technology-Enhanced | | content focus – specifically, how well the Smarter Balanced tests and items/tasks will address the expectations embodied in the Smarter Balanced Content Specifications and the Common Core State Standards.  CRESST did not conduct a close analysis of Smarter Balanced performance and item specifications *because of limitations of the evidence***.** (p. 12) <http://www.hewlett.org/uploads/documents/On_the_Road_to_Assessing_DL-The_Status_of_SBAC_and_PARCC_Assessment_Consortia_CRESST_Jan_2013.pdf>  *The CCSS were not specifically developed for assessment and contain a great deal of rationale and information about instruction.* Therefore, . . ., Smarter Balanced distilled from the CCSS a set of content specifications expressly created to guide assessment development. Within each of the two subject areas at grades 3 through 8 and high school, there are four broad claims, and within each claim there are several assessment targets. . . . Because of the breadth of the individual claims, the targets within them really define the nature of the performance expectations within these statements. (p. 9) <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/ItemSpecifications/GeneralItemSpecifications.pdf> |



From Step 2

From Step 1 1

|  |  |  |
| --- | --- | --- |
| 4. Draft items consistent with general item-writing guidelines and the particular test specifications | | |
| Feature of Step | Explanation of Step | Evidence gathered about SBAC |
| developing the items and tasks described in the specifications | See graphic above,( p. 18) | In the mathematics content area, items designed to elicit mathematical graphs, equation, or expressions are generally suitable for automated scoring in its current state, although test delivery considerations must be attended to carefully. In English language arts, the extended writing tasks are in most respects suitable for existing AI [artificial intelligence] scoring capabilities, but will likely need to be augmented by human judgment where it is necessary to assess whether a text is well reasoned or well suited to the intended audience. For both content areas, the short, textual constructed-response items represent the greatest challenge for state-of-the-art capabilities for AI scoring at the present time. Because they are intended to be scored based on general evidence of students’ reasoning processes rather than referencing of specific concepts, accurate and valid AI scoring for these items will require significant technological advances. In order to mitigate the risk associated with incorporating tasks that are challenging for AI scoring into the assessment design, *the Smarter Balanced Consortium may consider constraining selected item types to better accommodate the technology.* [In other words, limit the design of items to the capabilities of the computer scoring software, rather than the capabilities of the student. |
| 5. Conduct content and bias reviews of items by **knowledgeable and trained** **external reviewers** | | |
| Feature of Step | Explanation of Step | Evidence gathered about SBAC |
| Reviewing and revising items and tasks for content, its accessibility, its sensitivity to cultural, religious, and ethnic issues, and potential bias. | Content reviews will be performed by educators. Specialists in accessibility, sensitivity and bias will perform reviews for these additional features. Based on the reviews, revisions to items and tasks may be made.  Information is added to each item that specifies how to support the accessibility needs of students, including how content is to be read-aloud, or how to present content in braille or American Sign Language. Accessibility information is added by specialists with knowledge of specific access needs.  