Colorado Content Collaboratives
Technical Steering Committee Meeting

Marriott Courtyard Denver Airport
6901 Tower Road -- Denver, CO 80249

August 2, 2012

Notes

Welcome and review of agenda and meeting objectives

Jo O'Brien, Colorado Department of Education (CDE)

- Introductions:
  - Technical Steering Committee Members:
    - Sue Bechard, Inclusive Educational Assessment
    - Tim Brophy, University of Florida
    - Laura Goe, ETS
    - Kristen Huff, New York Regents
    - Jacqueline Law, Colorado Springs School District 11
    - Todd Morse, Academy District 20
    - *Absent: Guillermo Solano-Flores, University of Colorado at Boulder*
    - David Webb, University of Colorado at Boulder
  - CDE Participants:
    - Joyce Barrett, Exceptional Student Services
    - Bill Bonk, Accountability and Data Analysis
    - Sed Keller, Educator Effectiveness
    - Toby King, Educator Effectiveness
    - Dianne Lefly, Assessment, Research & Evaluation
    - Candy Myers, Exceptional Student Services
    - Tricia Miller, Vision 2020
    - Angela Norlander, Assessment, Research & Evaluation
    - Jo O'Brien, Assessment, Research & Evaluation
    - Britt Wilkenfeld, Educator Effectiveness
  - Participating Guests:
    - Bob Good, Denver Public Schools
    - Patrick Mount, Thompson School District
    - Margie Ruckstuhl, Harrison School District
- Update on the participation of the National Center for Improving Educational Assessment (NCIEA): The NCIEA contract with CDE is being repurposed to support the work of the Content Collaboratives. The Technical Steering Committee (TSC) will move forward to find practical solutions to combining

assessment measures and make recommendations for how student growth may be interpreted and calculated.

- The TSC is now shifting from establishing quality criteria for a single measure to determining how multiple measures could be combined to showcase a body of evidence. Then, making determinations as to how growth calculations could be created to inform an educator's evaluation.
- CDE will move forward with constructing the technological infrastructure of the resource bank and establish definitions for common assessments terms.
- Today, the Technical Steering Committee Members will respond to predetermined questions that focus on strategies for combining measures and determining student learning over time. Colorado school district assessment leaders will be present to ask questions and provide feedback. The focus of the conversation will be in support of the following objectives:

Primary Meeting Objectives:

1. Provide recommendations for guidance that should be given to Colorado school districts which will utilize the combination models.
2. Provide advice to CDE on the appropriate and inappropriate combinations of multiple measures that can be used to determine student growth.

David Webb

- Discussion Question 1: *What is the best advice regarding how to select and weigh measures which either leverages the accuracy of standardization or the more open-ended student demonstrations of mastery?*
    - David Webb: Pros and cons to standardization. Pro: Want comparability (across teachers, schools, districts, states). Con: Loss of autonomy and standardized scoring is not always precise. Best advice: Involve teachers. Assessments will need to be agreed upon; teachers will need to be engaged to make those decisions and buy-in to system. The 50/50 split indicates that professional practice is not valued over student growth or vice versa. Teachers must be engaged in the assessment process to an extent to which they never have before. High performing countries do this. Professional practice improves as teachers become more engaged and have a better understanding the use of assessment data.
    - Tim Brophy: The University of Florida is using teacher cohorts to develop arts assessments, improve practice, and raise the professional level of being an educator.
    - Todd Morse: The vision is clear, but how do we take the first steps toward that vision; how do we stay out of the weeds while moving toward that vision?

- Toby King: Raising the professional nature of teaching through this process creates a sense of integrity.
- Margie Ruckstuhl: In Harrison School District, peer scoring had anomalies, i.e. leniency. As a result we have moved to blind-scoring, and scorers provide input on scoring rubrics.
- Sue Bechard: Experience with scoring portfolios. Calibration of scoring rubrics is an on-going process; re-calibration over time is essential. No one scores their own. This is a very effective professional development opportunity. Districts cultivate expertise in scoring and become resources for their colleagues.
- Jacqueline Law: How do we deal with shrinking budgets and lack of expertise in small districts?
- David Webb: Consider "trading" with other districts; use resources from other regions. Use the BOCES or other regional resources. Use technology for virtual scoring.
- Bob Good: Concern over objectivity and subjectivity. Don't be wrong. Don't be inconsistent. Focus on the information you really want to collect. Formative practices don't require a high level of standardization, nor should they. The conceptual shift that needs to take place is providing evidence to determine what the educator done to have a causal effect on student performance.

