Illinois Assessment and Accountability Technical Advisory Committee (TAC)

DRAFT Meeting Summary and Action Items
December 11-12, 2019
University of Illinois, Chicago
1240 W. Harrison St., Suite 1535A

UPCOMING MEETINGS

February 27-28th (Full day on the 27th, half day on the 28th)

June 4-5th (Full day on the 4th, half day on the 5th)

INTRODUCTORY NOTE

This document presents a summary of the key discussion points from the IL TAC meeting held in Chicago on December 11-12. This document follows the written agenda, unless the topics were addressed in a different order during the meeting. In that case this meeting summary follows the order of discussion.

This summary does not present verbatim minutes and, in general, avoids attribution of comments to specific individuals. The meeting summary is intended to document the salient discussion points which are often the result of input from multiple individuals. Broad TAC-level recommendations, with a focus on recommended analyses, are presented when such recommendations emerge.

Please note that the complete agenda for this meeting and all materials provided to the TAC for review can be found on ISBE's website.

ATTENDEES

ISBE: Rae Clementz, Shu-Ren, Brenda Dixon, Allison Sherman
TAC Members: Laura Hamilton, Mike Russell, Jim Pellegrino, Jeff Broom, Diana Zaleski, Ellen Forte
Center for Assessment: Erika Landl, Chris Domaleski, Damian Betebenner
New Meridian: Tracy Gardner
Pearson: Aimee Boyd

ILLINOIS ASSESSMENT OF READINESS (IAR) – COMPARABILITY

Topic 1: Transition from Paper to Online

Damian Betebenner from the Center for Assessment provided an overview of the procedures used to calculate growth in IL and a summary of the results observed in 2018-2019. Growth by achievement plots were presented to illustrate projected performance for students starting at different achievement percentiles in Grade 3 and the assumption of constant growth across grades (i.e., at the 10th, 25th, 50th, 65th or 90th percentile). The TAC highlighted the prominent dip in achievement in grade 6 for both ELA and Math and noted that students who begin far behind in Grade 3 need to make significant growth (far above average) to achieve proficiency by Grade 8. This was the case in both content areas, but more so

in mathematics for which greater than average growth (i.e., the 50th percentile) was necessary to maintain a classification of proficiency for students starting out at/near the proficient cut in grade 3.

The TAC discussed the potential impact and benefit of baselining growth (i.e., so that gains relative to a stable baseline year could be tracked) but decided that additional information was necessary to evaluate whether this option should be explored. Specifically, the TAC recommended that actual vs projected achievement be compared for students starting at different achievement percentiles as a way of evaluating whether growth might mean something different from year to year. If the residuals are large, or trajectories calculated in the same manner look extremely different from year to year, baselining might be something to consider. Additional evidence will be brought to future TAC meetings to support these discussions.

Subsequently, Rae Clementz described the procedures used to ensure schools transitioning from paper to online administrations in 2019 were held harmless for purposes of accountability. She indicated that system and indicator-level scores were reported as calculated, but transitioning schools that would have earned a lower rating in 2019 compared to 2018 were assigned their 2018 performance classification.

The TAC did not have any issues with this procedure and agreed it is appropriate to continue this practice moving forward given the small percentage of schools left to transition (less than 5) and the incentive to transition afforded by this policy.

Topic 2: ELA Growth Evaluation

Within each grade, there are two blueprints that support the development of ELA test forms. The blueprints are identical except for one of the writing tasks a student receives - literary vs. narrative. At a previous TAC meeting the use of two blueprints raised concerns about comparability, specifically as it relates to estimates of student growth. For this discussion, data was presented summarizing the difference between the mean student growth percentiles (MGP) associated with each form. For each grade, results were provided for the total population as well as each sub-group. Results showed no differences of practical significance (e.g., MGP of 5 or greater) for any grades or sub-groups.

Consistent with previous recommendations (see the report from June of 2019) the TAC suggested that future analyses be conducted to compare differences in MGPs (i.e., 2020 minus 2019) for students taking different sets of forms across grades (e.g., Form 1/Form 1; Form 1/Form2; Form2/Form1). This would provide information about the degree to which growth over grades is differentially influenced by the pair of forms a student receives.

