ALN
ASSESSMENT LEARNING NETWORK

LEARNING POINT

Measuring Student Growth: More than just choosing a model

The increased interest in measuring student growth in public education has spurred much discussion about the pros and cons of various models that can be used to quantify and communicate growth in student achievement. While characteristics of the various growth models are often debated, considerations about the nature of the data used in these models—the inputs—are discussed and acknowledged far less often. The result is a mistaken belief that any growth model can be used with any assessment data.

This is an errant belief. As much care must be taken in evaluating the quality of the data used for any particular growth model as is used in evaluating the model itself. Ignoring the requirements for data used in a growth model can lead to invalid conclusions drawn from the results of the growth modeling.

This *Learning Point* will cover three areas that must be considered when using a growth model. Two of the considerations relate to the nature of the data used for growth modeling. The third, attribution, is related to the interpretation of results from a growth model. Each of these aspects influences what can and cannot be inferred from the results of a growth model but is separate from the growth model itself.

> **"As much care must be taken in evaluating the quality of the data used for any particular growth model as is used in evaluating the model itself."**

From where did the data come?

The characteristics of the assessments from which data are gathered and fed into a growth model are important and need to be considered. Tests are built for a specific purpose or purposes; when they are used for something other than that for which they were designed, results might not be valid. As an example, a summative test developed to assess proficiency may not provide data that is particularly useful for measuring growth. The test will have many items that assess the domain around the Proficient level ("domain" refers to the body of knowledge, skills or abilities being measured or examined). For summative assessments, the point of proficiency is the most important point on the score scale. Test items will be chosen so as to measure very precisely around the point of proficiency at the expense of precision at points far above or below proficiency on the score scale. What is important, for this particular type of test, is whether someone is above or below proficiency. Determining exactly how much above or below a student is doesn't carry as much import.

What are the implications of this for a growth model? Students at different points along the learning continuum will be measured with differential precision. Students near the point of proficiency will be measured with greater precision than students far from the point of proficiency—either

What is a "growth model"?

A student growth model attempts to capture and communicate changes in a student's, or possibly group of students,' achievement in a specified content area over time.

far above or far below. Using data with differential precision as the inputs for growth modeling can result in differences in our confidence in the results of a growth model.

Using a test designed to measure consistently across the entire range of ability in a content area may provide data that is more useful for growth modeling. A test of this type is built differently from a test built to measure proficiency, and thus might not work well as a test of proficiency. Characteristics of the test and how it's built are important considerations when thinking about your growth model.

To learn more

Proficiency and Growth: What's the Difference?
(MAC, 2017)
https://bit.ly/LP-Proficiencyvsgrowth

A Practitioner's Guide to Growth Models
Castellano, K.E. & Ho, A.D. (CCSSO, 2013)
http://tinyurl.com/yb74y2o6

Growth Models: Issues and Advice from the States
http://nces.ed.gov/programs/slds/pdf/guide_growth-model.pdf

Thinking About Improvement in Student Test Performance
http://tinyurl.com/improvement-Gullen

What are the characteristics of the data?

A second consideration relates to the distribution of data that is fed into the growth model. Some growth models may also have requirements for certain aspects of the model. If a linear growth model is chosen, data must demonstrate a linear relationship. Additionally, assumptions about the "shape" of the data used in the model may be made. Assumptions about the normality of the data and independence of observations are common requirements for some growth models. Yet, the actual distribution may not approximate a normal distribution because the test used might instead show a positively or negatively skewed distribution.

Some growth models also make assumptions about "internal" aspects of the model. One of these requirements might be that a model must estimate with equal precision across the input scale. Technically known as homogeneity of residuals, this must be evaluated after the growth model is fit. The more technical the growth model chosen, the more technical the requirements of the data.

Different growth models will have different requirements. Some models make only a few assumptions that need to be verified, while others will make many—some very technical—requirements of the data and the model fit.

What about attribution?

One hoped-for benefit from the use of growth models is the ability to attribute student growth, or lack thereof, to some specific program, school, or even teacher. These are commonly referred to as "value-added" models. Note that these types of models are looking at and claiming two distinct things: First, did growth in student achievement occur? And second, can that growth, or lack thereof, be attributed to something specific? All value-added models will have a growth model component. Not all growth models have to be value-added models. For some, it might be sufficient to capture and report the amount of student growth.

Adding a value-added component to a growth model involves adding many layers of complexity. This is because determining attribution is essentially evaluating causality. A value-added model attempts to claim that a specific program, school, or teacher caused a specific amount of growth for students. The determination of causality is more of a research design question than a statistical question. Research design principals such as random selection, random assignment of students to teacher(s), and others are important considerations in determining causality and assigning attribution. In public education, there are often difficult hurdles in meeting these requirements that must be dealt with. If these issues are not addressed, claims of attribution are weaker and may not be valid at all.

In summary

Growth models for student achievement range from fairly simple to incredibly complex. Different growth models will have different requirements for the assessments that provide the data to the model and for the ways in which that data is distributed. Whatever growth model is chosen, the assumptions made by that model must be evaluated. If they are not met, the results of the growth model and their interpretations are suspect.

In addition to determining growth, the ability to attribute that growth to something specific is becoming more desirable to some stakeholders and policy makers. This evaluation requires more than just statistical modeling. A statistical model is important for quantifying growth, but the design of the study used is important for valid determination of attribution of that growth.

The Michigan Assessment Consortium's Assessment Learning Network (ALN), is a professional learning community consisting of members from MI's professional education organizations; the goal of the ALN is to increase the assessment literacy of all of Michigan's professional educators.