➢ Discussion Question 2: *What are the best design combinations of tier 1, 2, 3 types which fairly and accurately might be combined into a final annual body of evidence?*

- David Webb: Identify measures that will offer teachers the right information to improve instruction. Answer this question and the body of evidence will be easy to identify. National/State initiatives should provide information to teachers about the kinds of evidence they need to adjust instruction.
- Sed Keller: Example: use formative practices to identify misconceptions; and have students engage in the solving of misconceptions. Provide more models and support for teachers to better understand the use of the information they are receiving from assessments.
- Jo O'Brien: The goal of the resource bank will be to connect the tier 1 and tier 2 assessments to formative practice.
- Toby King: Teachers need to be able to track how they have been able get to where they are – what methods and practices have the used and applied?
- Sue Bechard: What is the role of the student? Formative practice must include student participation.

- Jo O'Brien: The guidance should include the high value we have on formative practice and how it connects to the tier 1 and tier 2 assessments that can be used for some kind of judgments.
- Bill Bonk: Let's not forget that standardization is a positive value. There are opportunities for demonstrations of success within the parameters or constraints of standardization.
- David Webb: The use of information gained from assessments is the most important component to consider.

Kristen Huff

- Discussion Questions 3 and 4: *Which different modes of quality assessments are better paired than others (multiple choice, constructed response, performance tasks, etc.). How do an evaluator and educator pre-determine the right balance of reasonable measures in a way that fosters insight and improvement, not blaming or gaming?*
    - Kristen Huff: This reform effort is about assessment literacy with practical restraints in the classroom. Most important role we can play as assessment leaders is to provide opportunities for productive conversations to find innovative solutions. The purpose of multiple modes should be driven by use of the assessment information you hope to get. Example: college entrance requirements --- GPA's (varies across schools), SAT, ACT, essays, letters of recommendation -- they paint a portrait; some measures are consistent, some are not. The best way to deal with inconsistencies is to indicate the priorities. Standardization is not bad, it has technical value. Balance is key. Because of the judgments that are related to assessment there is a fear of "getting it wrong" and of subjectivity. We need to break through these fears and embrace the subjectivity by being transparent about value judgments. These judgments are protected and validated by procedural transparency. Guidance needs to include what and why we value performance tasks, for example. Be precise about why we value various types of assessments.
    - Jo O'Brien: What are the procedural validity checks?
    - Margie Ruckstuhl: Decision point: What are the combinations of assessments that we need to develop student growth evidence?
    - Kristen Huff: Districts have expertise that should be leveraged. Additional expertise can be leveraged from neighboring districts and the State. Questions to consider: Do we have the right people making decisions? Do we have a shared understanding about the trade-offs that exist in making decisions? Are the decisions that consensus-based? Are the participants (via documentation) supportive of the decisions? Every decision will have trade-offs. Focus on the values in

your district and stand behind it. And, be willing to change and grow as you learn more.

- Laura Goe: Small district capacity is limited. Guidance documentation and training needs to be in place. Professional development budgets need to be re-purposed to support this work.
- Kristen Huff: There are technical assistance funds available at this time to assist districts. This is a place where CDE can be helpful. The commitment to this work at the district level has to be there.
- Toby King: Districts are struggling to make sense of their "pie charts" of assessments. Many don't make sense. The guidance from CDE will be critical. Let's not let perfection get in the way of progress.
- Kristen Huff: This leads us back to the need for assessment literacy.
- Laura Goe: Guidance needs to be a productive model so it doesn't become a checklist without an understanding of what it means. The guidance needs to be a productive process, not a compliance issue.
- Jo O'Brien: To sum up – we have a lack of capacity among districts as to what the technical components of an assessment system should look like.
- Bob Good: You can find subjectivity in every type of assessment. We want to focus on having this process be less mechanistic. What is the process for these judgments to be made? It needs to be more sophisticated than what we currently have.
- Todd Morse: We need to make room for professional judgments.
- Jacqueline Law: Small districts need to participate in these discussions.
- Jo O'Brien: How do we draw meaning from these judgments? How do we show value? How do we combine these measures to show a body of evidence?
- Kristen Huff: There is no one way to combine the multiple modes. CDE needs to provide guidance/considerations to districts to help them determine their values.
- Jo O'Brien: We need to create the right questions.
- David Webb: Looking at a districts' current selection of measures, in this case there a many measures, the question is whether they measure what you really want to know. Alignment and consistency of measures to the district's values has to be the priority.
- Margie Ruckstuhl: Some districts simply want a template and are willing to learn along the way. Other districts need an opportunity to defend and improve their existing systems. Harrison created templates that are content-based but have common elements.
- Toby King: The TSC needs to come up with a few templates that districts can consider and use.

