

***Teaching and Learning Branch
Accountability Resources Workgroup***

**DCAS Technical Advisory Committee Meeting
Sheraton Suites, Wilmington, Delaware
December 13 & 14, 2012**

Meeting Summary Report

Day One: December 13, 2012 (Thursday)

In attendance:

DTAC Members:

Dr. Robert Linn	University of Colorado (Chair)
Dr. Ronald Hambleton	University of Massachusetts
Dr. Suzanne Lane	University of Pittsburgh
Dr. Richard Luecht	University of North Carolina
Dr. Scott Marion	National Center for Improvement of Education Assessment
Dr. Martha Thurlow	University of Minnesota

American Institutes for Research Staff:

Dr. Harold Doran	Ms. Christine Loew
Mr. Tom Glorfield	Dr. Gary Phillips
Mr. Kevin Murphy	Dr. Joshua Smith
Dr. Caroline Lang	Dr. Dipendra Subedi
Dr. Ming Lei	

DDOE Staff:

Mr. Brian Touchette (Acting Director)
Dr. Katia Foret (Mathematics)
Ms. April McCrae (Science)
Dr. Qi Tao (Technology)
Ms. Denise Weiner (English Language Arts)
Dr. Liru Zhang (Psychometrician)

Other Participants:

Dr. J.P. Beaudoin Research in Action

Introduction

1. *Welcome and Introductions*

Mr. Brian Touchette, Acting Director of the Accountability Resources Workgroup, welcomed the group and introductions were made. TAC members were advised of the recent retirement of Dr. Linda Rogers and Dr. Michael Stetter.

2. *Agenda and Review Materials, Meeting Notes of the July 2012 DTAC Meeting*

Dr. Scott Marion motioned to accept the summary and recommendations from the July 12–13, 2012, meeting; Dr. Suzanne Lane seconded this. Members had previously been provided with electronic media containing the meeting review materials.

3. *Delaware Educational Policy Update (Mr. Brian Touchette as proxy for Ms. Mary Ann Mieczkowski)*

a. DDOE Administrative Update

The DDOE received approval from all interested stakeholders for amendments to the teacher evaluation plan. The original AYP system will be adhered to for the moment but is being reevaluated as confusion could result if two different systems are utilized.

The original plan for *end-of-course (EOC) assessments* was that they would replace ninth and tenth grade assessments. However, the Secretary has decided that EOC assessments should be optional. This decision was taken because of various complications—for example, accountability ratings if a student took this examination while in middle school; students transferring from a charter to a public school. The mandatory requirements remain as US History and Algebra II or Integrated Math III.

The DDOE recently contracted with Caveon™ to evaluate test security. Strengths were noted, but improvements were recommended for both DDOE and district procedures. DDOE recently implemented mandatory training for all test administrators of the DCAS and the DCAS-Alt1. The training is given online through the department's secure Integrated Management System (IMS). The system does not permit participants to receive their security certification until they have completed the course. DDOE is working toward further security refinements with Caveon.

Discussion Summary

One of Dr. Hambleton's concerns around the adaptive testing program is the exposure of items for test security and the fear that someone could copy and share test items because of high-stakes uses, such as teacher evaluation. He wondered if Caveon had made an effort to look at statistics across districts to ascertain if certain schools or districts performed better on certain items. Members were reminded of AIR's presentation at the previous TAC regarding their test security software, which is designed to flag what may be suspicious occurrences. DDOE has now commissioned a security manual from Caveon and will further examine security protocol based upon any recommendations. In addition, AIR noted that they produce a benchmark report for each school that determines if a district performs better/similar/worse than what is conditionally expected within that benchmark.

4. *Delaware Assessments Updates (Mr. Brian Touchette)*

a. 2012-2013 DCAS Administration Schedule

The DCAS fall assessment window was successfully completed. This first test in the current school year is considered the starting point for students' to show fall to spring growth and additionally counts toward the new teacher accountability system. Based upon feedback from the field, the DDOE revised the student-testing schedule for this school year. Students will have two test windows (a fall window with one test opportunity and a spring window with two test opportunities). The test windows were expanded to accommodate end-of-course assessments and block scheduling.

The alternate assessment, DCAS-Alt1, testing was also successfully completed.