The Bias and Sensitivity Guidelines document describes the rules agreed upon by the Smarter Balanced Assessment Consortium states for achieving fairness in test content. Only items that are in compliance with the Guidelines will be included in the Smarter Balanced assessments. Therefore, the Guidelines will help ensure that the test content is fair for test takers as well as acceptable to the many stakeholders and constituent groups within the Smarter Balanced states. | . . . while built-in accommodations may be easier to accomplish, *there will still be the validity challenge of establishing the comparability of accommodated and non-accommodated versions of the test. Similarly, and particularly in the short run, while technology-based assessment offers many new opportunities, care will need to be taken that the technology manipulation required by the assessment does not unintentionally add construct-irrelevant barriers for some students, particularly those with less access and less facility with technology. In addition, performance tasks typically include substantial reading or other linguistic demands that can provide unintended obstacles to English learners and those with low reading ability being able to show what they know.* Developers will need to be creative in designing tasks, items, and accessibility mechanisms that minimize construct-irrelevant, unintended demands. (pp. 18-19) <http://www.hewlett.org/uploads/documents/On_the_Road_to_Assessing_DL-The_Status_of_SBAC_and_PARCC_Assessment_Consortia_CRESST_Jan_2013.pdf>  The next generation of assessments will have to address these [bias] concerns by greatly expanding and refining the portfolio of EL-responsive access and accommodation practices. Even incremental validity gains are valuable. Learners on large-scale tests have faced increasingly stringent expectations for demonstrating the validity of inferences made using accommodated scores, and current practices fall far short of the necessary standard. A National Research Council committee recently concluded that existing research on accommodations for ELs and students with disabilities was inadequate to demonstrate the comparability of inferences from scores derived from accommodated versus standardized tests, or to demonstrate which accommodations produce the most valid estimates of performance (National Research Council, 2004).  . . . the first line of defense against the inclusion of unfair materials in the Smarter Balanced assessments is the judgment of trained test developers who follow these Guidelines. *Judges may, however, miss potential fairness issues that sophisticated statistical analyses later find (Bond, 1993). Therefore, both judgmental and statistical evaluations of fairness are required.* (p. 6) <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/Guidelines/BiasandSensitivity/BiasandSensitivityGuidelines.pdf> |

|  |  |  |
| --- | --- | --- |
| 6. Pilot test items **on small groups** of students and revise items using information from pilot test results prior to more formal field testing | | |
| Feature of Step | Explanation of Step | Evidence gathered about SBAC |
| *Prior* to more formal field testing; small scale, less formal tryouts, particularly of those item types that are difficult and expensive to assess (e.g., writing prompts, constructed response, performance tasks) | During piloting, *on small groups, students perform the items and tasks*. How students perceive and work through specific items would provide useful information, and results are analyzed to be sure that the items and tasks work as intended*. Items that do not function as intended are reviewed to help determine whether or not an item is eligible for future use and revised as needed*.  Proposed accommodations should also be examined during the pilot test stage so that necessary adjustments to accommodation policies could be identified and incorporated prior to a full-scale field test. Accommodations should be research-based. They should not simply be adapted from state policies for students with disabilities (Abedi, 2001; Rivera & Collum, 2006 | It appears that large groups of students participated in the pilot study, using convenience samples of entire schools, and that the pilot study was testing online test administration as well as test items, as implied by this statement in the Test Administrator User Guide, “The Smarter Balanced Scientific Sample Pilot Test (Secure Pilot Test) is open only to schools and districts that confirmed their participation. Schools and districts that volunteered for the Open Pilot Test will receive information in March 2013 about how to access that test.  The Smarter Balanced Training Test is available only to schools and students participating in the Secure Pilot Test. The Training Test allows students and teachers to practice using the testing software prior to participating in the Secure Pilot Test. “ (p. 7)  “For the Pilot Test administration, Test Administrators must use the Test Administrator Interface (TA Interface) to create test sessions and manage students’ online testing progress during the session. Students must use the secure browser to access the Student Testing Site.” (p. 7)  <http://sbac.portal.airast.org/Pilot_Test/resources/Smarter_Pilot_TA-UserGuide.pdf>  Results of the pilot study were not reported to participants. |
