



ACADEMIC GROWTH: RECOMMENDATIONS FOR THE SENATE HELP COMMITTEE

The Center for Assessment's response to a Request for Information

INTRODUCTION

The National Center for the Improvement of Educational Assessment (Center for Assessment) appreciates the HELP committee's focus on academic growth. When developed and used appropriately, growth measures provide essential information to support student learning and to monitor academic progress at the aggregate level, including schools, districts, and student groups.

The Center for Assessment has led the development and implementation of student growth models in collaboration with states and other partners. More than 20 years ago, the Center's co-founder, Dr. Richard Hill, developed the approach known as value tables, which remains in prominent use today. Later, Dr. Damian Betebenner developed the Student Growth Percentile (SGP) model, which is the most widely used growth model in the United States.

Consistent with the Center's principles, the ideas behind value tables and SGPs are open-source and freely available. At the same time, the Center does not advocate using any one student growth model. Rather, we work with our state clients to design and implement a range of approaches aligned to the state's context and priorities. Our interest is in making sure measures of academic growth are technically defensible and useful for their intended purposes.

POLICY GOAL #1: FAIR MEASURES OF SCHOOL QUALITY

1. What have states learned from developing different measures of student growth?

Key points: There is no one "right" growth model for all purposes. Choices among approaches should be guided by clearly identified purposes and priorities. In particular, when using growth to gauge school quality, it's important to decide on the right balance between growth and achievement.

The Center for Assessment has worked with more than 35 states over the past 20 years, helping them implement student longitudinal growth models. Early on, this work was informed by a [framework for considering school quality](#) developed by Dale Carlson, the Center's first Board Chair and a former president of the National

Council on Measurement in Education. In this framework, Carlson argues that we need to consider at least four related metrics for evaluating school quality:

- Achievement: How did the students in this school perform on the tests this year?
- Progress: How did the average achievement in this school compare to prior years?
- Growth: How much did students improve their performance compared to last year?
- Acceleration: Is this school improving students' growth rates?

Carlson's first two categories refer to *achievement* at a fixed point in time and over time. These quantities are used to address questions about the average achievement of students at a school (i.e., status) and whether that average is increasing over time (i.e., improvement). Carlson's last two categories refer to *academic growth* at a fixed point in time and over time. These quantities are used to address questions about the average growth of students at a school (i.e., learning) and whether that growth/learning is increasing over time (i.e., acceleration). Growth/learning is privileged—and has been the basis of our work with states for over 20 years—because it is what we believe the education system can control. This is the premise of our response to this RFI.

Carlson's framework reminds us to understand the meanings and differences among the four categories when defining school quality. Student longitudinal growth is critically important because it is something the education system, its districts, schools, teachers, and parents can control to help all students meet key achievement targets vital to long-term success. Schools maligned for low achievement can demonstrate exemplary rates of student growth and should be seen as successes, whereas a school lauded for high achievement may show low growth that indicates a need to remediate its students' low learning levels.

... there is no "right" growth model for all purposes and hence for all states.

The Center for Assessment has developed the two most widely used growth methodologies currently used with state assessment systems, including English language proficiency testing. [Center for Assessment experts developed value tables](#), a contingency table approach for summarizing student longitudinal progress over successive years. According to a [2024 survey of state accountability systems by the Education Commission of the States](#) (ECS), value tables are still used in 10 states.

Dr. Betebenner developed and implemented student [growth percentiles](#) (SGPs) that bridge norm-referenced growth approaches with criterion-referenced growth-to-standard approaches. The methodology is currently used in at least 24 states, according to the ECS survey.

Our efforts to help states choose the best growth models for their goals have taught us that even though there are wrong ways to create a growth model, there is no "right" growth model for all purposes and hence for all states. States have different values, policy contexts, and intended goals for their growth models. The federal government should not advocate for a single "best growth model." We have worked with many state advisory committees to help them recommend growth model options for their state (the process we use is described in this [paper](#)). Among other considerations, we help state leaders and their constituents identify what they value in a potential growth model by working through the following questions. Technical properties are an important consideration, but far from the only consideration.