The State Board of Education approved the DCAS-Alt1 confirmatory performance standards (cut scores) for reading and mathematics and the cut scores for science and social studies.

b. 2012-2013 PSAT, SAT, and NAEP

To promote NAEP participation, the DDOE recently launched a poster contest aimed at students. In addition, fourth- and eighth-grade teachers will receive a calendar containing previously released NAEP items. It was hoped that these measures might encourage the use of NAEP items. The DDOE had also spent time to educate various stakeholders (e.g., State Board of Education, statewide directors of assessment, curriculum, special education, etc.) on comparison items from NAEP, SAT, DCAS, and Smarter Balanced Assessment Consortium (SBAC) with items similar in concept but with different expectations for rigor.

c. Smarter Balanced Assessment Consortium (SBAC)

Delaware has been pleased with the interaction with SBAC. They have been responsive to our comments and concerns. However, DDOE is dealing with potential policy implications due to differences with performance levels, interim assessments, etc.

The Secretary wants an assessment that will give schools feedback on students' knowledge and performance on the common core standards. The SBAC pilot does not offer test scores and results for students. Delaware is considering the administration of a separate test that will offer a scoring mechanism.

Discussion Summary

Comments were made around the fact that since NAEP is a low-stakes test and uses samples of students for state-level results. It should not be seen as surprising that the number of participants is generally low. However, TAC members were advised that Secretary Murphy's expectation is that all testing mediums should connect toward college readiness. The DDOE has used Race to the Top funding to allow for mandatory PSAT for ninth grade in the fall and SAT for eleventh grade students in the spring. Delaware ranks in the middle nationally for SAT results and has pushed hard for a high inclusion of students. It was conceded that as states adopted the Common Core State Standards and the state assessment should measure the CCSS, it would be interesting to see if and how NAEP will be adjusted. NAEP is a great resource in educational areas and for schools.

Dr. Hambleton questioned Dr. Zhang on psychometric studies planned for the SBAC test. She confirmed that the items be linked to a vertical scale and would provide more information at the next meeting.

2012-2013 DCAS Summary – Fall Administration (Dr. Liru Zhang)

Discussion and Suggestions

Dr. Zhang provided an overview of fall DCAS results, including details of quality control. The subject of smoothing was discussed due to the lower performance of students in lower grades than their peers in higher grades. It was agreed that any smoothing during standard setting was rhetorical rather than statistical. Members of the TAC debated what consensus and/or conclusions should or should not be made around the percentage of students in the grades who exceeded or did not meet expectations. This was particularly notable in grades 3 and 6. Dr. Doran indicated that AIR's system looked at very large variances but was not geared to

something he thought of as minimal in nature such as this was. Dr. Zhang suggested that the data here did not support the assumption that students' scores should increase in the spring test administration, because they would have received more instruction on the appropriate curriculum. Dr. Marion indicated that the subject of instructional sensitivity should be studied, although Dr. Lane doubted that this could be investigated in any detailed level. The big question is how the assessment is sensitive to instruction. It was agreed, however, that more analyses should be done on the dynamics of the population taking the test, such as the fall-spring growth and matched student performance to the previous grade. Dr. Zhang will present at the next meeting.

2011-2012 DCAS Technical Report (Dr. Harold Doran)

Dr. Doran provided a general overview of the 2011-2012 Technical Report but advised that final approval had yet to be given by DDOE.

Dr. Doran's PowerPoint submitted that the test items are well matched to the blueprint, although there is some evidence that the item bank requires more items to target students with higher levels of achievement. Dr. Doran's analyses indicated that item exposure rates are low. Dr. Linn asked why the marginal reliability reported on slide nine was relevant. An in-depth conversation followed on evidence of validity, reliability, fairness, comparability, and for adaptive testing of the report.

Discussion Summary – Technical Report

In general, technical reports are the subject of many TAC meetings. In order for the TAC to provide useful feedback, validity arguments should be made with respect to the claims of the purposes and intended uses of test scores. Technical reports should start with a cadence framework that should outline what claims and commonalities are being made. One example of this would be teacher evaluation and what evidence to support the use of test scores for this. In this case, any audience should be able to follow the argument being made.

Dr. Hambleton appreciated the simulation results within the technical report, but it was challenging for him to find the relevant information that he had been tasked to report on. He requested that materials pertaining to the same subject should be collated in a more orderly and efficient manner. In addition, Dr. Hambleton was critical of the lack of references and citations within the report, together with the lack of argument to support a claim. A frame of reference should be provided.

Mr. Touchette advised the TAC that DDOE and AIR would be meeting to talk about the TAC's recommendations and determine appropriate resolution(s). If there are suggestions that cannot be implemented, the TAC will be advised accordingly.

