

Evidence of Lack of Psychometric Validity and Reliability of AIR's UTAH SAGE Test and the FSA Based on Official Statements and Correspondence

March 21, 2015

Psychometrics & Validity

3/4/15 Senate Appropriations Subcommittee on Education

<http://thefloridachannel.org/videos/3415-senate-appropriations-subcommittee-on-education/>

Testimony that begins at 56:00

Senator Simmons: Commissioner I have a couple questions based upon what is occurred over the last few days and I'm looking at it from the point of view of the Deborah P versus Turlington case which I gather you're aware of, that requires in order to place sanctions or standards or the possibility of depriving a student of, for example, advancement is a property interest is protected. The requirement is is that these tests are administered must be psychometrically valid and tested and must be properly administer. With the public admission by AIR that there have been errors, at least there's two these have already been tested do you still intend to impose upon them any kind of standards of accountability or is this in fact going to be a situation which basically wait and see how the outcome of the testing is.

Commissioner Stewart: It did not affect the ability for individuals to be able to answer and it did not affect the quality of the assessment so we are aligned with the assessments to the standards and those students are able to respond appropriately.

Senator Simmons: Since Thursday, as I understand, the requirement that the test be psychometrically valid, has the testing already occurred of this test to determine that it is psychometrically valid?

Commissioner: Yes it has

Senator Simmons: Who did the psychometric testing?

Commissioner Stewart: AIR

Senator Simmons: And has anyone reviewed AIRs psychometric testing... it being the same entity that has admitted that it is made errors and taking full accountability for those errors.

Commissioner Stewart: Yes they are absolutely has been reviewed we have our own team that is the test development team that has reviewed that and we are certain that the content of the test is absolutely psychometrically valid and reliable.

Senator Simmons: When did they test it?

Commissioner Stewart: we can provide that information to you. I don't know the exact dates, but we can provide the information to you

Senator Simmons: and what is the size of the grouping to which they tested the students?

Commissioner Stewart: check but happy to provide that to you as well

Senator Simmons: How often was it was it done?

Commissioner Stewart: happy to provide you that as well.

Senator Simmons: Were they Florida students?

Commissioner Stewart: That I can answer, no they were not.

Senator Simmons: They were not Florida students?

Commissioner Stewart: That's correct.

Senator Simmons: Are these the Utah students

Commissioner Stewart: Yes Utah students did to experience these some of these questions and it was field-tested there.

Senator Simmons: Thank you.

Alpine District Utah School Board Members sent a letter on September 18, 2014. In that letter they asked for certain documentation pertaining to the validity and reliability of the SAGE test. Here is the excerpt with the full letter attached:

Finally, we have the following questions related to the validity and reliability of the SAGE testing. We understand that this information would not be protected by copyright, and therefore, could be provided to us, as elected officials.

1. Normative Sample details (who took the test).
2. Coefficient Alpha reliability.
3. Content description validity.
3. Differential item functional analysis.
4. Criterion Prediction Validity.
5. Construct Identification Validity.
6. Other types of validity scales/constructs that are applicable only to CAT test designs.

When Commissioner Stewart made the statement that AIR provided the psychometric validity of our FSA test and that the validity was based on Utah SAGE test, red flags popped up for us. We reached out the Wendy Hart, Alpine School Board member to see if she had ever received any reports. Below is her response.

Subject RE: Assessment questions

Date:: Monday, March 9, 2015 11:49:49 AM Eastern Daylight Time

From:: Wendy Hart <whart2000@hotmail.com>

To:: Suzette Lopez suzettelopez@earthlink.net

You can show them the letter that we sent to Judy Park at the State Office of Ed requesting the evidence of the reliability. If elected local school board members can't get access to it, then who can? We can file a GRAMA request (open records for Utah) for them, and see. But I'm pretty sure it doesn't exist.

My former State Board member said that the first year wasn't adaptive because they needed to have a baseline to create the adaptive capability. My guess is this is why they looked at ranking the questions based on the number of students that got each question right, instead of the content of the question. That seems like a skewed adaptability to me. In theory, the harder questions would have fewer correct answers, but also the poorly written questions would, as well. If it's a 'criterion referenced test' where they determine proficiency based on the types and kinds of questions asked, then the content of the questions should be at issue in determining proficiency, which they apparently weren't, just how many answered correctly.

Oh, this State Board member said it wasn't a public meeting but she and at least the current Board Chair were there in a meeting with AIR. I'll try to get more information. I'd bug the current Board chair, but he's in negotiations for the next 3 days with the Legislature to get \$30M to get us out of the waiver.

I'll ask a couple of other board members. Perhaps one of them would be willing to go 'on the record' for you.

We knew of the letter and in an effort gather the response, Suzette Lopez sent an email to Wendy Hart, Alpine School Board member on March 18th, 2015 asking if she had received a response to her letter. The below is the response.

