Colorado Content Collaboratives
Technical Steering Committee Meeting

The Warwick Hotel: Executive Lounge (15th floor)
1776 Grant Street, Denver, Colorado 80203
1-303-861-2000 http://www.warwickdenver.com/

May 10-11, 2012
Agenda

Thursday, May 10
8:30   Breakfast

8:45   Welcome and review of agenda and meeting objectives
Jo O'Brien, CDE and Scott Marion, Center for Assessment

Primary Meeting Objectives:
1. Provide advice on the scoring approaches and thresholds with the latest version of the content review tool.
2. Provide feedback on the quality of reviews and assessments being "partially recommended" or "recommended" by Cohort 1.
3. Provide advice on the improving the assessment search specifications for the content area research staff.
4. Assist CDE in thinking through the context of use—particularly for determining student "growth"—in the evaluation of assessment technical quality.
5. Provide recommendations on a design for field testing the selected assessments in pilot districts.

9:00   Overview of work accomplished by the CCCs
Jo O'Brien, CDE and Angela Norlander, CDE

Cohort 1 met for the final time during the first week of May in Colorado Springs. Cohort's 1 work focused exclusively on the identification and selection of assessments deemed to reflect high quality content and to provide useful information for teachers. A sample of these assessments will be tested out in the pilot districts before the final decision is made to include these assessments in the resource bank. This overview will describe the set of activities pursued in the three meetings since the TSC last met and present the strengths, limitations and any outstanding concerns about the process pursued with Cohort 1.

10:00   Review of revised version of content review tool and accompanying summary
Dianne Lefly, CDE

CDE has modified and streamlined the content review tool from the version reviewed by the TSC in February to help simplify and facilitate the review process for the CCCs. Scoring criteria were also added next to each question to report on a summary of scores earned across each dimension assessed. Dianne Lefly will walk through the four quality criteria represented in the tool and present a summary overview of assessments reviewed relative to the quality criteria.

Questions for TSC:
1. Should the scoring weights in the tool be adjusted to emphasize items in the tool that should be prioritized over others?
2. Does the summary picture tool provide information at the right grain size to inform users about the strengths and weakness of assessments being selected in this process?

11:00   Reviewing sample of assessments:
Angela Norlander, CDE and Elena Diaz-Bilello, Center for Assessment

We have assigned TSC members and staff into one of four groups to review a sample of assessments and completed review forms in six content areas

TSC will first be asked to use the tool to review a sample of assessments in the six content areas. Once these reviews are completed, TSC members will compare their reviews with the reviews already completed by the Collaborative members. In addition to this comparative review, we want the TSC to identify the strengths and limitations of the different type of assessments being selected in each content area by the CCCs and the quality of the reviews being generated by each content group. While reviewing these assessments in workgroups, consider the following questions:
- In addition to the modifications recommended by the CCCs reviewing each test, what other modifications/considerations should be considered if a select number of these assessments are to be field tested by the pilot districts?
- To what extent do the reviews provide sufficient or compelling evidence that the assessments have been reviewed in a technically defensible manner?

12:00   Lunch

1:00 **Whole group debrief on review of assessments and completed tools**
Scott Marion, Center for Assessment

Provide input and feedback to CDE on:
- The apparent accuracy of the completed reviews,
- Quality and level of evidence provided on the review tools,
- An overall sense about the quality and type of assessments that are at least partially recommended for inclusion in the resource bank by virtue of the CCC reviews, and
- The types of assessments, e.g., complex performance tasks, that may not make it through the review screen.

2:00 **Structuring the Resource Bank**
As mentioned at the previous meeting, CDE is creating a "resource bank" to help support the work of districts as they implement the educator evaluation requirements. The work of the collaboratives is meant to support the assessment component of the bank and much of the work that we have been reviewing today is designed to populate the resource bank. CDE is interested in discussing with the TSC how to think about the overall structure of the bank. For example, much of the selection efforts discussed thus far have focused on having CCC members evaluate nominated assessments for inclusion into the bank without a particular structure like one would use a test blueprint to select items to populate a test form. In essence, CDE would like to engage the TSC in helping them think through a "blueprint" for the assessment resource bank to help guide the selection/development of types (e.g., standardized tests, performance tasks) and numbers of assessments.

