CDE | COLORADO
Department of Education

Meeting Minutes – Thursday, March 24th, 2016 (9:00-noon)

201 East Colfax, State Boardroom Denver, Colorado

TAP Members:

Norman Alerta
Carol Eaton (call-in)
Grant Guyer (call-in)
Jacqueline Law (call-in)
JulieMarie Shepherd (call-in)
Elena Diaz-Bilello
Grant Guyer (call-in)
Linda Barker (call-in)

Audience:

Adam Van Iwaarden (NCIEA)
Damian Betebenner (NCIEA)

CDE Representatives:

Marie Huchton
Dan Jorgensen
Alyssa Pearson
Josh Perdue
Kelly Stritzinger
Joyce Zurkowski

Welcome

Norman Alerta/Dan Jorgensen

- The meeting was held in the State Boardroom at 201 East Colfax Avenue and was called to order at 9:14.

- Norm Alerta announced that we have two new TAP members, Elena Diaz-Bilello and JulieMarie Shepherd (CASB Representative).

'Transitional' Growth Percentile Discussion

Marie Huchton/Damian Betebenner

- An overview of growth trajectories that impact growth percentile calculations was presented. Also, the new calculation methodology based on going from reading/writing to combined ELA scores was explained. The utilized regression model included additional prior scores to account for the assessment change and to maximize the use of pertinent data. For example, both reading and writing will be included from prior years to calculate English language arts growth. The multiple options for High School math (prior year trajectories) are a challenge for calculating growth. Each trajectory has to be calculated during this initial year creating many computational scenarios.

- The score distributions for TCAP and PARCC were examined. Also, correlations were calculated between years from TCAP 2013 to TCAP 2014 and TCAP 2014 to PARCC 2015. In both cases, obtained correlations were around or exceeding 0.8 across content areas. These strong correlations argue for convergent validity of the assessments. The ELA scores were correlated to reading and writing separately and the many math end-of-course tests introduced some problems with restricted sampling, but the correlations for grades 9 and10 still exceeded 0.6

0. TAP members reported not being surprised by the modest decline in correlations as the assessments become more subject specific as you progress. For the 2016 growth percentile calculation conversation, the reduction in r-values along with the nuanced content may result in a substantive impact. CDE will need to decide whether to just use prior year PARCC scores or to also include prior years of TCAP data. It was mentioned that it is important to look at range restriction as group self-selection to different assessment trajectories likely exists. In effect, we may be truncating a distribution that previously was related to a higher correlation.

- Participation concerns exist for assessments. For 8th grade and lower participation exceeded 85% but rates were under 80% for high school grades. This signals a need to consider representativeness of results depending on which level of results we are dealing with in our analysis. Since we will no longer have 10th and 11th grade PARCC it should help reduce the impact of future challenges at this level. It was asked if non-random refusal at the secondary level creates a less representative population. CDE reported that this appears to be true for high school grades, however, for 3rd to 8th grades it is only at the local level that non-random opt out patterns might be an issue (i.e. not at the state level). Per the Assessment Office at CDE, given this situation we can be confident in the state level results but not the local level results (i.e. except at the high school level where we still see significant differences in key demographic variables). It was pointed out that it's important for some schools with higher achievement to recognize that the accuracy of their results may be limited as some comparison schools are no longer available for comparison.

- The between year correlation coefficients are slightly lower than what we saw in the past but are still quite strong. Based on conducted analysis, it was expressed that CDE believes that the scale scores are meaningfully related enough to allow for growth calculations. This concluded CDE's growth presentation, the floor was opened for feedback from the group:
    o   CDE expressed that we are now preparing spring reports, 3rd to 9th grade results only. These reports won't be provided for accountability determinations just for informational purposes.
    o   It was asked what other analysis did CDE conduct, demographics? FRL? Highly impacted locations?
        ▪   It was expressed that CDE staff have been time limited but will continue to examine possibilities.
        ▪   It was pointed out that the strength of correlations indicates some consistency. Per TAP members, if we are showing bigger differences in subgroup performance it may raise potential calculation limitations.
        ▪   It was expressed that the first year of transition is going to be messy; even if we say test content is similar the implementation of standards will differ. Other members pointed out that local level analytics may help support work; at least the high level demographics could be helpful. In effect, it is believed that some questions need to be answered at the local level.
        ▪   CDE pointed out that a decision has to be made about sun setting prior assessment results.
        ▪   Dr. Betebenner pointed out that most states use more than one prior when available, immediately dropping to one prior would be contrary to what's typical. It can be done this way but its likely measurement error will have a more profound effect. Trade-offs with both approaches. Most states want to sunset after three years of PARCC data.