Subject: RE: Did Parks ever respond to your letter?
Date: Wednesday, March 18, 2015 10:32:04 PM Eastern Daylight Time
From: Wendy Hart <whart2000@Hotmail.com>
To: Suzette <slopez@parrotfishstudio.com>

No. I've decided we're going to do an Open Records (GRAMA) request for the validity and reliability study. I just have to get the info on how to do it from this one guy. I think we can expedite due to needing the info before the big testing window.

----- Original message -----

From: Suzette <slopez@parrotfishstudio.com>
Date: 03/18/2015 7:23 PM (GMT-07:00)
To: Wendy Hart <whart2000@Hotmail.com>

Subject: Did Parks ever respond to your letter?
Did they ever respond to your letter? (the letter attached).
Please advise,
Suzette

There is now an official letter attached from the three board members stating that they never received a response to their inquiries.

Criteria used to "Validate the test" according to the Florida Department of Education. This is in response to an email from Erin Muir, Legislative Assistant, Representative Holly Raschein, Florida House of Representatives, District 120

From: Palazesi, Benjamin [<mailto:Benjamin.Palazesi@fldoe.org>]
Sent: Friday, March 20, 2015 8:10 AM
To: Muir, Erin
Cc: Cooper, Tanya

Subject: RE: Help needed. re. Reliability and validity

Hey Erin,

The Department, in conjunction with its contractors, conducts validity activities throughout each year for all statewide standardized assessments, based on industry standard practices. Some of the activities occur prior to an assessment being given to ensure validity and some occur after the assessment to ensure ongoing validity.

Pre-assessment Validity Activities

Item selection. Prior to our selection of items for the FSA, the items were field tested in Utah to collect data on item validity and bias. Florida assessment staff and content experts at the Florida Test Development Center then selected items from the bank that aligned to Florida Standards and also met our standards for item quality. This review was conducted using the same criteria it has used since 1996, reviewing items for:

Content

- Does the test item measure the standard it is designed to measure?
- Is the item of appropriate difficulty for the grade level and specific content being assessed?
- Is the item worded clearly?

Bias and Sensitivity

- Does the item provide any unfair advantage or disadvantage for any group of students that could affect student responses?
- Is the item acceptable, considering the variety of cultural, regional, philosophical, political, and religious backgrounds throughout Florida?

Test construction. The FSA was constructed from items that meet all specifications, and includes a range of difficulty, so that the test can be scored and can differentiate among students' knowledge and skills (from the Florida Standards for that grade level). Tests must meet all of the criteria as described in the test design summary referenced below.

Scoring. For the writing prompts, review the material used to train test scorers, and oversee the training of the scorers and scoring supervisors. For all other non- multiple-choice items, validate the methodology used to score each individual item

Providing information to the field. The following information is provided to the field so that educators, students and the public gain familiarity with the information to be included on the test:

- Item specifications;
- Training tests (for computer-based assessments) and sample paper-based tests, so that students, educators and the public can see the types of questions that are on the test and the manner in which the questions will be asked;
- Information on policies and procedures, such as use of calculators and reference sheets in mathematics for selected grades;
- The Test Design Summary, which includes, for example, which standards are covered on the test and in what proportion, length of time of the test, and the number of items on the test and their range of expected difficulty; and
- Rubrics used to score the writing portion of the test.

Post-assessment Validity Activities

A full report is provided each December in the Technical Manual, which includes sections that capture pre- and post-assessment validity activities. Statistical studies of validity for each assessment are performed after each administration of that assessment and are summarized annually in the Technical Report. Last year's FCAT 2.0 Technical Report is attached.

Benjamin Palazesi
Governmental Relations
Florida Department of Education
Direct Line (850) 245-0780
Main Office (850) 245-- 0507

Statement from Vince Verges, Assistant Deputy Commissioner, Division of Accountability, Research, and Measurement, Florida Department of Education, on using the Utah test as the basis for FSA test validation

From: Suzette Lopez [mailto:suzettelopez@earthlink.net]
Sent: Friday, March 06, 2015 7:12 PM
To: Verges, Vince
Subject: Assessment questions

Good afternoon,

I have a few questions that I hope you can provide some clarifications on.

Question 1:

In the last Senate meeting, Commissioner Stewart advised the committee that AIR had validated the psychometrics of the FSA test. She also stated that it was validated using Utah students, no Florida students. That being said, that leads us to believe that the Utah SAGE test is the basis for the validation. I was hoping to get a copy of the industry standard/approved validation report that was provided to Florida, providing assurances and the evidence of the validity of the test. Can you please also advise who exactly signed off on that validation and when? Who provided the actual validation of the test (was it a subcontractor for AIR?)

From V.Verges: The SAGE test design itself is different from FSA, and thus is not the basis of validation for FSA. All of the documentation on the assessments is currently publicly available is on the FSA Portal :

<http://www.fsassessments.org/resources/?section=1-°- students-°- and-°- parents>. The Test Item Specifications and Test Design Summaries provide detail on the nature of the assessments. Also on the portal are training tests to familiarize students, parents, and teachers with the presentation of the assessment. These documents and training tests serve as components of the validity evidence. As with every statewide assessment, the annual technical report is prepared after the assessment has been administered and scores have been reported, and the report will serve as the key source of validity and reliability evidence. Documentation, evidence, and data that will be included in the technical report for this year's assessments is gathered throughout the year leading up to and throughout the annual assessment administration, scoring, and reporting cycle. The annual technical report will be available in December of this year.

