Illinois Accountability Technical Advisory Committee (TAC)
WebEx Meeting Notes
January 23, 2018
(9:00am- 12:00pm)

Attendees:
ISBE: Jason Helfer, Rae Clementz
TAC Members: Laura Hamilton, Mike Russell, Jim Pellegrino, Jeff Broom, Diana Zaleski, David Conley
Center for Assessment: Erika Landl, Chris Domaleski

Introduction
After role call the Center for Assessment provided the TAC with an opportunity to weigh in on the proposed-final version of the December 2017 TAC meeting summary. The TAC had no further edits; therefore, the status of this document will be changed from proposed-final to final.

Chris Domaleski reiterated the purpose of the IL Accountability TAC before summarizing the proposed agenda. During the first part of the meeting the Center intended to describe and compare results from two approaches to measuring growth – regression and value tables. The second half of the meeting would be focused on collecting TAC feedback related to the results observed and recommended next steps. Chris clarified that additional analyses based on PARCC SGPs would be conducted prior to the next face-to-face TAC meeting scheduled for March 6-7, 2018.

To kick off the growth discussion, Chris reiterated some of the key factors that should be considered when selecting a growth model for use within an accountability system, including: data and technical requirements, stakeholder priorities, and the growth inference the measure is intended to support (e.g., descriptive, predictive or attributional). He also summarized the TAC's December recommendation as to the characteristics that should be valued when evaluating different approaches to measuring growth. Specifically the state should value approaches that:

- are relatively straightforward to understand and implement
- teachers perceive as something they can directly influence
- minimize school level instability due to n-size
- minimize correlations with prior year status

- demonstrate availability of the full distribution of growth outcomes to schools of various demographic compositions (e.g., poverty, SWD, and ELL)
- are sensitive to changes in student achievement, particularly for students at the low end of the test-score distribution
- minimize ceiling and floor effects
- are reliable (i.e., provide for stable results across years in cases where the underlying performance of a school is not changing)
- minimize punitive aspects
- detect (not mask) important school level effects
- are robust to changes in state assessment and differences in test characteristics

Regression

After an overview of the procedures and data used to generate regression-based growth measures, results associated with the application of this approach were presented. Analyses were conducted using 3-years of PARCC data from a cohort of IL students who were in Grade 3 in 2015. These data provided for the calculation of growth measures in grades 4 and 5 for Math and ELA. For each student, standardized residuals transformed to T-scores with a mean of 50 and SD of 10 formed the basis of the growth estimates. Subsequently, average T-scores were calculated and reported for each school having an N-count of 20 students or greater. School level analyses included descriptive statistics and correlations between average school growth and factors such as the percent of students achieving proficiency in the prior year, percentage of students with disabilities, percentage of English learners, and percentage of students eligible for free and reduced lunch.

The TAC indicated that correlations between PARCC scores for adjacent grades would be helpful to support the interpretation of results, so these analyses were quickly produced over the break. Within each content area, correlations between PARCC scores for adjacent grades were around 0.80, indicating that prior year performance accounted for approximately 65% of the variance observed in student level scores.

Based on the review of all analyses, one TAC member suggested that extreme student cases should be evaluated to determine the extent to which they clustered within particular schools and potentially influenced regression results. Other recommendations included setting a minimum threshold around EL representation for the purposes of exploratory analyses (e.g., must have at least 5% of students be EL to include in school-level analyses) and better illustrating how between year

growth correlations vary for different sized schools (e.g., representing schools of different sizes with different colors in the scatter plot).

The TAC also pointed out the relatively narrow distribution of school growth scores resulting from the use of the regression approach, which could explain, in part, why school-level growth score were uncorrelated across years.

