# HOW SHOULD STATES CHOOSE A GROWTH MODEL?

## *Aligning Your Growth Model With Policy and Technical Values*

February 2025

**Scott Marion, Chris Domaleski & Will Lorié**

*National Center for the Improvement of Educational Assessment*

Center for Assessment

National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

HOW SHOULD STATES CHOOSE A GROWTH MODEL?

*Aligning Your Growth Model With Policy and Technical Values*

The National Center for the Improvement of Educational Assessment, Inc. (the Center for Assessment) is a New Hampshire based not-for-profit (501(c)(3)) corporation. Founded in September 1998, the Center's mission is to improve student learning by partnering with educational leaders to advance effective practices and policies in support of high-quality assessment and accountability systems. The Center for Assessment does this by providing services directly to states, school districts, and partner organizations to support state and district assessment and accountability systems.



## SUGGESTED CITATION:

Marion, S., Domaleski, C., & Lorié, W. (2025). *How should states choose a growth model? Aligning your growth model with policy and technical values.* Center for Assessment.

## PHOTO CREDIT:

Cover Photo: Allison Shelley/The Verbatim Agency for EDUimages

HOW SHOULD STATES CHOOSE A GROWTH MODEL?

*Aligning Your Growth Model With Policy and Technical Values*

# TABLE OF CONTENTS

# HOW SHOULD STATES CHOOSE A GROWTH MODEL?

## *Aligning Your Growth Model With Policy and Technical Values*

## INTRODUCTION

The Every Student Succeeds Act (ESSA) requires that state accountability systems include "another academic indicator" for elementary and middle schools in addition to academic achievement. All states except two use a measure of student longitudinal growth as their other academic indicator. Student growth measures are not used to check a federal requirement; they provide an enriched view of student and school performance (Dale Carlson's 2001 paper provides an elegant discussion of these views).

But how should states decide which growth model to use?

Over the past few years, the Center for Assessment has helped several states examine their growth models to confirm, revise, or replace them. We find that these decisions involve much more than technical factors. From a strictly technical perspective, there isn't a "right" or "wrong" approach.

To choose the right approach for their state, education leaders should consider their policy priorities, the intended uses of the growth results, the anticipated implementation challenges, and, yes, technical factors. These things can be difficult to reconcile. Sometimes, a state's values conflict, such as when elevating one priority works against another. For these reasons, we suggest that states and districts use a principled approach to establish their most important values. This will help them evaluate potential growth models against an agreed-upon set of criteria.

*From a strictly technical perspective, there isn't a "right" or "wrong" approach.*

## STARTING WITH HIGH-QUALITY BACKGROUND KNOWLEDGE

These examinations are most sustainable when a broad-based group of constituents is empaneled in study groups or task forces to provide recommendations to state leaders. Most practitioners and policymakers, however, do not live in the world of growth models—which can be very complex—so it is critical to ensure that these panelists have adequate background knowledge to engage in the work.

We have found resources such as Castellano and Ho's[1] (2012) *A Practitioner's Guide to Growth Models* useful for providing background knowledge that allows advisory panel members to meaningfully engage in deliberations about growth models.[2] This guide explains the key technical and

[1] We thank Andrew Ho for suggesting that we write up this approach for selecting growth models.

[2] We also rely on resources such as *Policymakers' Guide to Growth Models for School Accountability: How do Accountability Models Differ?* and *Considerations for Including Growth in ESSA State Accountability Systems.*

interpretative characteristics of some of the most popular families of growth models. The table below summarizes the major models and key questions each model is designed to address.

| MODEL | KEY QUESTION |
|---|---|
| **Gain Score** | What is the magnitude of progress on a vertical scale? |
| **Growth-to-Standard** | Is the student's progress on track to a significant target? |
| **Categorical (Value Table)** | Has the student transitioned from one performance category to another? |
| **Growth Percentile** | How does the student's performance this year compare to his or her academic peers? |
| **Regression or Value-added** | Statistically controlling for selected factors, has the student grown more or less than expected? |

*Adapted from Castellano and Ho, 2012*

This technical information provides panelists with helpful background to ground the discussions. But we suggest that technical factors shouldn't be the first thing study group members tackle. We have found that it helps to start by stepping back and considering key principles and characteristics that reflect the group's goals and values for the growth model.

***We suggest that technical factors shouldn't be the first thing study group members tackle. We have found that it helps to start by stepping back and considering key principles and characteristics that reflect the group's goals and values for the growth model.***

## WHAT DO YOU VALUE WHEN MEASURING STUDENT GROWTH?

There is no gold standard for evaluating growth measures, but the deliberative bodies we've worked with have found it helpful to consider the following questions/criteria associated with growth models:

1. What is the relationship between aggregate (e.g., school-level) student growth measures and prior achievement?
2. Does the model favor higher- or lower-performing schools, or does it treat all schools equally?
3. Does the model include student and school background characteristics?
4. Is the model simple enough to be easily calculated and understood, or is it complex and can only be computed by technical experts?
5. Is the model proprietary or open-source?
6. Is the model dependent on the specific test score scale or scale agnostic?
7. Does the model have well-documented technical properties and quality?

These questions address the most important policy, technical, and practical growth considerations consistent with the goals and design principles for the overall accountability system. Let's dive a little deeper into each one to show why they're important questions to ask.

### 1. What is the relationship between aggregate (e.g., school-level) student growth measures and students' prior achievement?