From: **Verges, Vince** <Vince.Verges@fldoe.org>

Date: Mon, Feb 16, 2015 at 9:30 AM

Subject: RE: Florida Statewide Assessment Technical Manuals

To: Darcey Addo <darceylivesinfl@gmail.com>

EXCERPT FROM EMAIL:

Ms. Addo,

As I mention in several responses below, evidence of validity and reliability, as well as any of the other psychometric properties of the assessment will be compiled along with Florida student performance information and publicly released in the annual technical report later this year, as has been done in the past, and is used to meet the requirements that you cite below. As mentioned previously, other such evidence of content validity is already publicly available on the FSA Portal.

· Commissioner Stewart testified to the Appropriations Subcommittee on Education on February 4, 2015 that the assessment (FSA) “does not require administration within the state in order to determine that it is valid and reliable; and it has been determined that it is valid and reliable.” Can you please provide the data to which Commissioner Stewart is referring in that statement? The statement I quoted from the Commissioner is found approximately 92 minutes into this meeting: http://www.flsenate.gov/media/VideoPlayer?EventID=2443575804_2015021112 [V.V.: The information will be compiled and released in the annual technical report, as has been done each year.]

Thank you. So I am understanding that as of now, there is no validity or reliability information and it will not be available until the baseline assessments are complete. FLDOE will be administering an assessment that has not been validated. [V.V. Your understanding is incorrect. Information we have will be compiled with student performance information in the annual technical report.]

None of the responses in the above emails from FLDOE confirm the that statements of Commissioner Stewart to Senator Simmons that the psychometric validity and reliability of the FSA was confirmed in Utah by AIR. After checking with Senator Simmons office, Dr. Effrem was sent the following document that Senator Simmons was sent in response to his questions in that March 4th Hearing. Instead of providing information about the reliability and validity of what was done in Utah per the commissioner's testimony, they sent a 175 page technical report on the FCAT 2.0 and EOCs dated December of 2014, the cover page and pertinent page 137 of which are attached. It contained a very interesting paragraph that shows the lack of reliability and validity even of the FCAT and EOCs much less the FSA:

Less strong is the empirical evidence for extrapolation and implication. This is due in part to the absence of criterion studies. Because an ideal criterion for the FCAT 2.0 or EOC assessments probably cannot be found, empirical evidence for the extrapolation argument may need to come from several studies showing convergent validity evidence. Further studies are also needed to verify some implication arguments. This is especially true for the inference that the state's accountability program is making a positive impact on student proficiency and school accountability without causing unintended negative consequences.

Dr. Effrem's response is as follows:

Aside from this not being about the same test, based on different standards, using a different developer and from a different state, the above quote completely negates the validity of the FCAT, which had been in place for years, much less the FSA which is still being put together on the fly.

If after the many years that we have had the FCAT 2.0, they STILL don't know if the test is adequately measuring and having a positive effect on student proficiency, that the use in accountability is valid, and that this whole system is not having unintended negative consequences, how in the world can they say anything about the validity of the FSA? There is nothing here to answer the questions about the Utah field test. In fact, the word Utah does not appear in this report!

There is no evidence of psychometric validity from the tests in Utah. This document does nothing to answer Senator Simmons' questions, and if I were him I would be angry that the questions of a sitting senator are being so blatantly disregarded.

The only positive is that the report they did send confirms the grave concerns about the entire testing and accountability system and validity of this test and shows that at a minimum, it should not be used this year to make any high stakes decisions, but truly making life changing decisions with any test is extremely problematic.

Thanks again for all of your and Senator Simmons' concern for Florida's school children and your help to us on this issue.

Associated issues: Here are links to letters from parents that served on the parent validation committee whose words were misconstrued about their findings when reviewing the SAGE test questions in Utah; a video from the head of the teachers union in Utah saying she's getting 100s of complaints from teachers; a video by another teacher shows her students struggling with the SBAC practice tests, not because of the math concepts, but because of the user interface; and finally, a review by a math teacher of the very significant and complex problems with the online interface that will significantly lower the scores just because of difficulties using the program [this teacher discusses the SBAC test interface, but it is the same as Utah's SAGE and the FSA developed by AIR]:

<http://www.utahnsagainstcommoncore.com/first-parent-member-of-utah-sage-test-review-committee-speaks-out/>

<http://www.utahnsagainstcommoncore.com/second-parent-member-of-utah-sage-test-review-committee-speaks-out/>

<http://www.utahnsagainstcommoncore.com/sage-parent-reviewer-shares-concerns/>

<http://www.utahnsagainstcommoncore.com/uea-president-on-problems-with-sage/>

<http://mathedconsulting.com/wp-content/uploads/2015/03/Common-Core-Tests-Fatally-Flawed.pdf>