Value Tables

Erika Landl provided the TAC with an overview of the process used to extend IL's ISAT value table so it could be applied to IL's PARCC results. The PARCC assessment is reported in terms of 5 levels whereas the ISAT only had 4, so two additional levels (5a and 5b) were added to support the calculation of growth. Within the value table, change in performance level from Year 1 to Year 2 is associated with a given score. Consistent with the ISAT version, the PARCC value table demonstrated the following characteristics:

- Maintaining proficiency (4a) for 2 years results in a value of 100.
- Maintaining performance at higher levels across two years is associated with higher points than maintaining performance at lower levels.
- Any movement upward from Year 1 to Year 2 results in a value greater than 100.
- Any movement downward from Year 1 to Year 2 (i.e., 5b to 5a) results in a value of 100 or less.

The same descriptive statistics and analytical results presented in the regression discussion were displayed for the value table approach. Results across the two approaches were extremely similar, so recommendations related to the calculation and presentation of results were essentially the same. The TAC acknowledged the fact that, for students at the extreme ends of the test score scale, regression to the mean and measurement error may influence the extent to which growth measures represent a true gain or loss in proficiency.

One TAC member noted that school growth scores generated using the value table approach exhibited more dispersion than those observed under the regression model providing for greater differentiation between schools. This did not, however, result in improved stability of school-based growth results. Correlations between years were essentially 0.

Comparison of Models

After reporting results for each approach separately, Chris presented a side-by-side summary of results across the two approaches, which were extremely similar.

Schools tended to be rank ordered in the way based on value table and regression results, and the technical characteristics of the resulting growth measures were extremely similar. That is, across the two growth measures (value-table and regression) correlations between average growth and school factors such as English Learner (EL) and Students with Disabilities (SWD) representation, poverty rate and prior year proficiency rate were in the same direction and of a similar magnitude.

The TAC indicated that there were different ways the approaches could be specified, such as including additional predictors in the regression model and/or changing scores within the value table, which could influence the consistency of results across models but, based on the current analyses most schools were rank ordered in a similar manner.

The last set of analyses focused on those schools demonstrating the largest change in relative growth standing when evaluated across the two models. Specifically, schools were flagged that demonstrated a z-score difference of .333 or greater across the two models for Grade 5 Math. The results of this exploratory analysis suggest that schools having high poverty rates and larger SWD/EL representation tend to earn higher growth scores under the regression approach; however this approach also favored schools having small N-counts, which would be most affected by measurement error due to pronounced variability across years.

Discussion

Overall the TAC thought the growth analyses provided useful, compelling exploratory information about each approach. However, the TAC members reiterated the importance of identifying a growth measure that is not just technically sound but serves to identify those schools IL is most concerned about. Since growth is weighed so heavily in the overall accountability model (50%) school growth measures will play a large role in determining which schools are ultimately flagged for support.

The TAC also suggested conducting sensitivity studies which demonstrate how changes to regression and value table specifications influence the consistency of results. For example, the regression approach could be re-specified to include additional independent variables (e.g., multiple years of PARCC test scores). Similarly, the value table could be modified to:

- reflect a "neutral-gain" model, whereby the same change in performance across levels always results in a consistent gain/loss regardless of location on the scale

- reflect different state priorities with respect to the type of change considered most important (e.g., assign the highest values to cells that reflect significant upward movement at the bottom of the score scale)
- account for characteristics of schools that tend to bias results (e.g., What does it look like for a high EL/high poverty/low proficiency school to perform well with respect to growth, and how can the model be designed to acknowledge that?)

One TAC member also stressed the importance of evaluating student attribution and the impact of mobility on school growth measures. One way to do this might be to look at the results associated with schools that only meet minimum N if the full academic year (FAY) criteria are lifted.

Next Steps

The Center will evaluate school growth data using PARCC SGPs and evaluate the impact of different regression and value table model specifications.

Future Meetings

March 6-7: 8:30-4:30 on the 6th; 8:30-12:00CT on the 7th
April 2-3: 12:00-4:30 on the 2nd; 8:30-4:30 on the 3rd
April 30: 12:00-3:00CT (WebEx)