- Bob Good: And, the templates must have guidance in order to avoid unintended consequences.
- Jacqueline Law: I support the idea of crafting a list of questions to defend your decisions.
- Kristen Huff: Transparency leads to precision. Rational for decisions is a necessary component. In New York, workshops were conducted to walk districts through a process for selecting and weighting assessments.
- Sed Keller: Can the TSC validate some of the models CO districts have already developed?
- Tim Brophy: The validation process is necessary to establish defensibility.
- Laura Goe: In the spirit of Kahn Academy, can CDE develop "making your pie chart" video?
- Patrick Mount: Highly support the validation process.
- Kristen Huff: Since there is no "right" answer it is necessary to create a variety of models, with rational and values explained.
- Bill Bonk: Provide an anonymous process for a community of evaluators to give feedback to districts.
- Toby King: What does it mean when, for example, 25% of an assessment is represented in the pie chart?
- Sed Keller: We may need to dig into an existing template (pie chart) to determine what the questions really need to be.
- Kristen Huff: What does the percentage reference mean?
- Toby King: It is a description of how a rating produced. This "student growth" rating is then combined with the professional practice component to come up with the final rating for an educator.

Laura Goe

- Discussion Question 5: *Who has thought deeply about designing professional bodies of evidence outside of education and what has their experience taught us?*
    - Laura Goe: There is much we can learn from other professions about how they approach to evaluation, but very little deep thought has been done on the topic. Robert Mislevy notes that data cannot rise to the level of evidence without a hypothesis. It's not just looking at test results; it is also about what the teacher did to support the student to produce the test results. Examples: The focus of health care is competency and mastery – not comparison. A lot more training and certification is required than in education. Significant efforts are made to support health care providers. Whereas, teachers are sometimes are looked at as expendable. In health care if a patient dies they look

at what the provider did/didn't do. Often the fact that the patient died is irrelevant if protocol was followed. In education we don't do that. We hold new teachers to the same level of accountability as 20-year veterans; other industry is not that way. In sales, the context of the territory is taken into account. Some areas require more or less of a quota. We don't do this in education. In sales, very little training and support is given. If you don't perform, you're out. Civil service workers – standards exist; evaluators choose key performance elements based on their job. Demonstrating work performance is taken into account, not just showing up every day. Police -- Positive evaluation includes public comment. Although, when there are unrealistic expectations, gaming exists – like in education. Evaluators are also judged by their supervisors – similar to principals. Evaluators are looked at for consistency. Evidence is gathered from the people being served, and whether or not the officer followed procedures and protocols.

- Bill Bonk: The incentive has been built in to maintain the status quo. The system has a bias-down component.
- Toby King: This is built into the state system assurances. We look for inconsistencies and do an audit. For example, a turnaround school with high performing teachers with low performing students. Or, it could turn out that it is evaluator error. With smaller districts, their data can be aggregated across similar districts.
- Margie Ruckstuhl: Harrison runs scenarios to adjust the extremes. Dealing with small "n's" is a problem but we've come up with some ways to equalize the issue.
- Jo O'Brien: What are the comments on the other industries?
- Patrick Mount: If I'm the parent then how do I justify my child not learning if the teacher is supposedly doing everything right?
- Laura Goe: It is both competency and protocol based.
- Kristen Huff: Often the higher the bar the less evaluation is needed. Because it is based on achievement. There is a relationship between training and evaluation.
- Jo O'Brien: Sales responses?
- Margie Ruckstuhl: Student achievement goals.
- Laura Goe: Perhaps it's about expectations. We don't want to have different expectations for different students, schools, districts. It's a dilemma.
- David Webb: Corollary in education – you don't want to stop teaching if your kids hit expectations.
- Todd Morse: Be careful about "adjusting" expectations.