Improvement of Technical Quality for DCAS

Part I: Machine-Scored Constructed Response (MSCR) Items (DDOE and AIR)

Background was provided by Dr. Zhang, which included a PowerPoint presentation.

Procedures for Item Development, Review, and Rubric Validation (Ms. Denise Weiner)

Dr. Hambleton commented positively on the idea of the statistics of MSCR items and would like it shown graphically to better depict the step values. Axes should be standardized. Dr. Doran explained that these were snapshots of AIR's master data sheets, which are used for data

reviews. Dr. Phillips mentioned that these were also used for archiving and for identification of problems.

Dr. Lane asked for an explanation on a situation where a student does well on the test but poorly on a specific item in the rubric validation process. The TAC raised the question of validity MSCR items. It should be clear what the MSCR item intended to measure because it costs more than a MC item and is very memorable. In terms of reversal step-value, the TAC indicated that each score point must be meaningful and a higher score point requires additional knowledge and higher cognitive demand than the lower score point. These issues should be considered in item development, item review process, and rubric validation.

Part II: Item/Passage Exposure Rate in Reading

Using 2011-2012 DCAS eighth grade reading as an example, DDOE and AIR conducted analyses of item/passage exposure. The background, results of analysis by test window, the two opportunities in the spring, and the potential reasons for over- and under-exposure were presented by Dr. Zhang, Dr. Subedi, and Ms. Weiner. A passage was considered over-exposed if the passage was seen by more than 40% of the students within a test window. A passage was considered under-exposed if the passage was seen by less than 5% of the students consistently across test windows. The results of the analysis suggest that 24% of the passages were over-exposed in 2011 fall, 13% in 2012 winter, and 22% in 2012 spring. About one-third of the passages were under-exposed consistently across test windows.

From content perspective, the range of performance indicators at item level could trigger the under-exposed passages in comparing the range of performance indicators for over-exposed passages. This analysis was limited and thus the findings are inconclusive; more analyses and observations need to be done in the near future.

Over-exposure is a serious issue in adaptive testing, especially when test results are used for high-stakes purposes. With limited time, money, and a small student population in Delaware, it is not feasible to significantly enlarge the reading item pool in a short period of time. The priority is how to efficiently utilize the current pool. The critical issue in efficient utilization of the reading pool is to identify the possible causes for the under-exposed items/passages in order to:

- Inform test development and item coding
- Reduce over-exposed items/passages
- Improve test security
- Derive valid and reliable test scores for high-stakes uses

Demonstration of Sample Items – Science – Ms. April McCrae

Ms. McCrae gave an overview of sample science items. She explained the evolution of the multiple choice and the challenge to incorporate items into this concept.

Demonstration of Sample Items – Mathematics – Dr. Katia Foret

Dr. Marion requested detail on the information a student would see when a test item included a grid. Discussion ensued around how the decision was made on what information should be provided.

Day One concluded at 5:20 p.m.

Day Two – December 14, 2012 (Friday)

In attendance:

DTAC Members:

Dr. Robert Linn	University of Colorado (Chair)
Dr. Ronald Hambleton	University of Massachusetts
Dr. Bree Jimenez	University of North Carolina at Charlotte
Dr. Suzanne Lane	University of Pittsburgh
Dr. Richard Luecht	University of North Carolina
Dr. Scott Marion	National Center for Improvement of Education Assessment
Dr. Martha Thurlow	University of Minnesota

American Institutes for Research Staff:

Dr. Harold Doran	Ms. Christine Loew
Mr. Tom Glorfield	Dr. Gary Phillips
Dr. Caroline Lang	Dr. Joshua Smith
Dr. Ming Lei	Dr. Dipendra Subedi

DDOE Staff:

Mr. Brian Touchette (Acting Director)
Dr. Liru Zhang (Psychometrician)

Other Participants:

Dr. John Smithson

General

July 22–23, 2013, were the agreed upon dates for the next meeting. At that time, the TAC will accept whatever materials and results are available.

Delaware SAT and PSAT Results (Dr. Liru Zhang)

1. Background
2. 2012 Summary
3. Comparisons: across-years and matched students

Dr. Zhang requested that the data presented not be shared with anyone outside of this meeting.

She explained that grade nine students take the PSAT, and that it is Delaware's requirement that all grade 11 students take the SAT. This was made possible by the use of Delaware's Race to the top funding. It was confirmed that, although DDOE has set target results, these were not used for accountability purposes.

