

<p>TAC Fall 2017 Meeting Notes</p>	<p>Meeting Date: November 9, 10, 2017</p>
<p>Participants: TAC: Jamal Abedi (Absent), Lyle Bachman, Aki Kamata, Tim Kurtz, Carol Myford, WIDA: Kei Bishop, Mark Chapman, Gary Cook, Carsten Wilmes CAL: Keira Ballentyne, Dorry Kenyon, Shu Jing Yen DRC: David Chayer SEA: Andy Middlestead (MI)—call in, Vince Verges (FL)</p>	<p>Location: Double Tree by Hilton Albuquerque, NM</p>

1. Assessment Use Argument

Dr. Bachman (TAC member) and WIDA assessment staff (Drs. Kim & MacGregor— call ins) presented on the AUA. Dr. Bachman gave a brief history of the AUA and it's relationship to the currently held validity theories and approaches. TAC members discussed sections of the AUA.

Suggestions:

- Think about creating white (policy) papers on the negative consequence of using ACCESS score information inappropriately.
- Think about adding WIDA's values (e.g., in brief paragraph form) to Claim 2.

Recommended to dos:

- Take out the backing and explanatory notes from the AUA. Provide it back to the TAC for review. Upon final edits, create a "Formal Discussion Draft." Then take it to the WIDA's assessment subcommittee
- Take the explanatory notes and create a consolidated (and edited) table with responsibilities (i.e., WIDA's, a State's, or joint) for each issue or note.
- Create a document outlining the intended and unintended uses of ACCESS 2.0 (including the annual summative, the screener, and Alt ACCESS assessments).
- Once English Language Proficiency Assessment Peer Review Guidance is published, create Peer Review critical element to AUA crosswalk.
- Bring back to the next TAC meeting a recommendation on how the AUA and ATR relate to each other.

2. Year-in-Review Report

Dr. David MacGregor gave a presentation on a draft Year-in-Review report. There were several comments about the formatting of the document.

Recommended edits and to dos:

- Think about changing the heading structure to be more FAQ-like
- Revise the Test Content Alignment Overview to make the relationship between online and paper clearer
- The title and description of Table 2 should be "Number of Students Tested." Also, put an explanatory note on the numbers provided indicating that this was the information reported to WIDA for the 2016-17 administration.
- Possibly add ethnicity male/female members as well. Can Table 2 be simplified?
- Briefly explain/define the attemptedness criteria in that section.
- Add in the G1WA the affect on reclassification, i.e., only 7 students potentially affected out of 40K.
- Pyschometrics brief section should be "Psychometrics in Brief."
- Describe the document's audience in the report itself.
- Add "Reporting Timeline" to Figure 1.

3. ACCESS 2.0 Research Agenda

Writing Plus Point Study

Dr. Dorry Kenyon (CAL) presented on the study. There was a rich discussion on this topic. Some reservation was expressed by TAC members over the use of the “+” designation and the agreement criteria being +/- 1 score point. The chief issues discussed were 1) does the scale now mean something different with the new detailed descriptions?, and 2) is exact agreement actually exact? There was also discussion about move from the “+” system to a purely number based system.

Ideas for Analysis

- Look at the frequency of +'s on series 401 vs. 402.
- Look at the fit statistics for the score points at the + levels on 402 to see if they get better than those on 401

Recommended To dos:

- Explore the intervals between "exact agreement" and agreement between the pluses. Are they similar?
- Explore the intervals between 2+, 3, 3+ vs 4, 5, and 6.
- Look at a cross-tabs of 2+, 3, and 3+ from 1st to 2nd rater to determine if +s are being used.
- WIDA and CAL need to meet and discuss the next steps on migrating to a writing scoring system that is numeric (0-9). As part of that conversation, a description of how the change in scoring would effect the relationship to the screener, given that screener scoring doesn't change.

Less-than-four-domain Composite Analyses

Dr. H. Gary Cook (WIDA) presented on the less-than-four domain composite analyses and asked the TAC for their recommendation on next steps.

Ideas for Analysis

- Run a propensity matching type of analyses with missing data, in this case domains scores would missing. Populate the missing data with the matched information and create a composite, which would be on the same scale as the original composite. This way a new scale would not have to be created.

Recommended To dos:

- Write a white paper showing several less-than-four-domain composite analysis options. In that paper add the benefits and drawbacks for each presented model.

Construct validity study

Dr. H. Gary Cook (WIDA) shared a PowerPoint on the project. The TAC has several questions about the model, e.g., Did we look at a bi-factor model? (Yes, we did). The TAC made several suggestions about running other analyses, which are listed below.