- Sue Bechard: There is a natural norming tendency. Learning disabilities often vary among districts.
- Toby King: Prefer to go to a criterion referenced system with teacher evaluation. It requires that districts put the right teachers with the students who need specific support. We need to be careful about teachers leaving lower performing schools to avoid negative evaluations.
- David Webb: Is the consideration of instruction embedded in development of the assessments?
- Laura Goe: No, assessments were not designed with the teacher's role in mind. They are not instructionally sensitive. That is why the teachers need to be involved with assessment development, selection, administration, scoring. Standardized features need to be in place, however. Focus needs to be on the evidence collected and what the teacher's contribution has been to the learning.

Sue Bechard

- Discussion Question 6: *What are other examples of mixing multiple measures which definitely do not accurately portray student learning gains?*
    - Sue Bechard: Some caveats to consider when assessing students with disabilities who participate in general assessments
        1. Triangulating data on the achievement of students with disabilities is difficult, even in a status model. In one study, five profiles of low achieving students with disabilities were developed. In some profiles, assessment data were in conflict with classroom work, teachers' evaluations/grades, and the level of coursework the students were taking (Parker, Gorin, & Bechard, in process).
        2. Relying on predictions/perceptions about achievement of students with disabilities may be misleading.
            - Teachers' predictions about student achievement often underestimate abilities.
            - IEP teams' predictions about students' need for supports/accommodations are often inaccurate.
            - Students' self-assessments are often erroneous, in both directions.
        3. Some assessments have not established validity for students with disabilities. The interpretation of assessment results must consider how students with disabilities are addressed in the assessments' development and administration policies.
            - Some assessments have not included students with disabilities in validation studies (especially if they have been developed prior to 1997).

- Different assessments allow/disallow different administration practices. Combining test results when some allow accommodations and others do not may not be appropriate.

4. Consider the characteristics of persistently low performing students with disabilities when selecting assessments.
   - Research has highlighted several common characteristics: poor metacognitive and executive functioning skills, difficulty generalizing, low reading fluency (decoding difficulties mask comprehension skills), limited working memory, difficulty with focused and sustained attention.

- Discussion Question 7: *What advice is best to consider when facing numerous student ranges of ability?*
  1. Pay attention to principles of universal design
     - Students with disabilities are a heterogeneous group, not only in their range of abilities, but in the impact their disabilities have on their learning and achievement.
     - Look for assessments with multiple methods of presentation and options for response. One approach cannot work for all students.
     - Look for assessments that are maximally accessible to the greatest number of students.

     Resources on UD :
     - http://www.cast.org/learningtools/index.html
     - http://www.cehd.umn.edu/NCEO/TopicAreas/UnivDesign/UnivDesignResources.htm
     - http://www.osepideasthatwork.org/udl/intro.asp
  2. Look at allowable accommodations
     - Look for assessments that allow sufficient accommodations that do not invalidate the target construct (which means it is very important to know what the target constructs are!).
     - Select assessments that have comparable lists of allowable accommodations. It is very difficult to ensure that teachers will follow different guidelines for every test. Therefore it is difficult to ensure that students are getting the accommodations they need without invalidating the assessment results.

     Resources on Accommodations :
     - http://www.eric.ed.gov/PDFS/ED531522.pdf
     - http://nichcy.org/research/ee/assessment-accommodations
     - https://apps.cehd.umn.edu/nceo/accommodations/
     - http://ccsso.org/Documents/2005/Accommodations_Manual_How_2005.pdf

3. Pay attention to cognitive complexity in content and process.
    - Look for assessments with a range of cognitive complexity/depth of knowledge. One can ask an easy question or a hard question about the same concept. A combination of assessments should cover an appropriate range of complexity.
    - Test items should only demand a level of cognitive complexity that is intrinsic to the targeted construct. (i.e., Science tests that use technical vocabulary not necessary to the item, multi-step tasks that can be broken down).

