STATE ACCOUNTABILITY SYSTEMS

MEASURING & REPORTING ACADEMIC GROWTH FOR STUDENTS WITH SIGNIFICANT COGNITIVE DISABILITIES

CCSSO
One Massachusetts Avenue, NW
Suite 800 Washington, DC 20001

CCSSO
Council of Chief State School Officers®

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, Bureau of Indian Education, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public.

AUTHORS:

Phoebe Winter, Meagan Karvonen, Juan D'Brot,
Jeffrey C. Hoover and Sandra Hopfengardner Warren

TABLE OF CONTENTS

INTRODUCTION

EXECUTIVE SUMMARY

PURPOSE OF THIS HANDBOOK

The purpose of this handbook is to outline key questions and approaches to describing academic growth for students with the most significant cognitive disabilities (SCD) and incorporating those growth measures into accountability indicators. Growth is defined as change over time for individual students or groups of students.

Growth measures are a way of quantifying or otherwise describing the amount of change in achievement over time, for individual students or groups of students. Growth measures are produced by applying **growth models** (e.g., categorical gain, student growth percentile) to assessment scores. **Growth indicators** aggregate information from growth measures for use in accountability systems, so states can monitor their progress toward achieving long-term goals. The relationship between growth models, measures and indicators is illustrated in Figure 1.

FIGURE 1: RELATIONSHIP AMONG GROWTH MODELS, MEASURES & INDICATORS*

*Test scores from more than two time periods may enter into a growth model. The minimum is two scores.

This handbook aims to help states make decisions about how to define and measure growth for students with SCD and how to incorporate growth measures into indicators used in accountability formulas. Given the short history of alternate assessments and rapid changes in the design of alternate assessments based on alternate academic achievement standards (AA-AAAS) in the past 20 years, there is very limited literature on growth measures based on AA-AAAS and few examples of their use in practice. In a 2021 survey of states conducted for this handbook, only 7 of the 25 states responding indicated that their accountability systems included growth on the alternate assessment in their accountability systems.

Including students with disabilities in accountability is a civil rights issue. Including students with SCD in the statewide assessments used for accountability is intended to ensure that these students' results are considered when accountability results are used to drive decisions about educational programming and resource allocation. Inclusive assessment systems support equitable treatment in reporting and accountability (Thurlow et al., 2001).

This handbook describes an overall process for making decisions about **growth measures** and **growth indicators**, using approaches that support valid inferences regarding, and uses of, the growth measure. It does not address methods for measuring **improvement** (i.e., school and LEA progress toward long-term goals as defined by indicators in consolidated state plans).

For example, a state might evaluate the improvement a school makes in educating Grade 3 students in math in Year 1 versus Year 2, as indicated by an increase in the percentage of students who achieve Proficient and higher ranking in Grade 3 scores. Instead, this resource focuses on individual students' changes in achievement over time. State progress toward improving how schools serve successive cohorts of students is an important topic, but it is outside the scope of this resource.

While the primary goal of this handbook is to describe how growth measures and growth indicators incorporating AA-AAAS data can be included in state accountability systems, states might use growth measures for other purposes — for example, in State Systemic Improvement Plans (SSIPs) required by the Individuals with Disabilities Education Act (IDEA) (see Blackorby, Taylor, & Wei, 2016). These purposes are addressed briefly in Chapter 3 and Chapter 6.

SUMMARY OF EACH CHAPTER

In **CHAPTER 1**, readers are given an overview of topics pertaining to the education and assessment of students with SCD. Historically, students with SCD had been excluded from statewide assessment; legislation in recent years has altered the experience and expectations for this population. Although students with SCD are now assessed using AA-AAAS, growth measures are often not applied to the test scores of students in this population. Understanding student, instruction and assessment characteristics is important for describing and understanding these students' academic growth. This chapter discusses the state-level definitions used to define who is included in this population, the academic instruction that is typically provided to this population, and the types of alternate assessments that are typically administered to this population.

In **CHAPTER 2**, readers are oriented to growth models, measures and indicators. As noted above, growth **models** are procedures for quantifying or otherwise characterizing academic performance of students over time. In this chapter, the growth models discussed include categorical gain, gain scores, residual gain scores, student growth percentiles, value-added and growth to target. The procedures for applying these growth models are briefly described, along with the data requirements for each model. Growth **measures** are the result of applying growth models to student test data. Growth measures describe changes in academic performance over time. Growth **indicators** summarize growth measures and can be expressed at varying levels for inclusion in school, district, or state accountability systems.