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 is it scale-agnostic?
7. Does the model have well-documented technical properties and quality?

We have found that states prioritize these seven considerations in different ways. For example, most states want a relatively low correlation between aggregate growth scores and aggregate prior achievement, to ensure they are capturing a different aspect of school quality rather than simply adding another achievement measure. We know we can reduce these correlations by including student and school background variables in a complex value-added model (VAM), but many states have backed away from such models for various reasons. First, state leaders have felt that the challenge associated with explaining the model's complexity to user groups was not worth the potential benefit. Second, many are not open-source, so it is hard to evaluate their technical quality. Finally, by statistically controlling for characteristics such as economic disadvantage, race, and special education status, the model effectively sets different expectations for different groups of students.

The overarching question in Policy Goal #1 is whether “fair” measures of school quality exist. We have found that it is not simple to determine what is fair because fairness isn't a statistic we test for; it involves asking questions about the values and goals that underpin evaluations of fairness. For example, we first ask, fair for whom? Since No Child Left Behind, fairness to students has meant helping students reach desirable levels of achievement (i.e., proficiency). Proficiency-based accountability embodies that notion of fairness. Schools educate students entering at widely varying levels of achievement, from well below to at or above proficiency. Policymakers generally want growth measures that track progress toward key milestones for students below proficiency while promoting continued learning for students at or above proficiency.

... fairness isn't a statistic we test for; it involves asking questions about the values and goals that underpin evaluations of fairness.

Yet fairness to schools does not necessarily lead to the outcomes we associate with fairness to students. It is well documented that student achievement is influenced by many factors beyond teachers' control, which is why some growth models that include background factors were popular when test scores were used in teacher evaluation systems. While the motivation to include such factors is understandable, their inclusion sets lower expectations for certain groups of students, ultimately obscuring the reality that individual students' goals will not be met. This reality will not change with new growth models. The key is to recognize these tensions and to build a system that is faithful to both.

Growth-to-standard models include expectations that students will reach the proficient or other defined performance level within a specified time period. Generally, these time frames are defined as “three years or by the end of a specific grade (e.g., 8 or 11).” Growth-to-standard models acknowledge the importance of growing towards a meaningful outcome rather than growth for growth’s sake. For example, these models are used consistently in annual English language proficiency testing, where helping students reach the point where they can access content in English is considered critically important. This can be considered fairer for students, but because such models tend to be more highly correlated with prior achievement than pure growth models, they might be considered less fair to school personnel who have students entering their school performing below their nominal grade level or facing adverse circumstances influencing their education and learning. However, growth-to-standard models are more highly correlated with prior achievement than typical SGPs or VAMs, so when used in a school accountability context, they end up weighing achievement considerably more than growth.

Combining these values into an accountability indicator system that informs and, ideally, improves the education system may be complex. The effort requires a reporting and communication strategy that extends beyond schools’ accountability ratings to identify the actions necessary to improve student learning.

Finally, all states have technical advisory committees that are regularly called upon to review whether the growth model and the larger accountability system are working as intended. The Center for Assessment coordinates technical advisory committees in 20 states and collectively serves on more than 40 technical advisory committees. We can attest that technical advisory committees often provide critical advice to state leaders on improving the technical characteristics and utility of states’ growth models.

2. Are there any kinds of federal support that would be useful to states seeking to implement new growth measures or revise existing ones?

Key points: Growth scores are only as good as the tests on which they are based. The United States Department of Education (USED) should maintain and enforce requirements to administer high-quality tests in grades 3-8 and high school. We suggest increasing the funding for the State Longitudinal Data Systems program. Finally, we propose increasing support for the development and curation of effective communication resources.

As advocates for productive uses of growth data, we recommend that the United States Congress and the U.S. Department of Education continue to require that states use high-quality assessments in grades 3-8 and high school in at least English language arts and mathematics to support the use of student longitudinal growth measures in elementary and middle schools, and perhaps into high school.