    Resources on Cognitive Complexity/Depth of Knowledge:
    - http://www.accessibletesting.com/Research%20Projects/arm.html
    - http://www.cde.state.co.us/cdeassess/UAS/AdoptedAcademicStandards/Depth%20of%20Knowledge%20for%20the%20CAS.pdf
    - http://ccsso.confex.com/ccsso/2010/webprogram/Session1381.html

- Kristen Huff: Focus on the intent of the complexity. Does it need to be cognitively complex or is the item simply obscure?
- Toby King: Guidance needs to include alternate assessment guidance.
- Bob Good: There is a tension between good instruction and individual accountability.
- Toby King: 191 requires individual and distributed attribution. Also, movement towards the goals of the IEP can be used.
- Laura Goe: Assessments that are closer to the classroom are more useful to special education students. Focus on classroom-based assessments that are standardized. Close to the classroom means that the teacher has an active role in the administration and use of an assessment.

4. Select assessments that can be embedded during instruction
    - Look for assessments that employ approaches that resemble what is done during instruction. The more familiar the assessment environment, the more students will be able to understand what they are supposed to do.
    - Make sure students have the opportunity to practice using the format, etc. demanded by the test. The interpretation of assessment results should not be confounded by novel assessment circumstances.
    - Curriculum based assessments used for progress monitoring can provide consistent ongoing data over time.

    Resources on CBM
    - http://www.easycbm.com/info/reading_assessments.php

- http://ncaase.com/publications/tech-reports

5. Look for assessments that are engaging to the students, including TEA.
   - In a number of studies, students with disabilities expressed more enthusiasm for technology-enhanced assessments (TEA).
   - Students are more likely to use accommodations that are delivered via computers with headsets than in situations that single them out.
   - Students are more likely to complete assessments they find engaging.

   Resources for TEA
   - http://measuredprogress.org/mp-and-sri
   - National Center on Assessing the General Curriculum: http://aim.cast.org/learn/historyarchive/backgroundpapers/technologies_supporting
6. Involve special educators in selecting the assessments

   Resources on research/growth models for special education

   National Center on Assessment and Accountability for Special Education

   "The purpose of the NCAASE (The Center) is to develop and test various approaches for measuring the achievement growth of students with and without disabilities."

   http://ncaase.com/about/research-agenda

   ETS/K-12 Research on Students with Disabilities

   https://www.ets.org/Media/Research/pdf/SPOTLIGHT4.pdf

- David Webb: Teachers use accommodations identified with standardized tests but often don't think about them for classroom assessments. There is also a need to really educate teachers regarding DOK with focus on the reasoning aspect versus what is "hard" to remember. The state model evaluation rubric for teachers requires the implementation of recommendations made by special education support staff.
- Bob Good: There is a difficulty in measuring student growth in the very low and very high quadrants. Mixing norm versus criterion based world makes growth discussions very difficult. It is complicated by state versus district measures.
- Sue Bechard: Begin with valid assessments, and the valid use of those assessments.
- Laura Goe: A mechanism in place for students who are not on the same path as other students. Assessments need to be aligned to curriculum and the right curriculum needs to be available to students as needed.

- Toby King: This is a discussion of mastery versus seat time. Should students be able to "test out" of course work?
- Sed Keller: This issue brings to light the need for student learning objectives (SLO's) as they can be individualized.
- David Webb: The Smarter Balanced Assessments and Dynamic Learning Maps alternate assessments will be computer adaptive and could support this work as well.
- Sue Bechard: We need to stop measuring what students don't know, but what they do know.
- Kristen Huff: There is an opportunity with the development of the new models we have been discussing that can address these concerns. Assessment cannot solve all education issues.