Discussion Summary

Dr. Lane asked if DDOE had looked at any changes in Delaware's college entrance patterns because of the SAT taking. Mr. Touchette advised that this was only the second year that this had taken place, but the department does look at surveys that the students complete if they are looking at colleges. This Student Success Plan is one indicator of a student's readiness for higher education. Dr. Lane concurred that it was important to track this information.

DCAS End-of-Course (EOC) Assessments

Summary Results for EOC-1 (Dr. Liru Zhang)

Dr. Zhang explained to the TAC currently the different courses required could be considered under the same DCAS end-of-course exams. The analyses results were summarized by course and by grade for each EOC-1 test.

Discussion Summary

For a more true comparison, the TAC advised that the information presented should only have included results for ninth graders. It is inappropriate to compare the results on EOC with the results on DCAS summative assessments.

Standard Setting for EOC-1 (Dr. Gary Phillips)

Dr. Phillips presented an overview of the standard setting process that was completed after the compilation of the 2012 operational data. Performance standards were recommended in Algebra I, Biology, English II, Integrated Mathematics I, and U.S. History. The demographics of the panelists are described in the technical manual. Participants were given impact data at the outset, including DCAS, PSAT, and PISA items. In order to compare the new EOC performance standards to the DCAS standards, AIR statistically linked the EOC to the DCAS adaptive tests to allow the panelists to see if these EOC standards were higher or lower than DCAS standards. The panelists were encouraged not to set the EOC standards lower than the DCAS standards, and in general, the panelists established rigorous performance standards. The State Board of Education approved the performance standards recommended by the panelists.

Discussion Summary

Dr. Marion suggested that it would have been interesting for each group of panelists to receive a different set of impact data after the event just to see how the results may have differed. Dr. Hambleton advised that DDOE has a responsibility to properly educate the State Board on these processes and more graphics to depict the results would have been useful.

Results of Alignment Study for EOC-2 (Dr. John Smithson)

Dr. Smithson joined the meeting by telephone to discuss the results of the alignment study. Algebra II and Integrated Mathematics III are the content areas most aligned to the Common Core State Standards. He explained the methodology used for this two-stage alignment study and what was accomplished at each stage. Dr. Smithson discussed the team process that comprised teachers and university personnel. Participants were selected due to his or her content analysis knowledge and their understanding of the taxonomy process. Dr. Smithson gave details of the alignment indices and how points were awarded.

Discussion Summary

Dr. Marian requested further clarification on items that go across different content descriptors, including how the targets/lists of content descriptors are populated. The TAC raised questions and concerns with regard to how the criteria were set for alignment and the threshold was determined for evaluation. The TAC advised that rationales and evidence must be addressed and included in the report. A revised version of the report would be available for review.

Consequential Validity Discussion (Dr. Suzanne Lane)

Dr. Lane advised on what key items should be included when a validity interpretative argument is to be used. In education, it was hoped that the assessment and accountability programs would benefit the actual instruction. Therefore, Delaware should decide upon what are the most important claims it wants to make with regard to student performance. One example would be to think about each of the testing windows, and to take the information gained from each and use it in an appropriate manner to determine appropriate instruction to meet students' needs. Mr. Touchette conceded that, although the scale scores gave teachers little information, they were able to access score reports and dig into the strands to ascertain student strengths. Modules had also been developed and built into the system, and from this, the DDOE could see if a teacher was actually accessing the information provided. DDOE also contracted with Compass Learning to provide individualized learning programs for high school students.

Discussion Summary

The TAC cautioned that although there was a large amount of available data, it should be remembered that not all are experts in the use and interpretation of test scores. Mr. Touchette agreed and advised that the DDOE's data coaches were working with schools' professional learning communities to teach how to use and apply data they receive. The state can argue the need for consequential validity, but it needs to include goals and claims of their assessment, and the approach also needs to be in agreement with policy.

DCAS Reporting System – Features and Functions (Dr. Qi Tao)

1. Background of DDOE Online Report System
2. DCAS – OR (public access online reporting)
3. Delaware Student Assessment Reporting & Analysis (DSARA)

As a follow up to an item at the last TAC meeting, Dr. Tao gave an overview of DDOE's reporting systems, specifically the Delaware Student Assessment Reporting & Analysis Application. The system, designed by Dr. Tao, offers summary results at various levels, as well as by sub-groups, matched-student match, gap analysis, and longitudinal analysis. Dr. Tao requested feedback on what charts and reports should be included to portray the data clearly.