- The SGP calculation/distribution process was discussed. The preferred distribution is a 'bathtub' shape. It is to be recognized that when the n-size is small the distribution is more problematic and less representative. This raises a need for a minimum n-count especially for less common math pathways. For example, in the case of 7th grade Algebra I the model wouldn't run (i.e. with 90% of students previously scoring advanced). In order to deal with technical complications, a minimum n of 2000 was adopted to calculate SGPs in any content area (see highlighted cells in presentation). This means not all students with two consecutive years/scores will receive growth percentiles. It is expected that with no 10th -11th grade scores the impact of the many pathways will be lessened.

- Floor and ceiling considerations were discussed. A buildup of scores at the lowest obtainable scale score (LOSS) with some clustering at the highest obtainable scale score (HOSS). Theta values were looked at as a possible modification to calculate SGPs. It may help but cases continue to exist with ceiling effects. Thus, a decision was

reached to continue calculating SGP from scale scores. However, default all cases in which students earn the HOSS to an SGP of 99. It was raised that CDE also needs to consider broader score bands (i.e. as opposed to just the HOSS) that may be impacted. <u>Marie agreed to examine other anomalous trajectories to see if they have concerning results.</u> It was suggested that she start with students in level five to see if there are other impacted cases.

- Dr. Betebenner pointed out that if we apply the minimum n-size of 2,000, we're not saying the model is wrong; instead it's not producing a model that's easily understood by stakeholders (i.e. parents). The state needs to identify the conditions where the model doesn't seem to produce a result for a student that is meaningfully interpretable at the student level. CDE can then apply these logic/recommendations for not providing growth model data. If the student is hitting the floor year after year we don't need to say how much they're growing, similarly, for the students at the top. We really don't need to know more in means of growth. The test isn't sensitive enough at those ends. The question was raised what would we do for students at the LOSS? One possibility would be to default these cases to an SGP of 1.

- Per CDE, the use of 99 was shown to have a minimal impact. Discussed aggregating school MGP and applied rules. Cross-year school MGP correlations were calculated to determine between year trends in growth. Correlations dropped for reading from .506 (TCAP to TCAP) to .427 (TCAP to PARCC). Still reasonably moderate to strong correlations observed. High schools showed the biggest drops. Participation impacted high schools the most but only modestly for E/M.

- Test mode differences were discussed as people have asked about the comparability of scores. Colorado mandated on-line testing in ELA so it minimized the potential impact. However, for next year a paper pencil option is required. It was asked if could put a 1/0 dummy variable to control for differences in the model itself. Dr. Betebenner said this is a viable option. Another option was to norm paper based assessment results against themselves; however, less than 2,000 students took the test so would fail to meet minimum n-size criteria. A concern was raised with the dummy coding option in that it may lead to additional requests for changes to the model.

- It was expressed that SGPs and results tend to appear representative across most content areas with some caveats for high school. When we have less than 85% participation we should interpret the results with caution.

The various options for releasing transitional growth data were discussed and included:
1. Do not release anything
2. Release student level growth data to requesting districts that ask
3. Released student level growth data to all districts
4. Release school, district, and aggregate growth data publically
5. Include growth data in spring informational performance frameworks
6. Do not include as part of fall 2016 official SPF/DPF results, but allow submission of data in request for reconsideration.
7. Include 14-15 growth data as part of 2-year calculations for fall 2016 official SPF/DPF results

- It was said that release would require strong communication to support it. Members were ok with potentially releasing student level growth data to requesting districts. It was expressed that it may have limited value at this point but could be helpful for planning. Concerns with including in the fall 2016 report but possibly keep it for appeal purposes/reconsideration.

