

SEA Meeting Notes

Name of Meeting	Psychometrics Subcommittee Meeting
Date	9/21/18
Time	1:00pm CT
Attendees	SEAs: Ana Karantonis (RI), Vince Verges (FL), Jasmine Carey (CO), Rebecca Izzo Manymules (BIE) WIDA: Gary Cook, Kei Bishop, David MacGregor
Note Taker	Jessica Davis

Topic # 1: Introductions
Information Shared: <ul style="list-style-type: none"> • SEA Lead: Andy Middlestead (MI) • SEA Members: Ana Karantonis (RI) Ann-Michelle Neal (UT) Brandon Loudermilk (SC) Eric Meredith (MT) Jasmine Carey (CO) Jonathan Rollins (GA) Katherine Edwards (MN) Barry Pederson (IL) Salih Binici (FL) Rebecca L. Izzo-Manymules (BIE) • WIDA Content Leads: David MacGregor, Kei Bishop, Gary Cook • Client Relations Lead: Kristen Burton
Topic # 2: TAC Meeting
Information Shared: <ul style="list-style-type: none"> • Draft Agenda <ul style="list-style-type: none"> ○ Welcome new members- Claudia Flowers and Greg Cizek ○ ATR redesign ○ Peer review ○ Speaking score differences 2017 and 2018 ○ The writing shadow scale ○ Spring verification analysis and embedded writing FT calibration ○ Year in Review Report ○ Research reports <ul style="list-style-type: none"> ▪ Speaking study report (paper and online rater study), ALT composite report study, Accessibility & embedded support research, Planned Reports, Writing study, Accommodation study • Believe we have signed contracts for both Claudia and Greg. Both will be joining the fall TAC meeting remotely.
Discussion Questions Does anyone have any concerns about TAC agenda? <ul style="list-style-type: none"> • FL-No concerns. Does subcommittee want to be in review scheme for research reports? <ul style="list-style-type: none"> • Gary-May not be able to provide reports to subcommittee before TAC meeting. • CO-Yes, want to review. • FL-Yes, before the TAC would be nice but realize that's not always possible. • TAC Meeting is November 8-9 in Orlando, FL.
Topic # 3: Peer Review
Information Shared: <ul style="list-style-type: none"> • Updated Timeline

- Late Summer/Fall 2018-WIDA is working on drafting narratives and providing evidence now.
- By end of November-Have all draft narratives and sample evidence ready to share with subcommittees. Subcommittees will have 1 month to review and provide feedback.
 - Standards-1
 - ADI-2
 - Psychometrics-3
 - AAE-5
 - ADI-6
- TAC and mock reviewers will also review and provide feedback. Looking to contract with peer reviewers.
- February 2019-All dashed sections of hybrid evidence.
- March 2019-Deadline for consortia-related submission. WIDA will consolidate, edit, organize, and submit narratives and evidence and associated evidence related to consortium-related critical elements. Was hoping for more time, tried to negotiate, but this deadline was given.
- Peer Review Webinar Schedule:
 - Wednesday October 3, 2:00-3:30 CT
 - Wednesday November 28, 2:00-3:30 CT
 - Wednesday February 13, 2:00-3:30 CT
 - Wednesday May 22, 2:00-3:30 CT
- Department of Education (Don and Deborah) has agreed to participate on webinars.

Discussion Questions

- FL-Fantastic, thorough plan. A lot of pieces there. CAL was involved, too? Yes.
- Gary-We have gone through all subsections and identified evidence that is needed. Spreadsheet identifies who is responsible for each section.
- Gary-There are areas that we will not meet. Section 3.1-we will not meet the alignment requirement because the last alignment study was done in 2011. We are in the works for committing WCEPS to conduct an alignment study. We will have a contract in place. Also, we have not done an alignment study for Alt. ACCESS, so we will do that, too. We will provide what we need to let peers know that we understand we do not meet, but explain what we plan to do in the future.
- Gary-Section 2.5-2.6-We do not currently do formal data forensics analysis. Will bring this up at the next TAC Meeting (Not in the fall). Discuss what our plan is for the future. We will not meet this requirement. Recognize that we need to do some work.

Topic # 4: ATR Redesign

Information Shared:

- Preliminary Schedule for Updating WIDA's Annual Technical Reports for ACCESS Paper & Online and Alt ACCESS
 - 2018-19-Planning for reorganization
 - 2019-20-Implement Part 1: Keep the front matter of current ATR, add Annual Tech Supplement topics
 - 2020-2021-Implement Part 2: Drop current ATR front matter, connect Annual Tech Supplement topics to AUA document
- ACCESS Technical Supplement Example Table of Contents
 - Brief description of the purpose and design of the test

- Test Design & development
- Item development
- Technical quality
- Inclusion
- Standards and reporting
- Designed in a way to support peer review material.
- A more detailed table of contents is available, but will likely change after bringing to the TAC.

Discussion Questions

- CO-Since online and paper splits into 2 reports, will there be anything in the ATR about comparability between online and paper?
- Gary-A separate, comparability report will be available each year.
- Gary-Instead of redoing equipercentile linking every year, monitor every other year. Building a bigger pool of students to do equipercentile linking. Linking constants are not changing that much. Right now, we have to set up different tables each year for states that choose to adjust paper based on online performance. Would like not to do this every year.
- CO-Understand, would like not to need these every year.