| 7. Field test items with a sufficiently large, representative sample of students including as feasible, students across the range of language groups and degree of English proficiency | | |
| Feature of Step | Explanation of Step | Evidence gathered about SBAC |
| The actual use of an assessment item or task by a selected *student sample of adequate size and demographic to meet acceptable research standards* | The field test is considered a “dress rehearsal” for the official, operational assessment. Items are included in final form and assessed under conditions as close to a live administration as possible.  It is essential that field testing involve a sufficiently large, *representative sample* of students. For example, sufficient numbers of English Language Learners (ELLs) must be included in the field test sample (across all field test forms), including as feasible, students across the range of language groups and degree of English proficiency | In 2013, the SBAC the assessment system was tried out to ensure it was working as intended. Modifications will be made based on findings during the field test  DESE documents indicate a convenience, rather than a carefully chosen representative sample of students, participated in the 2014 field test. |
| 8. Conduct Differential Item Functional (DIF) analyses on the field test data | | |
| Feature of Step | Explanation of Step | Evidence gathered about SBAC |
| Differential item functioning (DIF) to ensure that items are performing appropriately | DIF analyses should be conducted on Subgroups of students, such as, major ELL subgroups, as available and appropriate. DIF results should be reviewed to determine if certain item types and language features consistently disadvantage ELL subpopulations or ELLs as a group. | The effectiveness of the single most commonly provided English language accommodation for ELs—reading instructions aloud—is unproven (Shafer Willner et al., 2008). Distinguishing and deploying accommodations that provide direct and indirect linguistic support to ELs in testing situations is a critical priority for both policy and practice (Rivera et al., 2006). <http://www.wested.org/resources/the-road-ahead-for-state-assessments/> |
| 9. Assemble items into operational test forms that are consistent with the test specifications | | |
| Feature of Step | Explanation of Step | Evidence gathered about SBAC |
| Based on field test results, operational test forms are built consistent with the test specifications and blueprints | Smarter Balanced end-of-year, on-demand assessments will utilize computer-adaptive testing (CAT), where complex algorithms are used to customize the items administered  to each individual based on his or her ability level, which is inferred from responses to prior items. <http://www.cse.ucla.edu/products/reports/R823.pdf> | Both consortia have been optimistic about the promise of automated constructed-response and performance task scoring and have incorporated that optimism into their cost estimates for the summative assessment. Both are estimating summative testing costs at roughly $20 per student for both subject areas. In the absence of promised breakthroughs, those costs will escalate, there will be enormous demands on teachers and/or others for human scoring, and the feasibility of timely assessment results may be compromised. For example, Smarter Balanced has promised end-of-year results, including those from performance tasks, within two weeks of the end of its spring testing window,. . . Meeting these expectations also will require innovation in scoring services |
| 10. Conduct technical analyses (reliability, equating) | | |
| Feature of Step | Explanation of Step | Evidence gathered about SBAC |