Discussion Summary

Dr. Luecht requested clarification on the definition of special education. Dr. Tao responded that for clarification purposes, a link would be included to show the makeup. It was agreed that assessment data are a difficult subject to communicate accurately and this should be tailored to the appropriate audience. In the case of parents, for example, what are the five or ten questions that they typically want answered if they look at this data? What would educators want answered? Data reports should be adjusted according to the audience and tables are an effective way to do this.

Drs. Thurlow and Jimenez would like more detail on the alternate assessment data. They would like to see the DCAS-Alt1 percent proficiency ratings as a comparison with the regular DCAS by content and grade.

Summary and TAC Recommendations for the DCAS (Dr. Robert Linn)

DE Assessment Updates

- The TAC expressed an interest in keeping track of Smarter Balanced progress.
✓ **Action: Status update will be provided at the next DTAC meeting.**

DCAS Fall Results

- In looking at the fall and spring DCAS results, the DTAC wondered if the fall results might be deflated due to teacher evaluations and recommends that changes from spring to the following fall be analyzed for possible district differences that might suggest intentional deflation of fall scores to make it possible to achieve larger fall to spring gains.
✓ **Action: Will analyze the data and present at the next DTAC meeting.**

DCAS Technical Report

- SEM curves should be presented with performance-level cuts.
- It would be good to consider the relative efficiency of tests with and without content constraints at the performance-level cuts.
- The validity section should be organized around a validity argument that starts with claims (e.g. the intended uses and interpretations of test results) and present evidence that is relevant to those claims (e.g. dis-attenuation correlation). Special attention should be given to the use of test results for teacher evaluation.
- The DTAC would like to see the responses/actions that are made to the DTAC's recommendations at the next meeting.
- The section on reliability should be organized around the presentation of relevant results, such as the conditional SEM results; the adequacy of model fit; and the accuracy and consistency of pass-fail by achievement level classifications.
- Relevant citations should be added to the Technical Report.
- The Technical Report should be reviewed for statements that are too strong such as these statements "...multi-trait multi-method matrix approach (MTMM is exactly as expected...)".
- The Technical Report should include a justification of the vertical scales by referring back to the previous reports.
- The report should include a discussion of evidence relevant to the dimensionality of the tests.
- The exposure rates should be reported, including over- and under-exposures.
✓ **Action: Discuss the DTAC recommendations with Dr. Harold Doran. AIR will revise the 2011-2012 DCAS Technical Reports.**

Improvement of DCAS

Machine-scored items

- The DTAC recommended that consideration be given to the question of whether or not the machine-scored constructed-response (MSCR) items measure anything that could not be measured by simple multiple-choice items.

- What is the validity of MSCR items? The validity evidence of MSCR items must be collected and addressed.
- The scoring rubrics should be reviewed to see if adding one more score point to the rubric really adds up new information to the item and if all score points are equivalent, which may lead to the reversed step values.
 - ✓ **Action: DDOE will discuss with AIR on the issues related to MSCR items in terms of development, review process, and validation.**

Item exposure rate

- The DTAC recommended examining the information by passages and items that contribute to exposure rate.
- The DTAC recommended that additional exploration of under- and over-exposed items/passages in reading is needed. Perhaps item coding should be reviewed and reconsidered.
 - ✓ **Action: Exploration has been continuing for reading, which includes additional analyses and simulation study by DDOE and AIR. Discussion should be scheduled to explore the possible reasons, especially for under-exposure. Exposure rates should be analyzed for mathematics at the soonest opportunity.**
- Re-calibration issues should be considered further at a future meeting.
 - ✓ **Action: Re-calibration will be discussed at the next DTAC meeting**

Alignment for EOC 2

- There is the concern about using the mean index to distinguish good and poor alignment. It could be that the alignment is not very good in the states that provide this normative benchmark. Thus, being above the mean does not necessarily indicate good alignment.
- The second issue concerns the use of the test specifications to determine the subset of the CCSS that will be used as the target. When you say that a test is well aligned with the CCSS, most people will probably think this means the complete CCSS and not just the subset that is defined by test specifications that are surely narrower than the full CCSS. Both of these issues need to be considered when we revisit the alignment question.
 - ✓ **Action: Conversation was started with Dr. Smithson on the criteria and threshold used for evaluating alignment for EOC2. Revision on the report is expected.**

DCAS Reporting

- DDOE should specify questions and develop data displays to answer these questions.
- DDOE may consider having a focus group meeting to review the online report and improve it.
 - ✓ **Action: A focus group meeting may be needed to improve the DDOE DCAS online report for improvement.**