Topic # 5: Speaking Score Differences Between 2017 and 2018

Information Shared:

- Gary-A 5-6 scale score difference brings me concern.
- To attain same proficiency, student would have to score higher on this year's assessment than last year. This is a concern.
- We have done a root cause analysis (double, double checked!) and determined that:
 - the speaking score tables were correctly transferred to DRC;
 - DRC correctly applied the speaking scoring tables;
 - we found no anomalies in speaking score equating;
 - apart from the new speaking items, rater training materials did not change; and
 - DRC's raters' interrater reliability is actually better than last year.
- Follow up actions:
 - We are looking at our rater training materials to make sure they are appropriate. May want to add more borderline cases.
 - We are also looking deeper into our sampling methods for scoring speaking to get a better understanding of the variability associated with that.
 - We are bringing our findings to the TAC in November to get their advice.
- Speaking field test is embedded in ACCESS. Wait until have a sufficient sample of Speaking performance for field test analysis and item selection for the next year. In the following year, we conduct a verification study to see what kind of displacement, if any, there is.
- Large sample in verification analysis-no concern.
- Moderate sample in verification analysis -concern
- No sample in verification analysis-concern
- Have to nail down the Speaking test samples for scoring and verification study by the end of January.
- As states move back/forward reporting deadline, sample is impacted.
- CO-Is analysis done on verification sample from a demographic standpoint?
- Gary-We try to match demographics as much as we can, but we are constrained by the timeline for reporting.
- CO-If early states are not representative of Consortium, there is not much you can do.
- Gary-Right, this is what I believe is happening.

- Gary-Concerns about pre-equating. If we pre-equate, the sample you are pre-equating on is from a previous year. If the distribution of students is different, you are imposing a distribution of students on the following year.
- “Rubber band only stretches so far, and then it snaps.” Speaking is the domain test that we see this really pronounces.
- FL-What’s been the practical impact of this?
- Gary-Overall scale score difference across the Consortium was 2-3 scale score points. Has not really effected distribution of Overall Composite scores. This has generated political problems in a few NE states.
- Gary-Speaking is a construct that is hard to capture well. Have to make sure the scoring, tables, and technical problems are right.
- Gary-It is possible that this fluctuation is normal, but this is not something I have seen, which is why I am bringing this concern to the TAC.
- RI-Is the problem that it is not representative?
- Gary-Two-fold theory. First, we need better examples of Speaking performance levels. Second, early states may not have same distribution of Speaking performances as other states.
- RI-Do early states change from year to year.
- Gary-Generally, no.
- Gary-If state chooses to change its reporting window, the sample is impacted. Not sure how to solve this other than pre-equating.
- CO-Make everyone test earlier.

Topic # 6: Instituting a psychometrics QC review meeting at CAL

Information Shared:

- At the last subcommittee meeting, announced that the normal CAL QC meeting was going to remove the psychometric component of QC.
- We have talked with HumRRO to gage their interest. HumRRO said they would be interested.
- We have talked with the Assessment Solutions Group (have a contract). Contract with ASG is looking at CAL item development, DRC technology, and WIDA’s processes in test development and score reporting.
- We are planning on writing up psychometric QC specifications for groups to bid on in an RFI. If any subcommittee members have something like this drafted, please share with Gary.
- Plan to follow-up by creating a scope of work and RFI. Potentially engage in a contract.
- FL-Can provide copies to Gary.
- Gary-Will have to go through contract services since cost will be over \$6,000. Hope to have this done by sometime next year.

Topic # 7: Cluster-based Scoring Tables

Information Shared:

- We talked about the idea of having 1 scoring table per cluster.
- In the past, we adjusted the LOSS according to the grade.
- Creating 1 scoring table for DRC to consume for online adaptive engine.
- Would like to simplify scoring table.
- TAC did not object or have suggestions.
- Brought idea to DRC/CAL and discussed with subcommittee in June.
- Decided to implement cluster-based scoring table, fixing the LOSS at the highest grade level so no one falls back due to LOSS change. Everyone will remain the same or gain.

- This was implemented for both paper and online.
- From 403 administration, you will see scoring table will be 1 per cluster.
- Be aware of this when you run descriptives next year.
- This is talking about students at 1 or 1.1.
- So no impact to other cut points? Correct, just LOSS.
- Gary-In most cases, this primarily applies to the LOSS. Out of 2 million cases, this may have some impact, but doubtful there will be substantial observations.

Topic # 8: Retention of Speaking and Writing Responses

Information Shared:

- All Speaking and Writing responses since 2015 have been retained and not yet removed.
- In the second year, Speaking response are migrated to a slower server.
- We would like to delete Speaking responses from DRC's system in the third year, unless specifically requested by a state.
- No one has requested a response after the administration year.
- We do use Speaking and Writing responses for training.
- These responses take up a lot of space. We will need to either start charging to store responses or delete them.

Discussion Questions

- Is this a sufficient approach for SEAs?
 - FL-We rarely have to retrieve responses for content assessment.
 - Gary-Terabytes of data. This is expensive.
 - CO-Believe there is something in contract about deleting data. Not sure if Speaking/Writing responses count as PII.
 - Gary-Treats Speaking/Writing as PII.
 - CO-In contract, should be deleted after 45 days or after this is no longer needed for data analysis.
 - Gary-In my mind this would be 2 years.

Action Items/Recommendations:

- Write up a proposed retention policy and send to subcommittee for thumbs-up/down.
 - Target Date: TBD
 - Responsible Party: Gary Cook