| Rather than just computing overall reliability indices, many programs report the reliability at key points of the score scale, such as the cut score which divides proficient from below proficient performance. | Lower reliability for some subgroups will help identify areas of potential invalidity that states will need to address. | Smarter Balanced has been very transparent in posting all of its plans and the results of its contracts. Yet, because its computer-adaptive testing approach essentially individualizes test items to every student, it may be difficult to ascertain how well deeper learning is represented for every student or overall. The test blueprint will provide rules for item  selection and presumably, those rules will include those for representing higher levels of depth of knowledge, but this is yet to be seen. (p. 19)  <http://www.hewlett.org/uploads/documents/On_the_Road_to_Assessing_DL-The_Status_of_SBAC_and_PARCC_Assessment_Consortia_CRESST_Jan_2013.pdf> |
| 11. Conduct validity studies | | |
| Feature of Step | Explanation of Step | Evidence gathered about SBAC |
| States have an explicit obligation to *demonstrate the validity of their assessments for all stated purposes*, especially those accountability decisions that affect students and schools (AERA/APA/NCME, 1999). | The *focus of validity studies differs somewhat, depending on the stage of development* and maturity of the assessment program(s). For example, *when assessments are newly developed, validity studies should focus on defining the targeted constructs and ensuring that items align to the standards* that operationalize those constructs. *As tests become operational, the emphasis of validity studies should shift to accessibility (lack of bias), reliability, and generalizability* | The transparency of subsequent steps of the *test development and validation process represents yet another possible challenge to assuring the representation of deeper learning in the consortia’s efforts.* <http://www.hewlett.org/uploads/documents/On_the_Road_to_Assessing_DL-The_Status_of_SBAC_and_PARCC_Assessment_Consortia_CRESST_Jan_2013.pdf>  When assessments are newly developed, such as the SBAC, validity studies should focus on defining the targeted constructs and ensuring that items align to the standards that operationalize those constructs. Then, as tests become operational, the emphasis of validity studies should shift to accessibility (lack of bias), reliability, and generalizability. <http://www.ncela.us/files/uploads/11/rabinowitz.pdf> and <http://www.azed.gov/highly-qualified-professionals/files/2012/02/technical-adequacy-of-assessments-for-student-growth-082411.pdf> |

1. <http://www.moagainstcommoncore.com/documents> [↑](#footnote-ref-1)
2. <http://www.smarterbalanced.org/news/wested-selected-as-project-manager-for-smarter-balanced-assessment-consortium/> [↑](#footnote-ref-2)
3. <http://dese.mo.gov/communications/news-releases/missouri-education-department-chooses-vendor-assessments> [↑](#footnote-ref-3)
4. <http://www.ednetinsight.com/news-alerts/prek-12-market-headlines/sbac-selects-ctb-mcgraw-hill-to-develop-next-generation-of-assessments.html> [↑](#footnote-ref-4)
5. <https://www.sde.idaho.gov/site/commonAssessment/docs/Memo_Validity_Overview_2014-09-11.pdf> [↑](#footnote-ref-5)
6. <http://www.azed.gov/highly-qualified-professionals/files/2012/02/technical-adequacy-of-assessments-for-student-growth-082411.pdf> Test consumers must distinguish between mere claims versus substantiated research

   findings regarding the technical adequacy of any assessment instrument. Validity and reliability are important criteria for discerning technical adequacy of an assessment. For a definition of key assessment terms, see A Glossary of Assessment Terms in Everyday Language,

   <http://www.ccsso.org/Documents/2006/Assessing_Students_with_Disabilities_Glossary_2006.pdf> [↑](#footnote-ref-6)
7. <https://www.dropbox.com/sh/443a9u9wlc6btyx/AADmQHwSouyFKaelEpQvUt9fa?dl=0> [↑](#footnote-ref-7)
8. <https://www.sde.idaho.gov/site/commonAssessment/docs/Memo_Validity_Overview_2014-09-11.pdf> [↑](#footnote-ref-8)
9. <http://www.smarterbalanced.org/news/states-move-forward-smarter-balanced/> [↑](#footnote-ref-9)
10. <http://www2.ed.gov/programs/racetothetop-assessment/reports/sbac-year-3.pdf> [↑](#footnote-ref-10)
