

## Assessment of Essential Skills Review Panel

### Meeting Minutes

Friday, January 16, 2015

Oregon Department of Education – Basement Studio A

**Meeting Facilitators:** Cristen McLean

**AESRP Members Present:** John Bouchard, Ralph Brown, Steve Christiansen, Lori Cullen Brown, Sarah Cunnigham, Robin DeLoach, Melissa Goff, Dawn Granger, Kathy Hall, Tamika Hampton, Ellen Irish, Shay Mikalson, Laurie Ross, Tiffany Shireman, Jill Sumerlin, Larry Susuki, Marilyn Williams, and Michelle Zundel

**ODE Staff Present:** Cristen McLean, Steve Slater, Ken Hermens, Bryan Toller

#### I. Welcome and minutes

The meeting began at 9:05 a.m. when Cristen McLean welcomed the AESRP committee.

Motion to approve the minutes of the October, 9th 2014 meeting was made by Ellen Irish, Lori Cullen Brown, seconded, passing all ayes.

#### II. Update on Graduation Policy

Jennelle Ives explained that she has begun working with Kim Patterson and they are moving forward with the graduation policy review panel and would like to identify people from AESRP who are interested in participating. [Names were collected via the chat feature.]

#### III. Secure Work Sample Bank

Ken Hermens provided a quick update on history of secure work sample development, including that the ultimate outcome is that there will be 30 tasks in math, reading, and writing maintained in a secure manner and then five additional tasks in each skill area are to be released as exemplars. For the reading work samples, the content that is ready to be released consists of 14 reading tasks, 8 informational and 6 literary, with more emphasis on the informational because students are required to submit an informational sample. Ken explained that ODE received a couple hundred responses for each work sample. A large portion of these were double scored. Currently he is working on selecting student work and developing commentary for released tasks so educators can calibrate their scoring using these as exemplars to score locally. ODE also did a very large field test of a Smarter Balanced-like task, which includes multiple passages (four or five pieces of resource materials). These will be scored as both reading and writing work samples because there are some reading questions that are preliminary to students writing their essay.

Ken explained that there are 27 writing work samples ready and many of them are actually writing prompts from OAKS that have already been field tested but that can be released now because of the transition to Smarter Balanced. Currently, there are eight narratives, seven expository, eight persuasive and four imaginative. Some lessons learned from the field testing were that graphic organizers are very helpful, (things like T-charts and Venn diagrams) and also to specifically state

to cite evidence. When these were included in the work samples, students wrote more carefully and completely rather than just asking them open ended questions with blank space. Finally, the Smarter Balanced-like tasks were very time consuming for students to complete and in some cases, caused exhaustion and fatigue to affect the scores. A lesson learned is that longer work samples should be scheduled over multiple test sessions.

Bryan Toller explained that in math, there are 3 algebra, 1 geometry, and 3 statistics work samples ready. Lesson learned from the field test include that it was really helpful to field test because it clarified what the target rigor is and provided a full picture of how a task performs. Additionally, we learned that reflecting and evaluating are areas where students have to be trained. One of the goals and requirements of all of the work samples that are being written for the secure bank is that they can all be solved in different methods. Similar to reading work samples, the remaining math work samples will be piloted in individual classrooms.

Cristen McLean discussed the process of distributing the work samples. She described that in the fall, ODE talked with AESRP about student eligibility for accessing the work samples and collected feedback. ODE has also collected feedback from additional committees and determined that the secure work samples would be available only to students who are in the 12<sup>th</sup> grade for the 2014-15 school year who do not have a passing essential skills score using approved methods. Those students can receive up to two work samples in each content or essential skills area. So a student without a passing essential skills score in reading could receive up to two reading secure work samples. Likewise, in writing, up to two secure work samples. Cristen described that the secure work samples are different than other secure test content because they are scored locally so there are some unique considerations with regards to thinking about security on this topic. ODE is trying to find the right balance between a system that is too rigid (difficult to access and/or student inability to be successful) and a system that is too loose (tasks are not secure and may have been accessed prior to administration). Given this tension around security, here is what is already decided: secure work samples will be posted in an application in the secure district site and the application will be launched in mid-February. Access will be limited, anyone who is going to access the content will need to have read the security protocol and signed an assurance. The only way to see a secure work sample is by downloading it, and each assurance for a download will be tracked so ODE can follow up to find out which students received which work samples and we can understand the impact of the bank. In order to facilitate the decision of which work sample to download, a list of the task content will be posted for the general public and will include a summary and title. That information should be used to determine which sample to download and administer to a student. Also secure work samples may not be emailed. This is consistent with previous protocol. Cristen asked, "Should secure work samples be completed only in paper and pencil format?" Some members expressed concern about students not being able to word process their written responses. Cristen asked which improprieties should get reported clarifying that if there is too stringent of a policy it will become untenable since there is not additional FTE to support this. Some members supported reporting only the most severe improprieties including posting online, on social media, or releasing student information. Additional guidance may be needed to remind teachers that teaching to the work sample is not allowed for secure content. Cristen asked how modifications and translations can be completed securely and members responded that people completing translations and modifications must also sign security assurances. A member asked a question about whether speech to text was allowable for reading

