NESS2019 POSTER - STUDENT GROWTH PERCENTILES

History of the Colorado Model

Over the last 20 years there has been a push for "accountability" in secondary education. In addition to measures of "proficiency" there has been a great deal of interest in measuring educational progress or "growth".

One of the most widely used growth measures is the Student Growth Percentile (SGP), known originally as the Colorado model, which was one of a small group of pilot programs funded by the U.S. Department of Education at the start of the twenty-first century.

In 2005 the SGP was approved by the U.S. Department of Education as an acceptable measure of "adequate yearly progress". Under tight deadlines to apply for the "Race to the Top" funds, many states adopted the SGP as a way to satisfy Race to the Top's requirement for a growth measure.

SGP was an easy choice because the developers released it as a free, open source R package (the "SGP" package). Originally more than 20 states adopted the Colorado model, and to this day it is used in Massachusetts and Rhode Island.

Subsequently, many of the developers of the Colorado model ascended to positions of power, and a whole industry grew up around the SGP which is now thoroughly entrenched and resists efforts to replace the SGP.

Issues with the Colorado Model

The Colorado model is designed as a solution to the problem of estimating educational progress given a series of assessment scores.

- It is a "one size fits all" solution that does not take into account specific test characteristics
- It uses quantile regression, which treats the underlying measures as continuous.
- It produces an enormous number of predictions, one for each of the 99 percentiles of the score distribution for each student.
- It uses smoothing (cubic splines) to produce an approximate cumulative distribution function from which percentiles are derived.
- Based on these approximations (and the assumptions they require), it claims to provide interval estimates for individual student growth scores.

Because of its complexity, few people understand how the SGP is computed, and even fewer realize the extent to which it is based on approximations.

Part of the industry that surrounds the SGP involves web-based dissemination of the results with attractive graphics that create the impression of precision and reliability.

A psychometrician who works at ETS told me that every independent researcher who has looked at the SGP has concluded that for an individual student, the 95% confidence interval for their score is 0 to 100. This is much wider than the confidence intervals claimed by the SGP developers.

Motivation for the Poster

Recent developments in statistical theory and computation have opened up the possibility of directly computing a growth measure from a student's individual test question responses using an Item Response Theory (IRT) model.

I think there is a good chance that this could be done for a Rhode Island grade cohort (11,000 students), and possibly even for a Massachusetts cohort (70,000 students).

We propose a hierarchical Bayesian approach to model the sequence of ability parameters (θ in psychometrics terminology) computed from a series of assessments for an individual student as a multivariate normal using a 2-parameter IRT model.

The θ value for the most recent assessment would be used to derive a growth measure by comparing it to the conditional expectation of θ given the prior assessment data under the multivariate normal distribution assumption.