Student-level academic growth models gained traction once states established longitudinal data systems. These data systems were supported by the State Longitudinal Data Systems (SLDS) grant program. State data systems need regular updates, especially as the volume of available data

Misconceptions about the distinctions among achievement, progress, growth, and acceleration ... continue to hinder clear interpretations of school quality.

is increasing dramatically. Without longitudinal assessment data, growth calculations become impossible, which magnifies the potential harm that misconceptions about assessment data, based on achievement or achievement trends, can have. We advocate for increased funding for the SLDS program. Increased funding could ensure data systems are based on the latest technology, enable them to incorporate additional data elements, and help expand them to reach pre-K through postsecondary institutions.

The technical side of growth analyses is mature. SGP, VAM, and value-table approaches used by almost all states have changed little in the last 15 years. Though advancements are possible, communication has been and continues to be the biggest missed opportunity. Misconceptions about the distinctions among achievement, progress, growth, and acceleration, as outlined by Carlson, continue to hinder clear interpretations of school quality. Growth reports are often buried in school accountability reports, which contributes to this confusion.

One of the biggest challenges with growth models is accurately communicating the meaning of the results and supporting constituents in using them effectively. Some states have done this well, while others have struggled. Given the stability of growth model analysis techniques, we suggest that the federal government curate effective communication resources and use cases, including visualizations, for the most commonly used growth models to help states communicate their models to users and support their effective use.

3. Are there design changes to state assessment systems that would support the creation of higher-quality growth measures, and are there any federal policies standing in the way of such innovation?

Key points: While current federal requirements help ensure broad content coverage, they may constrain innovation, such as in content-referenced growth research. We also support limited, strategic use of off-grade items to improve growth measurement at the upper and lower ends of the achievement distribution without undermining grade-level assessment.

Some changes should be avoided. We caution against shortening summative tests, which could reduce reliability and distort growth estimates. Also, fall-to-spring growth measures should not be used for consequential accountability purposes given their well-documented risks and unintended consequences.

In general, continuing to require states to use assessments that meet rigorous technical standards is probably the most important thing the federal government can do to support high-quality growth models. Current alignment requirements mean that the full set of content standards must be represented on the test. This is often accomplished through a “domain-sampling” approach, in which items are distributed to cover the necessary standards. The current alignment requirements, however, might unduly restrict the creation of assessments that support measuring content-referenced growth in “big ideas” (e.g., proportional reasoning) across key grades. We suggest that the federal government allow state assessment pilots to best support content-referenced growth approaches or other innovative methods for measuring student growth.

The special education community has long advocated against off-grade-level accountability assessments to measure the performance of students performing well below grade level. This is based on concerns that off-grade-level assessments will drive instruction toward off-grade content, exclude students from their grade-level peers, and lower expectations. While we understand the potential unintended negative consequences of allowing off-grade testing, we suggest allowing a few

states to pilot assessments with a limited number of below- and above-grade items to get more precise estimates of student growth for students at either end of the achievement continuum. This can be accomplished without abandoning measures that primarily reflect performance on grade-level content. One way to do this is to ensure that any off-grade content is clearly linked to grade-level content. For example, using learning progressions to adaptively route students to off-grade-level items; using multistage test designs to allow accommodations that alter grade-level constructs (e.g., audio comprehension and oral reading fluency for tests or reading comprehension); or linking alternate assessments and general assessments to form one seamless assessment.

We also recommend caution when allowing states to shorten their state summative tests. This can reduce overall reliability and conditional standard errors of measurement to the point that growth estimates are compromised. Additionally, we recommend ensuring that tests can represent the full distribution of student achievement and avoid “floor” and “ceiling” effects.

Finally, we oppose allowing fall-to-spring growth calculations to be used for consequential purposes, which could be feasible with through-year assessment systems, for use in accountability systems. The risks and unintended negative consequences associated with fall-to-spring accountability growth measures have been well documented, tracing back to the early days of Title I accountability, where school personnel encouraged students to “try a little less” on the fall tests.

POLICY GOAL #2: EMPOWERED FAMILIES

4. What have states learned about how best to communicate information about school-average growth to families, such as through school report cards?

Key points: Prioritize communication that supports appropriate interpretations and uses for students. Effective growth communication must go beyond reporting numbers and be tailored to different audiences, using clear visuals and contemporary communication tools rather than relying on technical guides alone.