work samples and Ken clarified that it is. The last question was whether districts should be able to securely store a copy of the secure work sample with scored student work. Cristen explained that ODE's protocol for all secure content other than this is content must be destroyed immediately after administration. With work samples, though, they are scored locally. Some members thought destroying immediately was appropriate but questions were raised about how much information the district needs in the event of an appeal. The student work can't be separated from the secure work sample in many cases so some members suggested allowing districts to securely store the work sample and student work until the student has graduated.

#### **IV. Evaluation of Official Scoring Guides against Common Core**

Ken provided a quick recap of the work on the Writing scoring guide, including that ODE engaged a panel of people from across the state suggested edits for a new draft then created two new drafts, one that is argumentative/informational and then a separate narrative one. The most significant changes on the scoring guide were at the 5 and 6 levels. Specific traits that were most impacted, were ideas and content, organization, and citing sources where the emphasis shifted from the formatting of an essay to the integration of resource material. In most cases the changes were more related to terminology than to substantive alterations. In the fall AESRP meeting, ODE provided a comparison of what the actual edits were to the writing/scoring guide. ODE also developed a Smarter Balanced-like task in order to test out the scoring guide. Ken described training that has been conducted on the scoring guide and results from initial applications of the scoring guide seem positive. The next steps for testing out the scoring guide include training raters on the scoring guide and applying it at least one of the score sites for the state writing assessment of senior retests where Ken will include some papers that haven't been touched yet and have them apply the new version of the scoring guide to those papers first and do double scoring. Once this is complete and the information has been scanned, ODE will conduct some comparisons and analyses.

Ken introduced the research questions that are guiding this scoring guide and stated that additional suggestions are welcome.

- Where do the revisions on the draft guide result in more or less rigorous expectations? Since raters have scored all of these already using the existing current official writing guide, there are a lot of scores that can be compared on the exact same papers to scores given using the draft version of the scoring guide to see if there's a difference in the level of performance.
- How does the inclusion/exclusion of individual trait scores impact passing rates? Right now ideas and content, organization, sentence fluency, and conventions are the 4 required traits. What if students were required to have a minimum score in citing sources or use of sources piece? What if voice or word choice was added?
- Thirdly, and this was brought up by one of the panelists in the fall, what would a compensatory model look like, would it be workable and desirable? Compensatory meaning a composite score making a passing or no passing decision as opposed to the current system where they have to get at least a 4 in each of the required traits. So if there's a composite score of 20 with single scoring, would conventions be double weighted? Would certain traits require a 4 even if the composite score was a certain number?
- Lastly, are there particular elements that should be weighted or have a minimum score associated with them beyond others? Should something be accepted as passing, something

that gets a 3 in conventions or potentially lower than that, when the student can get a 5 in ideas and content and in organization, should that compensate? Should a paper that gets a 3 in ideas meet? Is it important for the ideas to be up to a certain proficiency level? So a lot of things to consider in that respect.

Ken also said that the middle school version has not been started yet but that he's hoping to revisit this in the spring. A member asked about the predictive value of the Writing scoring guide on Smarter Balanced Assessment and Ken explained that the revisions make the scoring guide more similar in terms of content to Smarter Balanced.