- It was asked, do we want to include 'transitional growth' in multi-year framework? It was stated that if the state was producing both one year and two year and giving school/district the higher of two then you mitigate the risk of limited data.

- CDE expressed that the spring reports would be informational only with no ratings or points. CDE wouldn't have a media release. The release was discussed as having pros- and cons- but may help Principals from being blindsided in the fall. A recommendation was to release the reports with a guidance document that included 'use with caution' along with caveats to the data. This would allow for principals to be somewhat prepared for any changes.

- TAP members expressed concern with the timing of the release of spring reports. However, members expressed that they would like to see spring reports with growth information. Also, student level growth data was requested as soon as possible. For the student growth data files, CDE asked if we notified District Assessment Coordinator's and Superintendents that it's available would that work for everyone. We could then release on request. Everyone responded yes with the caveat that we need to include an explanation of the data and that it's not for 'official' purposes. CDE will continue to explore the best process for release.

- CDE asked for verification that default to 99 is ok and then we will release. We will also examine other bands that may have ceiling effect. <u>Prior to release we will first have an internal TAP release for additional feedback. The goal is to release the files to TAP members via district assessment coordinators early next week. Once we receive and review any additional feedback we will decide how it will be released to schools.</u>

Performance Framework Redesign Considerations
Marie Huchton/Alyssa Pearson

- Requested Feedback concerning a number of questions:
  - What is the TAP's recommendation for the official name of the super-subgroup?
    - ACEE recommended, combined subgroups; TAP members, no matter the name will need to explain that there's no double-counting.
    - CDE said they will request feedback from the field when the spring report is released.
  - How should READ Act Bonus Point targets be operationalized based on new state assessment?
    - Based on data, at 50th percentile there's only two schools that meet target. At 15th percentile only 10 schools meet the target.
    - It was expressed that the decision should be criterion referenced, such as movement to level two. A mean scale score cut point would need to be based on some reading ability as opposed to an arbitrary distribution decision. We would think about individual kids and what's meaningful in a criterion referenced manner. The recommendation is being to first set up student success criteria to earn credit for a school. Each student on a READ plan either met or didn't meet criteria. Next, we would look at proportion among eligible groups between schools and then use that distribution to set cut points for schools. It was expressed that using the performance levels of the assessment seems appropriate.
  - Is subgroup information concerning catch-up growth useful?
    - It was expressed that it is helpful for planning purposes and as part of teacher evaluation metric in some cases. However, not deemed good for points on performance frameworks.
    - <u>CDE will have it available and then solicit feedback as part of the spring reports.</u>
  - Should we keep participation rates that exclude parent refusals from calculations but continue to hold schools and districts accountable for other non-participants?
    - About 55,000 students didn't test but weren't parent refusal. A number of districts/schools wouldn't make participation rates based on these codes.
    - The possibility of using participation rates as a carrot and not a stick was discussed. In effect, this could be initial criteria for distinction rating. It was expressed that the state board has created an issue in that we can no longer convince community to take tests. It is harder to get participation. CDE response is that the state board position is about parent refusals.
    - CDE anticipates that once the coding issues are corrected it probably won't be as much of an issue as its thought that many of these are parent opt outs.

- o CDE requested the thoughts/considerations of the TAP regarding setting sub-indicator targets that they would like brought to the State Board of Education.
    - CDE has heard strongly from the field that a single set of norms is preferred. This was recognized as being contrary to the expressed position of some TAP members.
    - It was requested that TAP members e-mail any additional thoughts they have on this topic for consideration at the state board meeting on the 14th. We would also like any feedback concerning interest in maintaining previously recommended indicator ratings or change.

- The formal sharing of information was concluded with a TAP member mentioning that we need to continue to monitor the paper vs online issue as districts may push towards all-paper testing especially if they think it's beneficial.

Public Comments, Action Items, & Adjourn Meeting

Norman Alerta/Dan Jorgensen

- The opportunity for public comment was provided with no attendees choosing to speak. The next TAP meeting is currently scheduled for May 19th, 9-noon with the location to be determined.

Meeting Adjourned at 12:30