Tim Brophy

- Discussion Questions 8 and 9: *What are examples of such combinations with real measures in Elementary, Middle, and High School? What advice is best to consider when facing numerous academic standards in a grade?*
    - Tim Brophy: **What are examples of such combinations with real measures in Elementary, Middle, and High School?** Some disclaimers: (1) We do NOT have longitudinal, well-researched examples of combination models in the arts, primarily due to a lack of rigorous, reliable, valid measures that are aligned directly with state standards and benchmarks; (2) Florida's Performing Fine Arts Assessment project is producing item banks that are associated with Next Generation Sunshine State Standards benchmarks. The benchmarks are associated with courses. If it goes as planned, the completed system will draw down interim and summative assessments from these item banks for specific courses, and these results will serve as the source for student achievement scores for determining teacher effectiveness – target date is the 2014-15 academic year

      Florida examples –
      <Hillsborough County (Tampa) – email from music supervisor Melanie Faulkner>:
      We are entering the third year of our new teacher evaluation system called Empowering Effective Teachers based on Charlotte Danielson's Framework for Teaching. Teachers' final evaluations are comprised of 40% student achievement (aka test scores) and 60% observations.

      Student achievement:
      As we have discussed previously, we already have written tests (response questions) that are administered to grades 1 - 5. For the past two years, tests have only had 20 questions which in one sense is not very many, but in another sense is pretty realistic considering

the amount of student contact time. This coming year we will have 25 questions and will adjust schedules to accommodate the additional questions. For Art and Physical Education, time was not really an issue, but it was for Music. The first year, there were 9 - 11 questions, depending on the grade level, with listening (rhythms, melodies, excerpts, instrument sounds, etc.) since we were trying to get as close to performing as we could. Because of time constraints, this past year we went to 2 - 7 listening examples, again depending on the grade level.
It actually is extremely similar to what the FPFAA is trying to accomplish. The tests are based on our district curriculum (which are still revising to align with NGSSS) and vocabulary.

Tests are known as "district tests" rather than end of the year tests since they may be administered mid-way through the second semester. First grade is a pre-measure and not counted toward teacher evaluation. Grades 2 - 5 do count toward teacher evaluation. Since there is no pre-test, they are considered predictor tests.

This past year our tests were online through Achievement Series. We learned a lot! Students were great taking the tests online (Art, Music and PE), however, the teachers were less comfortable. Bandwidth was an issue depending on the time of day and unfortunately, the version our district has does not support audio so a separate CD has to be played for musical examples.

Teachers read the tests and answer choices aloud to every grade level to help meet the need of testing accommodations, especially with our high ELL population. Other accommodations were met through small group testing.

Observations:
Teachers are observed by their school administrators and peer evaluators. Peer evaluators are music teachers who were selected through an application process, trained in the observation rubric, and then calibrated with trainers. The rubric is overall good although administrators often have a challenge in determining how it looks in a music class. Teachers receive a minimum of two formal observations and two informal observations as well as "pop-ins". The number of observations depends on the teacher's ratings. Content Supervisors are called in to observe both formally and informally when needed.

There are four domains in which teachers are rated: Planning and Preparation, The Classroom Environment, Instruction, and Professional Responsibilities. There are 22 components within these domains that may be rated Requires Action, Developing, Accomplished, Exemplary.

Overall, this system has been good for improving instruction, however, there are challenges as well.

<End of Faulkner email>

Polk County (Lakeland), Orange County (Orlando), and Pinellas County (St. Pete) are all using FCAT scores and attributing a portion of the gains to music teachers. Exact formulas are not known. Each district is using an observation tool and the results are combined with the scores for the total evaluation. In all three counties it ranges from 40-60 to 50-50 combinations.

Measuring Teacher Effectiveness report January 2012 – Bill and Melinda Gates Foundation – "Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains". An outside group of researchers are conducting the study – this is a report on year 2. Report on five instruments in their report:

- Framework for Teaching (or FFT, developed by Charlotte Danielson of the Danielson Group),
- classroom assessment scoring system (or cLass, developed by Robert Pianta, Karen La Paro, and Bridget Hamre at the University of Virginia),
- Protocol for Language arts Teaching Observations (or PLaTO, developed by Pam Grossman at Stanford University),
- Mathematical Quality of Instruction (or MQI, developed by Heather Hill of Harvard University), and
- UTeach Teacher Observation Protocol (or UTOP) developed by Michael Marder and Candace Walkington at the University of Texas-Austin).