#### **V. Transition to Smarter Balanced: Methodology for Essential Skills**

Cristen introduced that AESRP and ODE are trying to identify the level of rigor on Smarter Balanced as compared to the "meets" level on OAKS. The methodology ODE introduced in the fall described a direct linking methodology where ODE would take student performance on OAKS and link that with Smarter Balanced field tests. Following that analysis, ODE proposed to convene a qualitative panel of educators and go through qualitative standards verification. The timeline for that approach was that scores would be adopted in August, at the earliest. During the discussion in the fall, many AESRP members identified challenges with this approach surrounding concerns about the heavy reliance on field test data and lack of operational data. Cristen explained that since the fall meeting, ODE has continued to reflect on this feedback. ODE has improved the methodology to include two additional analyses that draw upon operational data and will only employ qualitative standard setting if there is not reasonable confirmatory evidence across the three quantitative analyses. Thus, in addition to the direct linking, there will be a second analysis to embed OAKS items in the operational Smarter Balanced and the third method is an equi-percentile method, where ODE looks at the percentage of students who earned a "meets" score on OAKS and identifies an equal percentage of students passing on Smarter Balanced and what that score would be. Each method will be completed independently and will include analysis of the subgroups as well as the total population with a lens towards equity. If there's not strong confirmatory evidence, ODE would then engage professional judgment, and that would allow ODE to identify additional insights and look at the Smarter Balanced achievement standards given some additional resources. The timeline for this is that the scores would be adopted in October, at the earliest, and the strengths are that overall this higher quality data allows us to have stronger inferences and more confidence in our conclusions. In order to accomplish this approach ODE requests a subgroup of AESERP be formed to be available for 2 to 3 meetings, between April and July, to provide input on what the analyses are revealing. Additionally, another full AESRP meeting will need to be added in August, in order to bring a recommendation to the state board as a result of these analyses.

##### **1. Have we missed anything?**

A member asked whether embedding items would make the test longer and Steve Slater clarified that embedding items to field test them has been standard practice in Oregon for a while. For this administration, the estimate is the additional items would add between two minutes and eight minutes for a test with each student only seeing one to six additional items. A member asked about whether there would be information released across spring as a result of the initial analyses and Steve said he would need to discuss this with management, but because this is a phased series of analyses the first results would not be appropriate to make

decisions based on, they would likely be shared with AESRP to provide a first look at how well the two scales were lining up. A member asked about how this achievement level setting was different from previous achievement level setting processes on OAKS for systems accountability. Cristen clarified that once the achievement level is set on Smarter Balanced for Essential Skills, it cannot be raised without adequate notice. A member asked about why the same analyses that were used to support the adoption of the Smarter Balanced achievement levels couldn't be used for Essential Skills. Steve and Cristen explained that the analyses presented there were to explain how the OAKS scale lined up with the Smarter Balanced scale not the other way around and that those analyses exclusively used field test data, which AESRP expressed concerns about in the fall. Some members expressed concerns about the timeline of receiving the final achievement levels (adopting the achievement level in October rather than August) because students will already be enrolled in course, such as remedial courses focused on work samples, at that time. Additionally, districts will likely use all other available assessment methods while waiting on the results and then may find that the student had actually passed on Smarter Balanced thus the additional testing was unnecessary.

2. Is there anything we should add?

A member asked for clarification to be added as to what will happen if there is additional feedback about the methodology or analyses and changes are needed. Cristen explained that the subgroup will form an important part of the process by providing an early review and helping identify if adjustments are needed. Additionally, she explained that the qualitative standard setting process will be employed if the analyses are not producing confirmatory results.. The timeline proposed is the earliest possible date given the date when ODE will receive operational test results and when the State Board meets. If ODE needs to employ the qualitative standard setting process, the timeline will be longer.

[Members volunteered to be a part of the subgroup via the chat feature.]

**VI. Essential Skills and Local Performance Assessment Manual**

Cristen explained that ODE will be releasing a mid-year update to include guidance on the secure work samples.

**VII. Updates**

Cristen McLean provided participants with updates on a variety of projects that ODE is leading that are related to assessment but not directly related to the Essential Skills. Cristen thanked participants for their engagement and contributions.

**VIII. Adjourn**

The meeting was adjourned at 11:15 a.m.