Ideas for Structural Analyses

- Run multilevel model with grade level and domain/standard level with all tier items together (incorporate missing by design data)
- Consider looking at a bi-factor model for SR item types
- Randomly subdivide the data into two sets: run a model for each subset and compare one to the other to validate the model
- Run models that mirror what you report (e.g., RLSW)

- Run a model fixing loadings to be equal because we are using Rasch model

Future research projects

There was a rich discussion about future projects. Here are ideas expressed by the TAC regarding possible research topics:

- Dr. Bachman (TAC) suggested looking at research on educating stakeholders on assessment literacy. What can be done to more effectively engage state stakeholders in understanding test use, decision-making, and understanding?
- Dr. Kamata (TAC) suggested creating a methodological series that looks at a variety of issues, e.g., the creation composites.
- Dr. Myford (TAC) urged going through the AUA and finding holes in the claims and warrants. She raised several questions. Do families understand the decisions (Claim 2) and consequences (Claim 1) of the assessment scores? How are multiple-sources of evidence enacted by states? How are multiple-source decisions made? What evidence do we have that we're adequately sampling the domain by test items or types of tasks? Another thought would be to work at clearly defining the domains (domain analysis) and see if the tasks provided are a sufficient sample.
- Mr. Kurtz (TAC) asked, "How are students, once exited, re-identified as ELs?" He also expressed interest in understanding how non-ELs would perform on ACCESS. Mr. Kurtz's concern is about false positive students rather than false negatives.

All TAC members agreed that the next step should be to look at the AUA and identify areas lacking evidence.

Recommended To dos:

- Bring to the next TAC meeting (April 2018) the "evidence" holes in the AUA. This is first step. A follow up step would be to create a research agenda from AUA.

4. ACCESS 2.0 Activities

WIDA Screener Technical Brief

Dr. Kei Bishop (WIDA) presented preliminary analyses on the ACCESS 2.0 screener results. Note that the data used to do the analysis was NOT a census dataset. The TAC provided several suggestions about things to analyze or add to the technical brief.

Recommended To dos:

- Add a section describing in the report identifying that reported scores attempt to minimize the false positive outcome (maybe in the Scaling section).
- Creating an interpretive guide for the screener.
- Examine the possibility of doing test-retest reliability studies.
- Look at local educators' perception of how screener scores adequately identify ELs and whether they're performing well or not in classrooms. Also, explore what is used to make determinations that a student is or isn't an EL.

ACCESS 2.0, series 402 (i.e., SY 2017-2018) operational plans

Dr. Shu Jing Yen (CAL) shared briefly on Series 402 activities. There was rich discussion, and the following information was requested.

Recommended To dos:

- WIDA should follow up with CAL on the potential of doing post-equating to examine 401/402 RL adaptive

engine drift.

- A more detailed (documented) criteria for RLSW item selection, including both qualitative and quantitative procedures should be provided.
- Mr. Kurtz requested documentation outlining the characteristics that determine how CAL decides whether a writing item moves forward qualitatively to field testing. The report back to the TAC could be briefly describing the activities and processes used to move an item forward (possibly a slide in a PPT).
- WIDA should check with DRC on how operational/FT writing read behinds and validity sets are created and applied. When CAL/WIDA get aggregate information about SW scoring, what does that information represent?
- Documentation outlining characteristics of appropriate writing responses, if none exist could such a document be created?

Future enhancements

Dr. Shu Jing Yen (CAL) shared briefly on upcoming enhancements, specifically to field testing listening and reading items/folders. The TAC had several questions. Are there FT position effects? Are there folder sequence effects? Are subsequent years' folder parameter estimates affected by FT folder placements? The TAC asked that this topic be readdressed at the next meeting.

Recommended To dos:

- Bring the field test folder design and potential folder sequencing or location effects up at the April 2018 meeting.

Progress on embedding Writing Field Test design

Dr. Yen (CAL) shared briefly on the the embedded writing field test design. She discussed the possibility of creating a new “alternate” writing scale and shared some analyses (with current data). Dr. Yen stated that she will be will be doing comparative analyses between the current and “alternate” writing scales. She asked the TAC, given the embedded field test design, would a new writing scale be required our could the current scale be continued? The TAC discussed this issue and expressed concern about the classification differences in writing, literacy and overall composite scores. Their preference was to maintain the current writing scale and not create a new on. There was also discussion about

Recommended To dos:

- Move forward with the current writing scale. Check the classification characteristics between the Original and Alternate scales options for the writing domain and associated composites.

5. Paper and Online Differences

Drs. Bishop and Cook (WIDA) asked the TAC how WIDA should move forward with ACCESS 2.0 mode differences. After discussion, the TAC suggested that WIDA move forward with an equipercentile approach to adjust for mode differences. Dr. Kamata (TAC) suggested that WIDA conduct this approach across years. Dr. Yen (CAL) suggested using a Bayesian approach to selecting the populations. Once the population has been defined, check to see if the population sample is maintained. Dr. Bishop (WIDA) shared about a speaking study that will look at rater scoring differences on speaking tests. The TAC provided several recommendations to that study.

Recommended To dos:

- Continue with equipercentile mode adjustments between ACCESS 2.0 paper and online sores. Explore

the cross year and population effects in created concordance tables when data become available.

- For the speaking study, look at a counterbalanced design with video & audio recordings.
- TAC suggested looking at the descriptive statistics of schools that have large differences in Speaking between paper and online

6. The Future of ACCESS 2.0 Paper

Dr. Mark Chapman (WIDA) provided an update on the future direction of the paper version of ACCESS 2.0, specifically the consolidation of tier B/C forms and the creation on additional paper forms for grade clusters 6-8 and 9-12.

Next Meeting: April 17-18, 2018; Location: New York, NY