11. <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/10/SmarterBalanced_FieldTest_Report.pdf> [↑](#footnote-ref-11)
12. <http://edsource.org/2014/smarter-balanced-tests-are-still-a-work-in-progress/69828#comment-15759> [↑](#footnote-ref-12)
13. <http://www.edweek.org/ew/articles/2014/11/17/13sbac.h34.html?tkn=MWYFiGvymGxUQP4DD7DfEWwqn7OqwkiWFQrw&print=1> [↑](#footnote-ref-13)
14. <http://www.teapartymedia.net/20141214/CRESST%202012.pdf> [↑](#footnote-ref-14)
15. <http://www.smarterbalanced.org/news/smarter-balanced-to-convene-member-states-and-education-experts-in-los-angeles/> [↑](#footnote-ref-15)
16. <http://www.plainsite.org/dockets/tgx6rejn/missouri-western-district-court/state-of-missouri-et-al-v-the-mcgraw-hill-companies-inc-et-al/> [↑](#footnote-ref-16)
17. <http://www.law360.com/cases/51ba48526db4021733000001/dockets> [↑](#footnote-ref-17)
18. <http://dese.mo.gov/communications/news-releases/missouri-education-department-chooses-vendor-assessments> [↑](#footnote-ref-18)
19. <https://americanprinciplesproject.org/wp-content/uploads/Sauer-v.-Nixon-Temporary-Restraining-Order-copy.pdf> [↑](#footnote-ref-19)
20. <http://www.smarterbalanced.org/news/smarter-balanced-chief-state-school-officers-meet-to-advance-assessment-system-design/> [↑](#footnote-ref-20)
21. <http://www.democratandchronicle.com/story/news/2014/03/10/teachers-union-sues-performance-reviews/6256961/> [↑](#footnote-ref-21)
22. <http://www.washingtonpost.com/blogs/answer-sheet/wp/2014/04/30/houston-teachers-sue-over-controversial-teacher-evaluation-method/> [↑](#footnote-ref-22)
23. <http://newyork.cbslocal.com/2014/11/04/long-island-teacher-sues-state-over-grade-claims-evaluation-system-is-flawed/> [↑](#footnote-ref-23)
24. <http://www.takepart.com/article/2013/04/16/florida-teachers-union-sues-state-teacher-evaluations> [↑](#footnote-ref-24)
25. <https://www.dropbox.com/sh/443a9u9wlc6btyx/AADmQHwSouyFKaelEpQvUt9fa?dl=0> [↑](#footnote-ref-25)
26. <https://www.sde.idaho.gov/site/commonAssessment/docs/Memo_Validity_Overview_2014-09-11.pdf> [↑](#footnote-ref-26)
27. Regulations: Part 300 / D / 300.304 / c / 1 / iii [↑](#footnote-ref-27)
28. NCLB, 2001, §200.2(b)(4)(I,ii) [↑](#footnote-ref-28)
29. www2.ed.gov/policy/edsec/guid/saaguidance03.doc [↑](#footnote-ref-29)
30. <http://www.gpo.gov/fdsys/pkg/FR-2010-04-09/pdf/2010-8176.pdf> pp. 37806-37807, 37809 [↑](#footnote-ref-30)
31. <http://www.washingtonpost.com/politics/how-bill-gates-pulled-off-the-swift-common-core-revolution/2014/06/07/a830e32e-ec34-11e3-9f5c-9075d5508f0a_story.html> [↑](#footnote-ref-31)
32. <http://www.gpo.gov/fdsys/pkg/FR-2010-04-09/pdf/2010-8176.pdf> [↑](#footnote-ref-32)
33. <http://media.columbiamissourian.com/multimedia/2009/06/30/media/Common_Standards_MOA_FINAL_signed.pdf> [↑](#footnote-ref-33)
34. <http://dese.mo.gov/sites/default/files/qs-esea-waiver-mo-flexibility-request.pdf> [↑](#footnote-ref-34)
35. <http://www.ctb-e.com/emails/pdfs/Asst_Webinar_10_28_10_final.pdf> [↑](#footnote-ref-35)
36. Linn co-authored a CRESST 2013 critique of SBAC’s development <http://www.cse.ucla.edu/products/reports/R823.pdf> [↑](#footnote-ref-36)
37. <https://www.sde.idaho.gov/site/commonAssessment/docs/Memo_Validity_Overview_2014-09-11.pdf> [↑](#footnote-ref-37)
38. <https://dese.mo.gov/sites/default/files/MAPFieldTest.pdf> [↑](#footnote-ref-38)
39. <http://dese.mo.gov/sites/default/files/DESELegislativeQ&A.pdf> [↑](#footnote-ref-39)
40. <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/ItemSpecifications/GeneralItemSpecifications.pdf> [↑](#footnote-ref-40)
41. <https://www.dropbox.com/sh/443a9u9wlc6btyx/AADmQHwSouyFKaelEpQvUt9fa?dl=0> [↑](#footnote-ref-41)
42. <http://montereycountyschools.blogspot.com/2013/01/more-on-proposed-revamping-of-testing.html> [↑](#footnote-ref-42)
43. <http://missourieducationwatchdog.com/why-this-missouri-momteacher-is-refusing-the-sbac-test-for-her-children/> [↑](#footnote-ref-43)
44. <http://dese.mo.gov/sites/default/files/am/documents/CCR-15005_Change%20in%20Spring%202015%20State%20Assessment_0.pdf> [↑](#footnote-ref-44)
45. <http://www.gpo.gov/fdsys/pkg/FR-2010-04-09/pdf/2010-8176.pdf> [↑](#footnote-ref-45)
46. <http://www2.ed.gov/programs/racetothetop-assessment/reports/sbac-year-3.pdf> [↑](#footnote-ref-46)