ILLINOIS ASSESSMENT OF READINESS (IAR) – MATH ADMINISTRATION TIMES

Aimee Boyd from Pearson presented the TAC with a variety of analyses focused on identifying patterns in the math administration time data. To support these analyses, students within each unit were categorized into one of 4 groups based on their location in the administration time distribution (0-75th, 76th-87th, 88th-92nd and 93rd-100th percentile). For each time percentile group and unit these analyses included the evaluation of:

- average scaled scores;
- demographic representation;
- distribution of time spent on items associated with different sub-claims and total points possible;
- mean time spent on items by sequence on the test (e.g., before last 5, last five, last item); and
- average omit rates.

Rae Clementz reminded the TAC that these analyses were spurred by concerns from some stakeholders that high performing students did not have sufficient time to perform as well on the test as they could have if they didn't feel rushed. ISBE is trying to determine if sufficient data exist to suggest a change in the time allotted for each unit or if the state's policy related to timing should be altered (e.g., allowing students who are deeply engaged to have an additional 5-7 minutes). While the current policy limits test taking time to 60 minutes per unit, the system does not prohibit ongoing work after time expires. However, ISBE emphasized that the data the TAC reviewed was not the final data set used for reporting and does not reflect whether administrations were later invalidated.

The TAC reviewed and discussed these analyses in turn, noting the difficulty finding consistent patterns across units, grades and student groups. That being said, the TAC did identify a few relatively consistent trends:

- In general, average scaled scores did not differ dramatically across the percentile score groups.
- Students with language barriers tended to take more time than students without these barriers.
- Students spent more time on sub-claims and items associated with constructed response items than they did on other types of items.
- The administration time for some examinees exceeds the 60-minute limit defined in policy.

In the end the TAC did not believe that the current data provided enough information to suggest changes should be made to the state's existing administration time policy. They recommended that additional analyses be conducted to inform this decision, specifically those addressing the following questions:

- To what extent is a students' time percentile group classification stable over units?
- To what extent are the students that do not finish (e.g., omitting the last 1 or 5 items) the same across units?
- What does the distribution of student performance look like within each time percentile group (with respect to scaled score distribution and representation of each performance level)? It is possible that some of these distributions are bimodal?
- What are the characteristics of students in the highest timing percentile group? Does it tend to be the same students? What are their demographic characteristics – do they tend to be white and high ability? To what extent are these students isolated in particular schools?
- What are the typical omit rates associated with constructed response items on a high-stakes summative assessment?

To address these questions and more effectively interpret the results the TAC suggested:

- including a ratio reflecting the performance of each percentile group relative to that of a fixed reference group (all students), and
- creating stacked bar charts which indicate the representation of each performance level within a given time percentile group.

The TAC also discussed the potential negative implications associated with lengthening the administration, including increased costs and logistical/procedural changes that would be difficult for all, but especially under-resourced, schools to address.

ILLINOIS SCIENCE ASSESSMENT UPDATE

Rae Clementz provided the TAC with background information about the current IL Science assessment (ISA). The test was first administered in 2015, the year IL adopted the NGSS, but results for 2015 and 2016 were not reported until February of 2017.

The tentative plan for the 2020 Science assessments in grades 5 and 8 is to use a combination of items from the current IL bank, items provided by Oklahoma and new items developed by IL educators. Plans for the Grade 11 assessment are not clear as the construct represented by the new science blueprint is significantly different from that tested by the existing ISA. The United States Department of Education (ED) may provide a waiver allowing for a census field-test in grade 11 for 2020 in lieu of an operational test, but this assumes new items will be available.

The TAC reviewed the Science blueprints and reiterated the importance of taking adequate time to design, pilot and administer test items that reflect the state's vision for Science. They cautioned that a quickly developed, transitional assessment of the type described might not well represent the quality characteristics and measurement targets intended by the new standards. This might send an unintended message to the field about the state's priorities for Science.

The TAC expressed an interest in learning more about the current item development process as well as the specifications and reporting structure for the 2020 assessment. Importantly, ISBE should develop a more detailed blueprint and test specifications before going too far down the road with item development and/or acquisition. Quality item development requires a clear understanding of the claims the results are intended to support, so the TAC wondered how/if claims and PLDs were being considered throughout the development process. To answer this and other questions, ISBE recommended that those leading the content development effort – from ISBE and Southern IL University – present to the TAC at its February meeting.