Effective communication about growth should begin by describing students’ growth for students, parents, and teachers. Parents want to understand how their students are progressing, and educators want information to make sense of student learning over time. In addition, aggregated growth information helps identify patterns and progress for student groups and schools.

When it comes to reporting, simply placing a number on a report is insufficient. Such approaches are not just ineffective but can also lead to misinterpretation and misuse of data. It’s important to ensure that reports that include information about academic growth clearly support appropriate interpretations and uses, while explicitly cautioning against unsupported interpretations or uses.

More effective communication strategies are multifaceted and designed to meet the needs of different audiences, including parents, educators, and policymakers. Good practices make use of compelling visuals and a range of communication tools, such as videos, interactive websites, webinars, and online modules, to support understanding and build user capacity. Technical manuals and interpretation guides have not been adequate to support defensible interpretations of growth results; newer technologies with close support offer greater potential to help users understand student growth data and use it effectively. That said, no matter how compelling the communication approach, it is still difficult to grab parents’ attention amid the daily onslaught of information.

5. How can the federal government support cross-state learning about communicating information to families?

Key points: *States already have some ways to learn from one another, but stronger federal leadership could help scale what works. This can be accomplished by supporting trusted, nonpartisan organizations to study how people actually interpret and use growth, and by developing practical, open-access resources to ensure growth information is understood and used appropriately.*

There are already some good mechanisms in place to support cross-state learning, such as the state collaboratives coordinated by the Council of Chief State School Officers (CCSSO), notably its accountability systems and reporting collaborative. However, there are opportunities for the federal government to do more to advance cross-state learning. One option would be to commission nonpartisan, highly respected organizations, such as the National Academy of Sciences, Engineering, and Medicine (NAEM) or the National Academy of Education (NAE), to conduct research on effective reporting and communication practices. Other organizations, such as the National Parent Teacher Association (PTA) or the National Association of State Boards of Education (NASBE), may also be valuable partners, offering insights into the priorities and communication strategies that resonate with constituents. Such initiatives should include assessment experts and communications professionals, but they must also include representatives from districts and states who can both share their best practices and respond to new ideas proposed by experts. Moreover, all outcomes from this work should be widely distributed and licensed as open source to maximize use and impact.

There are opportunities for the federal government to do more to advance cross-state learning.

Future initiatives should also encourage research that goes beyond collecting stakeholder preferences or “wish lists.” We need studies that examine how different audiences interpret growth and how they intend to use the results. Approaches such as cognitive laboratories, where study participants verbalize their thinking, or structured usability studies could help determine whether users can recover the intended meaning from reports and resources and use them as intended. In a cognitive laboratory, people are observed as they read and use information, allowing researchers to understand where designs are clear and where they cause confusion.

Additionally, we need to develop practical guidance to help constituents use growth information to support student learning and school improvement. In many cases, the challenge is not access to data but a lack of clear direction on how to use it well. Well-designed resources can help close this gap and improve the likelihood that growth information is used productively.

6. How have states addressed the challenge of measuring growth in elementary schools, where statewide assessments typically do not begin until 3rd grade?

Key points: *We have not seen strong evidence of promising practices for using K–2 assessment results for growth or accountability. These assessments are generally not as high-quality as those in later grades. There is limited research evidence about the stability and meaning of growth estimates for young learners, and existing studies show mixed results when linking K–2 measures to later outcomes. More research is needed before K–2 assessment results are used for consequential purposes. The primary focus of these assessments should remain on supporting early learning, especially for students who start behind.*

We advise caution when using K-2 assessment results. These assessments are rarely comparable statewide, since districts use different assessments. Moreover, compared with assessments in the upper grades, there is less certainty about the quality and interpretability of the information these measures provide. When K-2 assessments are used, it is essential to carefully examine their technical quality and ensure they meet the same core expectations as assessments in later grades. In addition, there is limited empirical evidence about the characteristics and stability of growth estimates for young learners.