2012-2013 DCAS-Alt1 Reading and Mathematics – Fall Administration (Dr. Ming Lei)

Discussion Summary

Following the presentation on the 2012-2013 fall administration by Dr. Lei, Dr. Thurlow asked how participation of the ethnicity categories in the DCAS-Alt1 compared to that of the regular DCAS. Mr. Touchette agreed that the African American participation on the alternate assessment looked large when compared to this same category participating in the regular assessment. Dr. Jimenez requested that the disability categories be labeled correctly. Dr. Thurlow requested follow up on the results of those labeled with the term “learning disability” since sometimes the term is used incorrectly, which results in the student being incorrectly identified. Dr. Jimenez suggested that DDOE should look over the actual numbers—by disability—in the various grades by year and compare these numbers for the following year to see if the numbers remain constant. Another concern was for the 8th grade students who were regularly moved into the regular classroom by their parents for diploma purposes. Mr. Touchette agreed that the interpretation of a potential disability was crucial but had been unsuccessful in having the regulation changed to allow for better control.

At the request of the TAC, the training for second-raters was explained. Dr. Thurlow asked if information was available on those students assigned a second-rater. Dr. Lei responded that a summary was available and that second-raters were randomly assigned.

2011-2012 DCAS-Alt1 Technical Report (Dr. Ming Lei)

1. Evidence to support validity argument for intended uses of test scores in high-stakes accountability and teacher evaluation
2. Evidence to support the reliability of test results in high-stakes decisions (e.g., reliability, SEM, and accuracy and consistency of classification)
3. Evidence about the statistical characteristics (e.g., scale validation)
4. Evidence to support meaningful interpretations of test scores (e.g., measure what students with disability know and be able to do)
5. Evidence to support the fairness of DCAS for all students (e.g., administration, scoring and stopping rules, second rater, DIF)

Dr. Lei then presented the DCAS-Alt1 technical report.

Discussion Summary

Clarification was requested on the participation rates. It was confirmed that the scale is set by grade band but is actually being reported by grade. As with the regular assessment, this report also needs to be centered around appropriate validity evidence and cited appropriately.

With reference to the reliability study, Dr. Hambleton requested that standard deviation be included in future tables. In addition, the mean standard error or “n” size should also be included in future comparisons to demonstrate the stability of the means.

It was confirmed that information on test reliability classification accuracy was not included in the manual. TAC members concluded that the format of the DCAS-Alt1 technical manual was well presented. Of special interest was the section on the development of new items. However,

an appendix, the classification of items, and the cut scores were not included. Also omitted was the alignment study, and there was also no data on standard setting.

Questions were then invited on the Stopping Rule Analysis. DDOE commissioned a study to compare stopping rules and to look at higher stopping rule options as a student might score 3 points just by guessing. It was confirmed that the sample size was approximately 390 students. Dr. Lei explained how the data was manipulated for a 4 point if it was administered as a 3 point.

DCAS-Alt1 Standard Setting Report (Dr. Gary Phillips)

Dr. Phillips presented the highlights of the report. A diverse panel of 34 teachers, parents, and higher education leaders met to discuss and recommend performance standards for Science and Social Studies. The recommendations were subsequently presented to and approved for adoption by the State Board of Education.

Accommodations for Alternate Assessments (Dr. Martha Thurlow)

At DDOE's request, Dr. Thurlow shared detail of federal policy and future requirements; current state practices; challenges and issues and finally, recommendations for DCAS-Alt1. Her recommendations are denoted on slide 15 of her PowerPoint.

Alternate Assessment

Technical Report

- Update the categories for student with disability (SWD), e.g., severe mentally retarded.
- Investigate possible reasons for increases in use of Alternate assessment by district.
- Add standard deviations to the table to better interpret SEMs.
- In computing inter-rater agreement, cases should be removed where one rater gives a score and the second mark is not ratable.
- The validity evidence does not "speak" to stated purposes. Revisions should be made to provide evidence that does address the stated purposes.
- DTAC is skeptical that the vertical scale is justified with evidence.

✓ *Action: Discussion between DDOE and AIR has been started in terms of revision of the report and additional analyses. More details will be available.*

Other Issues

- ✓ The TAC commends the state for the setting of standards that are more reasonable than the performance standards that were set previously.
- ✓ There is a need to establish policies regarding the use of accommodations for the alternate assessment.

The meeting concluded at 4:30 p.m.