47. <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/10/SmarterBalanced_FieldTest_Report.pdf> [↑](#footnote-ref-47)
48. Doug McRae, is a retired measurement specialist who worked at CTB/McGraw-Hill — one of the nation's foremost education test publishers — for 15 years, first as a senior evaluation consultant and later as vice president of publishing. While at CTB/McGraw-Hill, he was in charge of design and development of K-12 tests widely used across the U.S. From 1994 to 2008 he was self-employed as a consultant who advised on the initial design and development of the California STAR Program and advising policymakers on the design and implementation of K-12 standardized tests in California. He has a Ph.D. in Quantitative Psychology from the University of North Carolina, Chapel Hill. [↑](#footnote-ref-48)
49. <http://toped.svefoundation.org/2011/03/07/common-core-groups-should-be-asked-plenty-of-questions/> [↑](#footnote-ref-49)
50. <http://montereycountyschools.blogspot.com/2013/01/more-on-proposed-revamping-of-testing.html> [↑](#footnote-ref-50)
51. <http://montereycountyschools.blogspot.com/2013/01/more-on-proposed-revamping-of-testing.html> [↑](#footnote-ref-51)
52. <http://edsource.org/2014/smarter-balanced-tests-are-still-a-work-in-progress/69828#comment-15759> [↑](#footnote-ref-52)
53. <http://www.edweek.org/ew/articles/2014/11/17/13sbac.h34.html?tkn=MWYFiGvymGxUQP4DD7DfEWwqn7OqwkiWFQrw&print=1> [↑](#footnote-ref-53)
54. <http://www.teapartymedia.net/20141214/CRESST%202012.pdf> [↑](#footnote-ref-54)
55. <http://www.wisegeek.org/what-is-computer-hacking.htm> [↑](#footnote-ref-55)
56. <http://www.encyclopedia.com/topic/Hacking.aspx> [↑](#footnote-ref-56)
57. <http://www.crn.com/news/security/18829740/student-charged-with-hacking-information-theft.htm> [↑](#footnote-ref-57)
58. <http://blog.tmcnet.com/blog/tom-keating/voip/college-student-could-get-5-years-for-hacking-voip-system-changing-gra.asp> [↑](#footnote-ref-58)
59. <http://www.myfoxatlanta.com/story/27864970/student-involved-in-school-hacking> [↑](#footnote-ref-59)
60. <http://pixelenemy.com/12-year-old-hacker-sells-information-to-anonymous-for-video-games/> [↑](#footnote-ref-60)
61. <http://thetechjournal.com/internet/web-security/hacker-sells-data-of-4-million-us-citizens.xhtml> [↑](#footnote-ref-61)
62. <http://www.aon.com/about-aon/intellectual-capital/attachments/risk-services/cyber_liability_higher_education.pdf> [↑](#footnote-ref-62)
63. <http://netdiligence.com/files/CyberLiability-0711sh.pdf> [↑](#footnote-ref-63)
64. <http://thejournal.com/articles/2013/10/17/how-little-data-breaches-cause-big-problems-for-schools.aspx> [↑](#footnote-ref-64)
65. <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/ItemSpecifications/GeneralItemSpecifications.pdf> [↑](#footnote-ref-65)
66. <http://www.ctb-e.com/emails/pdfs/Asst_Webinar_10_28_10_final.pdf> (slides 35- 44) [↑](#footnote-ref-66)
67. <http://commoncore.fcoe.org/sites/commoncore.fcoe.org/files/resources/SBAC-04PerformanceTaskSpecs-SC1%5B1%5D.pdf> and <https://view.officeapps.live.com/op/view.aspx?src=http%3A%2F%2Fwww.k12.wa.us%2FSMARTER%2Fpubdocs%2FSmarterBalanced_SampleItems_Webinar.pptx> [↑](#footnote-ref-67)
68. <http://sbac.portal.airast.org/wp-content/uploads/2013/07/Grade5MathPT.pdf> [↑](#footnote-ref-68)
69. <http://sbac.portal.airast.org/wp-content/uploads/2013/07/Grade8MathPT.pdf> [↑](#footnote-ref-69)
70. <http://www.edweek.org/ew/articles/2012/11/30/13tests.h32.html?tkn=WTQFpP2Mh9z4wbnoHTk0N7CZqB8JELOTFuZM> [↑](#footnote-ref-70)
71. <http://www.hewlett.org/uploads/documents/On_the_Road_to_Assessing_DL-The_Status_of_SBAC_and_PARCC_Assessment_Consortia_CRESST_Jan_2013.pdf> [↑](#footnote-ref-71)
72. <http://dese.mo.gov/sites/default/files/asmt-gl-tam-2015.pdf> [↑](#footnote-ref-72)
73. <http://www.ctb.com/ctb.com/control/servicesDetailAction?serviceId=410&p=services> [↑](#footnote-ref-73)
74. <http://www.kellyservices.us/us/ctb-mcgraw-hill-home/#.VQD5YvO9LCQ> [↑](#footnote-ref-74)
75. <http://ctb.appone.com/MainInfoReq.asp?R_ID=39933&B_ID=46&fid=1&Adid>= [↑](#footnote-ref-75)
76. http://ny.chalkbeat.org/2013/06/18/serious-glitches-with-electronic-grading-delay-regents-scores/#.VQSLf9EtGUk [↑](#footnote-ref-76)
77. http://www.wnyc.org/story/318148-department-ed-adjust-gifted-and-talented-tests/ [↑](#footnote-ref-77)
