

Operational Characteristics of Student Growth Percentiles Computed from the Massachusetts Common Assessment (MCAS)

J.DIMMICK, E.QUINN¹, A.SORACI, and M.YARMALOVICZ

Stonehill College
Easton, Massachusetts, USA

July 28th, 2015

ABSTRACT: We use the Item Response Theory (IRT) model for the Massachusetts Common Assessment (MCAS) test as a data generating process in a simulation study to examine the behavior of aggregated Student Growth Percentile (SGP) scores.

¹Corresponding author. email: equinn1@stonehill.edu

1. INTRODUCTION

The push for "accountability" in education and the use of concepts like "Adequate Yearly Progress" towards this end has produced a surge of interest in methods of tracking student achievement using a series of measures taken over the course of a child's education.

Compared to the Item Response Theory (IRT) approach to modeling psychometric measurement, the application of longitudinal "growth" models to the measurement of educational progress is relatively immature. This is evidenced by the numerous methods that have been proposed and the existence of federally funded "pilot programs" designed to spur development of these methods.

The issues of validity and reliability are significantly more complicated with these measures than with individual standardized tests. Factors like mobility of students between schools and the effects of family background produce variation in student achievement, yet teachers and schools have little control over these and cannot be held accountable for them [6]. Separating the effects of factors educators have control over from those they do not is a very thorny problem.

It is generally acknowledged that more research on these measures is needed, and this paper is motivated by that fact.

2. THE COLORADO OR STUDENT GROWTH PERCENTILE MODEL

While there is not a consensus on which of the many proposed methods of determining student "growth", a number of states have adopted the Student Growth Percentile (SGP) model developed by Colorado as part of a federally funded pilot program.

This is due in part to a number of desirable theoretical properties this measure possesses, such as invariance to monotonic changes in scale, robustness to outliers, and lack of correlation with prior achievement [3]. Another reason is the availability of free open-source software in the form of the SGP package for the R statistical system [2].

The SGP software takes a longitudinal record of scores from suitable standardized tests over as many as five separate years and uses it to compute the Student Growth Percentile, a single value, for each student.

State departments of education typically describe the scores for an individual student as the change in the student's achievement compared

to students with a similar history on statewide assessments. The Massachusetts Department of Education website contains an implementation brief titled "Using Student Growth Percentiles" [5] that states:

"Massachusetts measures growth for an individual student by comparing the change in his or her achievement on statewide assessments (e.g. MCAS, PARCC) to that of all other students in the state who had similar historical statewide assessment results (the student's "academic peers")."

Results of the SGP growth model for individual students are made available to educators and parents through a website.

We will focus on the SGP model with the Massachusetts Common Assessment (MCAS) mathematics test as the underlying standardized test.

3. MOTIVATION AND OBJECTIVES

Student Growth Percentiles are used by many states to evaluate the educational progress of students and the quality of teachers and schools. The computation of these measures is very complicated and so it is important that their operating characteristics be well understood given the high stakes.

The SGP method is not specific to any one test, but is designed to be applied to a longitudinal history of scores from any underlying standardized test. In this case, the underlying test is the MCAS mathematics exam.

A closer examination of the SGP calculation is provided on page 5 of the SGP Technical Overview [3]:

"Specifically, for each grade by subject cohort, quantile regression is used to establish 100 (1 for each percentile) curvi-linear functional relationships between the students grade 3, grade 4, grade 5, and grade 6 prior scores and their grade 7 scores. 4 The result of these 100 separate analyses is a single coefficient matrix that can be employed as a look-up table relating prior student achievement to current achievement for each percentile."

So, depending on the student's scaled score on the current year's test, one of 100 quantile regression models is selected. The percentile corresponding to the model with the largest predicted scaled score that

does not exceed the student's scaled score for the current year is defined as the Student Growth Percentile for that student.

Any change in the current scaled score will result in a different quantile regression model being used. Furthermore, the computations for even a single student require the scores for the entire grade cohort (70,000 students).

We find it troubling that the end result of this complicated process is often presented with no measure of its uncertainty.

Our objective in this study is to use simulation to estimate the variability of the SGP measures, using the Item Response Theory (IRT) model from the Massachusetts Comprehensive Assessment System (MCAS) as the data generating process in a hypothetical zero growth scenario.

4. THE DATA GENERATING PROCESS

For a given simulated student, their ability parameter (θ) together with the item response theory (IRT) parameters for each question determines the conditional probabilities of every possible answer on the MCAS given that value of θ . We used these probabilities to simulate the results of repeatedly examining the same student, which is the basis of our variance estimate.

The IRT parameter values we used were from the Massachusetts Department of Elementary and Secondary Education's 2011 MCAS and MCAS-Alt Technical Report [7]. Tables of parameter values for the mathematics test appear in Appendix H of this document.

Our intent is not to exactly replicate a particular group of student MCAS scores, but to generate a reasonable surrogate that can be used to model various growth scenarios. With that said, IRT models generally replicate the actual total score distributions well, and that is all the SGP uses [4].

REFERENCES

- [1] Betebener, D. W. (2009), Growth, Standards, and Accountability, The Center for Assessment.
- [2] Betebener, D. W., & Iwaarden, A. V. (2011). SGP: An textupR package for the calculation and visualization of student growth percentiles [Computer software manual].

- [3] Betebenner, D. W. (2011), A Technical Overview of the Student Growth Percentile Methodology: Student Growth Percentiles and Percentile Growth Projections/Trajectories, The National Center for the Improvement of Educational Assessment
- [4] Davey, T., Nering, M., and Thompson, T.(1997), Realistic Simulation of Item Response Data, ACT Research Report Series
- [5] Determining Student Impact Implementation Brief: Using Student Growth Percentiles, www.doe.mass.edu/eval/ddm/GrowthPercentiles.pdf
- [6] Progress on Development and Reporting Measures of Student Growth, The Council of Chief State School Officers (2010), www.ccsso.org/Documents/2010/State_Growth_Models_for%20School_2010.pdf
- [7] 2011 MCAS and MCAS-Alt Technical Report, ,
Massachusetts Department of Elementary and Secondary Education,
www.mcasservicecenter.com/documents/MA/Technical%20Report/TechReport_2011.htm
- [8] Nering, M., & Ostini, R. (2010). Handbook of polytomous item response theory models. New York: Routledge.