This question helps us evaluate whether growth is picking up on distinct aspects of student progress rather than simply amplifying the influence of achievement (i.e., proficiency rates) already in the accountability model. This criterion is usually evaluated by calculating the correlations between aggregate measures of prior achievement and student growth. Lower correlations indicate that the growth model picks up information distinct from student achievement.

### 2. Does the model favor higher- or lower-performing schools?

*All schools should be able to demonstrate favorable growth results when students demonstrate academic progress.*

Given the strong relationship between student achievement and student background characteristics (e.g., economic status), a growth model highly related to prior student achievement will likely be highly related to these background characteristics. In this case, the growth results would favor higher-achieving schools. This question is intended to evaluate whether schools that serve a greater proportion of economically disadvantaged students or English learners, for example, have noticeably different growth distributions than schools with fewer students in these groups. All schools should be able to demonstrate favorable growth results when students demonstrate academic progress.

### 3. Does the model include student and school background characteristics?

Value-added models generally incorporate student—and perhaps school—background factors into the growth model. When used for teacher evaluation, including background factors was thought to be a way to make the model fairer to teachers with different classroom compositions. Several concerns have been raised about these complex models. Building many of these demographics into a model (covariates) increases the complexity. The number and nature of "adjustments" resulting from using the various covariates can make the model opaque and inaccessible. Many are also concerned that such models implicitly set different expectations for different groups of students.

### 4. How simple, how complex?

Growth models vary considerably in their complexity. Some, like gain-score models, require users to subtract the score in year 1 from the score in year 2. Others, like value-added models, require technical experts to employ complex regression approaches (or something similar). Simplicity is appealing because most users can understand how the results were calculated. But simple models generally do not account for important factors that threaten technical quality and fairness (think of the "flat tax"). Complex or comprehensive models often address more stringent technical standards, but is that gain worth it if few can understand the model's inner workings? Others may accept the tradeoff, noting that we accept complexity in other areas. After all, we know how to use our computers effectively even if we can't explain how microprocessors work or take them apart and rebuild them.

### 5. Is the model proprietary or open-source?

Some growth calculations rely on proprietary models. The user (e.g., the state) cannot run the model without employing the company that owns it. Some might find it worth the tradeoff to use a proprietary model to get the exact model they want. Other models are open-source, where the code is made freely available to any user, generally under a Creative Commons or open-source license. At the Center, we make no bones about advocating for open-source solutions for transparency and knowledge transfer.

### 6. Is the model dependent on a specific scale or scale-agnostic?

Growth models can be scale-dependent or scale-agnostic. A scale-dependent model relies on a specific test score scale or achievement levels. For example, if a Smarter Balanced state used a gain-score model based on subtracting the prior grade score from the current grade and then translating these gain scores into indicator values, and then switched to a different test, it would have to redo its growth model. This is not impossible, but it could be very challenging if the state's new test did not use a vertical score scale. A scale-agnostic model (like student growth percentiles or value-added models) is based on the relationship between two sets of test scores. As long as the relationship between old and new test scores is similar to the relationship between year-to-year scores on the old test.

### 7. Does the model have well-documented technical properties?

The model's technical quality can be evaluated in several ways, such as the consistency and accuracy of its results, the comparability of the results within and across years, and the precision of its growth results throughout the distribution. The model should also be sensitive to detecting progress even for students who score among the lowest or highest in the state. Finally, the technical quality should be documented through independent analyses and/or peer-reviewed publications.

*The model should be sensitive to detecting progress even for students who score among the lowest or highest in the state.*

## THE PROCESS: BRINGING IT ALL TOGETHER

It is hard to keep all of these criteria in one's head at the same time. Some are more important—or at least more encompassing—than others. Almost all groups we've worked with have agreed that having a weak relationship with prior achievement, being open-source, and being of high technical quality were the most critical criteria for adopting a growth model. These groups often added another important consideration: Whatever model is chosen should be explainable to various audiences and minimize the chance of misconceptions about the growth results.

Each feature exists on a continuum (e.g., highly related to prior achievement to minimal relationship with prior achievement). Once the participants understand these features, each group member is asked to consider where they stand on each continuum. Participants can consider these or other important tradeoffs and indicate their preference for each feature using an online form or a sheet of poster paper, marking the place on the continuum that matches their position.

After the first round of placements, the facilitator engages the participants in a discussion of their preferences, particularly for features with the most substantial disagreements. Eventually, the facilitator and group members summarize the preferences for all of the features in a picture, such as the one below.

| | | |
|---|---|---|
| Highly related to achievement | | Unrelated to achievement |
| Favors low-performing schools | | Favors high-performing schools |
| Includes background factors | | Doesn't include background factors |
| Simple | | Comprehensive |
| Open-source | | Proprietary |
| Scale-dependent | | Scale-independent |
| Technically strong | | Technically weak |

We do not start investigating specific growth models until we agree on the criteria we will use to select one or more for further investigation. Having these touchstones is critically important for matching the selected model with constituent and policymaker priorities.

Panel preferences will often suggest one or two main models for further investigation. Using historical data, a state might then examine the extent to which the candidate model(s) yield results in line with expectations. Because all model decisions are ultimately a choice among alternatives, the results of an empirical study can provide additional support for why a particular model was adopted.

We encourage others to use or adapt this approach for either selecting a new growth model or validating an existing choice.


National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

www.nciea.org