Compared on two criteria:

- estimated the *reliability* with which trained observers were able to characterize persistent aspects of each teacher's practice, using thousands of hours of lessons collected for this project
- report the association between the observations and a range of different student outcomes: achievement gains on state tests and on other, more cognitively challenging assessments, as well as on student- reported effort and enjoyment while in class.

Findings:

- all five observation instruments were positively associated with student achievement gains.
- Reliably characterizing a teacher's practice requires averaging scores over multiple observations. (they found the highest reliability at four observations)

- Combining observation scores with evidence of student achievement gains and student feed- back improved predictive power and reliability. (The student feedback was collected using the Tripod student perception survey, developed by Ron Ferguson at Harvard University, which was the focus of our last report))
- In contrast to teaching experience and graduate degrees, the combined measure identifies teachers with larger gains on the state tests.
- Teachers with strong performance on the combined measure also performed well on other student outcomes: (I.e, their students reported higher levels of effort and greater enjoyment in class)
- Recommendations for practitioners:
  - First, to achieve acceptable levels of reliability with classroom observations, observers should demonstrate their ability to use the instruments reliably *before* observing
  - Second, to produce reliable feedback on a teachers practice, states and districts will need to observe a teacher during more than one lesson
  - Third, to monitor the reliability of their classroom observations and ensure a fair process, districts and states will need to conduct some observations by impartial observers and compare the impartial scores with the original scores

- Tim Brophy: **What advice is best to consider when facing numerous academic standards in a grade?**
  - Weight the importance of the standards with respect to the curriculum, frequency and number of minutes of instruction
  - Align assessments with the weighted standards so that the assessment data obtained best matches the weights
- Laura Goe: Different levels of reliability will exist with multiple observations and multiple raters.
- Bob Good: Opportunity to learn will garner a great deal of interest. We need to ask teachers -- where are we now in curricular delivery against the standards, and where do we need to be in 2-3 years? With that in mind, what would the minimum representation be in terms of assessment in order to support that goal? RtI needs to stop poaching students out of PE, music, and other classes. Many of these teachers are fearful of their programs being eliminated because the evaluation structure doesn't allow enough time for learning and evaluation to take place.
- Margie Ruckstuhl: In Harrison, there are specific templates for specific content areas. It has validated all teachers.

- Toby King: Leadership at the district and school level is responsible for making sure that the master schedule allows teachers and students to be supported.
- Tim Brophy: Ask ourselves what we can expect our teachers to teach in the time they have been given and hold them accountable to that. There is also a complication when considering outside instruction. Many students in the arts, for example, benefit from private lessons.
- Toby King: Shared attribution can provide motivation for all teachers to support improvement.
- Kristen Huff: Regarding number of raters function – time to train evaluators is a challenge. The Tennessee model is a good example.
- Laura Goe: The main challenge has been time.
- Tim Brophy: In Tennessee there are Race to the Top funds supporting this initiative. They are currently working on a sustainability plan. Teacher preparation programs are currently training their teachers for this new world.
- Margie Ruckstuhl: Harrison uses professional development days for scoring, and teachers are given only as many assessments to review as students they currently have. Double-blind scoring at the secondary level is unmanageable because the numbers are too high.
- Jo O'Brien: Student perception data? What are thoughts?
- Tim Brophy: This data is focused on instructional improvement. It is an indirect measure.
- Patrick Mount: Student perceptions don't change much after the first 10 days.
- Toby King: In Colorado student perception data is calculated in professional practice, not student growth.
- Sed Keller: Student perception data has a high correlation to student growth.
- David Webb: In the university system, students evaluate professors and those evaluations are part of the professor's re-appointment or tenure.
- Tim Brophy: This conversation is a shift from teaching centers to learning centers. The focus needs to be on students, as opposed the details of the process or assessments.

Wrap-up and next steps

- Next steps:
  - CDE to assemble practical guidelines for districts
  - CDE to construct examples of how to plot growth
  - TSC will receive the information in advance, plan a virtual meeting to discuss drafts

    - CDE will work with represented and other identified districts about what is currently being used, and create some possibilities for the TSC to react to
    - CDE will send out a draft of the glossary of terms for use in the guidelines
- Next in-person meeting: Wednesday, December 12, 2012.