We know of one study that attempted to link K-2 reading assessments to the grade 3 state reading assessment. The accuracy of grade 3 score predictions varied by student group. This is likely because the construct changed from word recognition to language comprehension. Note that this was not a growth study. It was an attempt to predict 3rd grade reading scores from K-2 assessments. Further research is needed to better understand these patterns before K-2 assessment results are used for consequential purposes. As early childhood is seen as critical to later student success, maximizing learning for younger students—especially those starting behind—should be a priority to improve later outcomes.

POLICY GOAL #3: INFORMED POLICYMAKERS

7. What changes to the National Assessment of Educational Progress or other federal data collection efforts would support a national focus on student growth?

Key point: We advise against changes to NAEP. It is not designed to measure traditional student growth, and attempting to do so would risk damaging the trend data that makes NAEP valuable. We support continued study, but the timeline is longer-term.

The required changes to NAEP to support traditional ways of measuring student longitudinal growth would be untenable. The National Assessment Governing Board (NAGB) had to vote last year to reduce the NAEP schedule because of new budget restrictions. Increasing the number of assessments to support growth measurement would triple NAEP's cost. Furthermore, supporting the measurement of students' longitudinal growth would require a radical change to the NAEP sampling methodology. NAEP is the gold standard of large-scale educational measurement because it does one thing (monitor long-term trends) and does it exceptionally well. Trying to change the purpose of NAEP could damage our ability to use 35-year trends and still not support student growth measures.

One potential breakthrough on this front will be presented by Senior Associate Dr. Damian Betebenner and Prof. Henry Braun of Boston College at the 2026 National Council on Measurement in Education conference. They will demonstrate an innovative framework for conducting longitudinal data analyses that unifies non-longitudinal data such as NAEP or TIMSS.

8. How have states most effectively used growth data to inform policy decisions?

Key points: There are very few clear examples of growth results informing policy decisions. This gap highlights the need for federal support to develop practical resources and demonstrate effective uses of growth data. Growth can serve as an effective outcome variable, but it's important to be careful about the policy tradeoffs of using models that rely on prior achievement versus those that include additional factors.

We know states use growth to improve accountability, but we lack clear examples of how state and district leaders have used growth specifically to inform policy. This is an area that could benefit from federal investment in initiatives to develop resources that help translate data into information and showcase promising practices for using growth well. For example, student growth data can serve as a much more valuable outcome variable in efficacy research than proficiency rates or changes in proficiency rates.

We need to make better use of growth data to support evaluation and continuous improvement. Both value-added models and student growth percentiles could serve as effective outcome measures in research and evaluation studies. However, SGPs and similar approaches that rely primarily on prior achievement (rather than additional covariates, such as race/ethnicity, economic status, etc.) may provide a clearer, more straightforward signal for policy decisions. We do not oppose the use of VAMs that include additional variables, but such models raise important policy questions, including whether they may mask effects that are important to detect or whether growth expectations are appropriate for all schools and student groups. These and other factors are critical when using VAMs to inform policy decisions.

9. How can federal policy incentivize states to focus on growth and remove any barriers to state innovation?

Key points: Measuring growth in high school could benefit from greater flexibility, as many states lack coherent assessment systems. Moreover, supporting collaboration and innovative accountability pilots could help states develop more effective practices.

More incentives and flexibility could support improved practices for measuring growth in high school. In particular, many states have found it challenging to extend growth measures into high school, often due to the absence of a continuous assessment system across grades. This is compounded by the fact that high school growth often uses the last assessment the student takes, so, within current accountability frameworks, achievement takes priority over growth.

Providing states with incentives and flexibility to develop a more complete and coherent set of high school assessments would be an important first step. We are thinking more along the lines of end-of-course tests rather than simply adding commercial survey tests that are not tied to the specific content that students are learning. In addition, supporting cross-state collaborations or shared resources could help states learn from promising practices as they explore more effective approaches to measuring growth in high school.

Finally, consider authorizing an innovative accountability pilot, not unlike the Innovative Assessment Demonstration Authority (IADA). This would allow states to experiment with different types of systems, including those that prioritize academic growth to support continuous improvement.



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

www.nciea.org