78. http://indianapublicmedia.org/news/istep-scores-delayed-glitches-49894/ [↑](#footnote-ref-78)
79. http://www.bipps.org/; Personal communication (December 16, 2014) [↑](#footnote-ref-79)
80. http://www.bipps.org/wp-content/uploads/2015/03/Selling-%E2%80%9CPerformance%E2%80%9D-Assessments-with-Inaccurate-Pictures-from-Kentucky.pdf) [↑](#footnote-ref-80)
81. http://nelsoncountygazette.com/?p=24371 [↑](#footnote-ref-81)
82. <http://dese.mo.gov/communications/news-releases/missouri-education-department-chooses-vendor-assessments> [↑](#footnote-ref-82)
83. <http://thejournal.com/articles/2013/10/31/brief-missouri-chooses-ctb-mcgraw-hill-to-administer-online-assessments.aspx> [↑](#footnote-ref-83)
84. <http://www.ednetinsight.com/news-alerts/prek-12-market-headlines/sbac-selects-ctb-mcgraw-hill-to-develop-next-generation-of-assessments.html> [↑](#footnote-ref-84)
85. <https://americanprinciplesproject.org/wp-content/uploads/Sauer-v.-Nixon-Temporary-Restraining-Order-copy.pdf> [↑](#footnote-ref-85)
86. <http://dese.mo.gov/communications/news-releases/statement-smarter-balanced-membership-ruling> [↑](#footnote-ref-86)
87. <http://blogs.edweek.org/edweek/state_edwatch/2014/12/missouri_must_stop_payments_for_common-core_test_judge_orders.html?r=1879966624&cmp=ENL-EU-NEWS3&preview=1> [↑](#footnote-ref-87)
88. <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/07/Freely-Available-Policy-Rationale.pdf> [↑](#footnote-ref-88)
89. <http://www.smarterbalanced.org/news/smarter-balanced-chief-state-school-officers-meet-to-advance-assessment-system-design/> [↑](#footnote-ref-89)
90. <http://www.gpo.gov/fdsys/pkg/FR-2010-04-09/pdf/2010-8176.pdf> [↑](#footnote-ref-90)
91. <http://press-pubs.uchicago.edu/founders/documents/v1ch18s16.html> [↑](#footnote-ref-91)
92. <http://www.senate.mo.gov/2020Committees/Reports/EducatedCitizenry2020FinalReport_Dec2010.pdf> [↑](#footnote-ref-92)
93. <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/ItemSpecifications/GeneralItemSpecifications.pdf> [↑](#footnote-ref-93)
94. These problems can cause significant harm in light of the dramatic consequences being assigned to student scores on such tests in the Missouri School Improvement Plan; that is, negative effects on student advancement, teacher retention, local district funding and a district’s overall ability to retain control of its schools. [↑](#footnote-ref-94)
95. <http://www.democratandchronicle.com/story/news/2014/03/10/teachers-union-sues-performance-reviews/6256961/> [↑](#footnote-ref-95)
96. <http://www.washingtonpost.com/blogs/answer-sheet/wp/2014/04/30/houston-teachers-sue-over-controversial-teacher-evaluation-method/> [↑](#footnote-ref-96)
97. <http://newyork.cbslocal.com/2014/11/04/long-island-teacher-sues-state-over-grade-claims-evaluation-system-is-flawed/> [↑](#footnote-ref-97)
98. <http://www.takepart.com/article/2013/04/16/florida-teachers-union-sues-state-teacher-evaluations> [↑](#footnote-ref-98)
99. <http://www.smarterbalanced.org/> [↑](#footnote-ref-99)
100. <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/ItemSpecifications/GeneralItemSpecifications.pdf> [↑](#footnote-ref-100)
101. <http://www2.ed.gov/programs/racetothetop-assessment/reports/sbac-year-3.pdf> [↑](#footnote-ref-101)
102. <http://www.ncela.us/files/uploads/11/rabinowitz.pdf> [↑](#footnote-ref-102)
103. Kansas withdrew from SBAC and the state of Kansas is developing its own statewide assessments. <http://www2.ed.gov/programs/racetothetop-assessment/reports/sbac-year-3.pdf> [↑](#footnote-ref-103)
104. <http://www2.ed.gov/programs/racetothetop-assessment/reports/sbac-year-3.pdf> [↑](#footnote-ref-104)
105. <http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-12-review-models-for-computer-based-testing.pdf> [↑](#footnote-ref-105)
106. Appendix K of Commissioner Nicastro’s letter and <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2014/10/SmarterBalanced_FieldTest_Report.pdf> [↑](#footnote-ref-106)
107. <http://www.smarterbalanced.org/news/wested-selected-as-project-manager-for-smarter-balanced-assessment-consortium/> [↑](#footnote-ref-107)
108. <http://www.ncela.us/files/uploads/11/rabinowitz.pdf> [↑](#footnote-ref-108)
109. <http://www.smarterbalanced.org/smarter-balanced-assessments/item-writing-and-review/> [↑](#footnote-ref-109)