Linking BEACON and Georgia Milestones

In 2020-2021, a linking study was conducted to examine the predictive relationship between DRC BEACON and the Georgia Milestones End-of-Grade assessments in grades 3 through 8 in English Language Arts and Mathematics. As an optional interim assessment, BEACON was designed to be administered throughout the academic year to provide on-demand formative information to support data-driven instruction. However, BEACON assessments can also provide evidence that indicates whether students are on track to meet end-of-year goals as measured by Georgia Milestones. Predictions, or forecasts, of students' expected performance levels on Georgia Milestones, conditioned on students' BEACON performance, were produced and evaluated in this study.

Linking Study Methodology

Data for the linking study were gathered from four administrations: 1) Beacon Fall 2020 (August – November), 2) Beacon Winter 2020 (December – February), 3) Beacon Spring 2021, (March – May) and 4) Georgia Milestones Spring 2021. Student data from all four administrations were matched and resampled such that the data used for linking were representative of the state in terms of achievement, gender, ethnicity, SWD, locale and RESA before the linking analysis.

Three linking methods were implemented and evaluated for technical stability and feasibility to establish the predictive relationship between BEACON and Georgia Milestones for fall, winter, and spring test administrations: linear regression, isotonic regression and equipercentile linking.

- Linear regression (Pedhazur, 1997) involves the estimation of the unique line that best fits a scatter plot of points based on paired observations of the predictor and its dependent criterion variable.

- Isotonic regression (Barlow, Bartholomew, Bremner, & Brunk, 1972), also known as monotonic regression, is the technique of fitting a free-form line to a sequence of observations such that the fitted line is non-decreasing (or non-increasing) everywhere, and lies as close to the observations as possible.

- Equipercentile Linking (Kolen & Brennan, 2004) defines two scores on test X and Y as equivalent if their corresponding percentile ranks or cumulative frequency distributions in any given group are equal.

The three methodologies for linking were evaluated before data collection by Georgia's Assessment Technical Advisory Committee (TAC) and approved as appropriate options to pursue in producing projected end-of-year Georgia Milestones achievement based on through-year BEACON performance. In evaluating the theoretical and technical rationales for the three approaches, some benefits to each methodology were considered. Linear regression is a common approach to prediction in educational settings, and assumes that there is a strong linear relationship between scores on the predictor variable and its dependent criterion variable. However, formative interim assessments like BEACON may not have a strong or consistent linear relationship with the Georgia Milestones summative assessments due to the differing purposes of formative and summative assessments. While formative assessments like BEACON administered throughout the year are low-stakes opportunities to inform ongoing instruction, state summative assessments inform policy, evaluation, and accountability processes. This differential purpose and student experience may introduce a weaker linear relationship or a non-linear relationship between interim assessment and Georgia Milestones scores, which suggests that the isotonic regression and equipercentile approaches may be more effective than linear methods.

Analysis and Outcomes

Links were established for all three of the methods outlined above using the 2020-2021 BEACON and Georgia Milestones student data. To evaluate the different prediction methods, the predicted Georgia Milestones student scale scores and achievement levels were compared with observed Georgia Milestones student scale scores and achievement levels. Note that perfect correlations and exact classification agreements were not expected here; as noted prior, the BEACON formative assessment system and Georgia Milestones summative assessment system have different purposes, and the goal of this analysis was to evaluate whether the links modeled indicated a meaningful strength of agreement between the projected and observed achievement. While no attempted approach failed to produce a stable link and good strength of agreement between projected and observed achievement, the isotonic regression approach yielded the most accurate results. The isotonic regression link produced the highest correlation (correlation range: .80-.92) between predicted scale scores and observed scale scores, and the highest classification consistency coefficients (exact/adjacent agreement range: 98%-99%) across all grades/content areas, by season.

Results from all three approaches were reviewed by Georgia's Assessment TAC. Based on the strength of agreement and technical and theoretical rationale, the isotonic regression approach was recommended as the most appropriate linking method to connect BEACON performance to projected Georgia Milestones achievement. Specifically, the probability of being in each Georgia Milestones achievement level given student performance on a corresponding BEACON assessment will be computed for each season. The achievement level with the highest probability is reported on BEACON individual student reports when full BEACON tests are administered. This link and reporting plan has been reviewed and approved by GaDOE's Assessment TAC.

Plan for Ongoing Maintenance

It is important to recognize that the 2020-2021 school year reflects an unprecedented set of circumstances in education. The degree to which students' opportunity to learn was impacted by the pandemic and the challenges of returning to school in a variety of new configurations necessitate revisiting these linking analyses in subsequent years. This study will be replicated with data from the 2021-2022 school year and the link between BEACON performance and projected Georgia Milestones achievement will be reevaluated and updated. Ongoing yearly maintenance of the technical quality of link will then continue.

References

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions: the theory and application of isotonic regression*. New York: Wiley.

Kim, D., Gao, F., Lewis, D., & Boughton, K. (April, 2010). A comparison of predictive methods for benchmark assessment. Paper presented at the annual meeting of the National Council on Measurement in Education. Denver, CO.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, linking, and scaling: Methods and practices* (2nd ed.). New York: Springer-Verlag.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). Fort Worth, TX: Harcourt Brace.