However, CDE recognizes that no matter how carefully they structure and populate the resource bank, districts will have control over the assessments they choose to use from the bank. Therefore, CDE would like to discuss with the TSC how they might best be able to guide districts in appropriately selecting/developing assessment to create a valid assessment system.

NOTE: A break will be taken at an appropriate time in this discussion.

4:30 **Wrap-up and adjourn for the day—dinner on your own.**

8:00   Breakfast

8:30   Considerations of use context—particularly for determining student "growth"—in the evaluation of assessment technical quality
Scott Marion, Center for Assessment

Many of the assessments identified by the CCCs and field tested by pilot districts will be used to evaluate student performance as part of the teacher evaluation system. The scores of some of these assessments may have an adequate score range and interval-like qualities while scores from other assessments will be restricted to course-grained ordinal scales (e.g., 4 or 6 point rubric scores). Nevertheless, the evaluation of an assessment used to measure student performance at a single point in time (status) may be quite different than an assessment used to measure "growth" over at least two occurrences. In populating the resource bank, CDE is wrestling with the very difficult challenge of trying to anticipate uses in a very local control state. The paper by **Goe, et al.** (separate attachment) provides an overview of how student growth may be conceptualized for non-tested subjects and grades and will help provide context for this discussion. On the other hand, the draft **Marion, et al.** paper (separate attachment) points out many of the limitations of using most non-state administered assessment to try to calculate a "growth" score. While the paper was written to address the challenges of determining growth in a Student Learning Objectives (SLO) framework, many of the technical issues relate to the discussion here. The document from **Indiana** (separate attachment) provides an example of how one state is addressing the issue of evaluating student performance results for non-tested subjects and grades.

Questions for TSC:
1. Should the technical quality evaluations of assessments included in the resource bank simply state that the assessment appears to have technical quality sufficient for measuring student achievement at a single point in time for a specific set of content standards?
2. Does the TSC recommend that CDE provide guidance and professional development for districts so they may better evaluate the technical adequacy of their selected analytical approaches for supporting inferences about growth? Does the TSC have suggestions for how best to approach this challenging task?

10:00 Design for field testing assessments in pilot districts
Angela Norlander, CDE and Elena Diaz-Bilello, Center for Assessment

A selected number of assessments identified by the CCCs will be field tested in pilot districts. In some cases, the field test will provide information about assessments where the technical quality is unknown and in all cases, the field test will help determine whether teachers across various types of districts value the information from these assessments. However, given that CDE does <u>not</u> have the resources to field test any more than a sample of assessments, CDE requests feedback from the TSC regarding the purpose and structure of a pilot test for these assessments in pilot districts.

<u>**Questions for TSC:**</u>
1. Since CDE will only field test a small portion of assessments, can the TSC help CDE clarify the types of information it should seek through the field test?
2. Following on #1, how should CDE determine the number and type of assessments that should be included in a field tested by level and content area?
3. If this is considered more of an "item tryout" rather than a field test, how should CDE think about the numbers of participating schools and students while ensuring that the field test includes appropriate representation of students with disabilities and English language learners?
4. Conducting traditional psychometric analyses may be a challenge give the voluntary and non-standardized nature of the field test, therefore, what types of analyses and what types of information does the TSC recommend that CDE should collect as part of the field test?
5. How should CDE use the field test results as part of the information provided in the resource bank?
6. Can the TSC help CDE identify potential partners for evaluating the processes and outcomes from field testing selected assessments?

11:45 Wrap-up and next steps
- Next meeting: Thursday, August 2nd

12:00 Box lunch and adjourn