LISTENING TOUR UPDATE

Rae Clementz provided the TAC with an overview of key questions and recommendations that surfaced during her state listening tour. In general, stakeholder comments were in 4 key areas:

- Weighting of academic and non-academic components (75/25) and specific indicators
- Establishing a measure of growth in high school now that the state has assessments in grades 9-11.
- Labels associated with school designations (i.e., low performing and underperforming)
- Policies related to assessment of all English Learners – 1st year English Learners had to take all tests, even though they were not counted in accountability determinations

Rae also indicated that stakeholders were particularly passionate about the inclusion/exclusion of chronic absenteeism as an indicator within the system. While signs suggest that the system will not change significantly in the near future, due to the need for additional data, the lower two tiers of the system will be renamed to Targeted and Comprehensive Support.

ACCOUNTABILITY SYSTEM CONSIDERATIONS

Topic 1: Evaluation of Procedures Used to Score Aggregate Growth

Chris Domaleski revisited the procedures used to establish an index score for growth on the 0-100 scale, and presented 2018-2019 growth index score distributions. Using the current procedures, schools with MGPs at or below 28 receive a growth index score of 0 and schools at or above 73 receive a score of 100. MGPs from 29 to 72 are linearly interpolated to assign a score from 1 to 99.

The TAC reviewed and discussed the average growth index scores associated with schools at different percentiles of the index score distribution. In general the TAC did not see a problem with the current scoring procedure as it serves to spread out schools and incentivize growth. Consequently, no modifications to the procedures (i.e., effective range) used to score growth were recommended at this time.

The TAC reiterated that an evaluation of projected vs actual growth over time should be considered to determine whether the distribution of performance is changing in a way that suggests procedures for baselining growth should be explored. The TAC also suggested that growth distributions be examined for selected student groups (e.g., English learners, IEP, and economic disadvantage). These analyses will be presented at a future TAC meeting.

Topic 2: Chronic Absenteeism

Erika Landl revisited IL's definition of chronic absenteeism (per state law), discussed how it is calculated within the state's accountability system and presented data summarizing the performance of IL schools on this indicator. Due to the broad range of factors that influence absenteeism (e.g., poverty, safety, family circumstances, weather) many schools perceive this indicator as being out of their control. To address this concern the TAC suggested that ISBE articulate its theory of action for this indicator and develop/promote an early warning system that informs local improvement efforts around chronic absenteeism. Training around chronic absenteeism provided through the Education Leader's Network (ELN) could also be beneficial.

Topic 3: Impact of Alternative Schools in the Traditional Accountability System

Chris Domaleski provided data summarizing the performance and impact of 18 alternative schools currently included in IL's accountability system. All 18 schools were identified in the bottom 5% of schools in the state (they comprised 32% of the bottom 5% of schools) and had an average overall index score of 3.65 with a range of 1-8. Due in part to the low performance of these schools, the threshold associated with the bottom 5% of schools (i.e., those classified in Tier 4) was 8.17. If these schools were not included the threshold would have increased to 12.

To understand what, if any business rules should be put in place related to the inclusion of alternative schools the TAC suggested that ISBE explore the implications of removing them from the bottom 5% of schools. Specifically, this analysis would determine the number of additional schools that would now be flagged as having a low performing sub group if the threshold associated with the bottom 5% increased. This data will be presented at a future meeting in combination with information about:

- the design of other states' alternative school accountability models; and
- any constraints defined within ESSA and IL state law related to the development of an alternative school accountability model.

Once again, the TAC stressed the importance of articulating a theory of action for the overall accountability system and each indicator so the rationale and hypothesized mechanism for change was clear. This type of documentation, in combination with increased teacher engagement and assessment literacy, is necessary to help teachers understand the value of each element of the system and how the results are intended to be used.

TAC DISCUSSION AND PLANNING

At the end of Day 2, each TAC member was asked to indicate the biggest priorities for future discussion and TAC consideration. The TAC members agreed that supporting the development of a high quality Science assessment should be a high priority for future meetings. Other suggestions were focused on supporting the state in helping districts/school effectively use and interpret data by having additional discussion focused on: the pros and cons of existing information systems (e.g. ,Ed 360), how/if schools can impact rates of chronic absenteeism, and the theory of action underlying the design of IL's accountability system and components.

Other topics proposed for consideration included: clarification of the procedures and criteria used to scale and equate the IAR, factors that influence school climate and culture and the definition and the potential impact of IL's proposed school quality and student success (SQSS) indicators (e.g., Fine Arts, CCR) in general and compared to other potential SQSS measures (e.g., social emotional learning, educator wellness).