In **CHAPTER 3**, readers are introduced to theories of action and the components of theories of action. A theory of action can establish the goals of an accountability system and describe how these goals can be achieved. Components of a theory of action include goals, purposes and uses, intended outcomes, mediating outcomes and underlying assumptions. Each of these components may help guide the development, documentation and evaluation of growth measures. For states without formal theories of action, the components are related to the type of information states typically have. This way, states can apply the principles involved in a theory of action to their decision-making, even if they do not have an explicit theory of action for their accountability system.

In **CHAPTER 4**, readers are introduced to applications of growth models and growth measures based on AA-AAAS scores. In this chapter, readers are guided through considerations to be made prior to applying growth models to AA-AAAS scores, along with key questions to consider when selecting a growth model to apply to AA-AAAS scores. These considerations and questions include examining technical characteristics of the AA-AAAS as well as the growth model, with the goal of selecting the growth model that is best suited for making the intended inferences with the scores from the specific AA-AAAS. Examples of these considerations are provided using the growth models previously described in Chapter 2.

In **CHAPTER 5**, readers are guided through considerations pertaining to the incorporation of growth measures for AA-AAAS scores as an indicator in an accountability system. These considerations focus on the requirements under Every Student Succeeds Act (ESSA) as well as concerns about how the growth indicator is designed and used. In this chapter, readers are guided through concerns regarding how to design a growth indicator as an additional academic indicator in their accountability system; the technical quality and reliability of the growth indicator; and the intended use and impact of the growth indicator based on the AA-AAAS. Each of these considerations emphasize using the decision-making process producing the growth indicator and the reliability of the individual growth measures to design a growth indicator that demonstrates both policy goal alignment and technical quality.

In **CHAPTER 6**, readers are provided with recommended practices for making decisions about growth models, measures and indicators for alternate assessments. The recommended practices are organized into recommendations during the *design*, *development*, *implementation* and *evaluation-and-revision* phases.

The *design* phase involves recommended practices for designing a growth measure that aligns with the state's theory of action, including identifying and involving key stakeholders in the design process. The *development* phase pertains to recommended practices for developing growth measures and including these growth measures in the accountability growth indicator. The *implementation* phase includes recommended practices for how growth measures are applied and reported. The *evaluation-and-revision* phase involves the review of the results using growth indicators as states engage in monitoring and evaluating their accountability system. A decision-making flowchart and vignettes illustrate how two hypothetical states could follow the decision-making process, and there are recommended practices for applying growth measures in other contexts, including IDEA results-driven accountability, monitoring, evaluation and research.

GUIDING PRINCIPLES AND CAVEATS

The following principles are woven throughout this handbook:

1. Policy uses and intended interpretations should be the primary driver of states' design decisions.
2. The technical adequacy of the AA-AAAS and its scores, the growth measure and the growth indicator must support the state's policy uses and interpretations.
   - ESSA growth indicators are intended to describe or predict growth over time for a group of students. The indicator is part of a larger accountability system and is not used in isolation for high-stakes program evaluation. States should seek a level of technical adequacy for its AA-AAAS growth measure that will allow it to be meaningfully incorporated into the state's growth indicator. This may not require using a complex growth model or the same model used for its general assessment.
3. AA-AAAS growth data may be reported at the individual student level (where families and teachers are key stakeholders) and reported at an aggregated level (to be used by a range of stakeholders). If the data support intended interpretations at the desired reporting level and also support the state's policy goals, states should consider including AA-AAAS in the academic growth indicator to call attention to the issue of equitable treatment of all students.
4. Use discernment with regard to the potential negative consequences of AA-AAAS growth data and/or unintended uses of the data. During the design and development phase, identify potential issues that may be associated with interpretations based on the alternate standards or extensions on the alternate assessment. Mitigate potential challenges associated with the technical constraints of the alternate assessment by matching an appropriate model to the data available. Collect data during operational use to evaluate whether the foreseen risks manifested and whether mitigation plans worked.
5. It is critical that states to pursue this work using an equity lens. As the field continues to evolve, remain open to new ways of thinking about and evaluating equity and new ways to use AA-AAAS growth data to evaluate equity of educational programs and outcomes. Equity should be considered both with regard to how students can access the assessment fairly and how all students are represented in the state's accountability system.
6. It is also important for states to conduct ongoing research that helps them evaluate the efficacy of their growth indicators — including monitoring whether indicators are having the desired effects — and to be prepared to make necessary revisions.

Throughout this handbook, recommendations are grounded in research when it exists. Where evidence is still needed, the authors recommend studies and criteria for evaluating outcomes. It is also recommended that states take a systematic and principled approach to this work, without rushing to apply a particular growth model to their alternate assessment. While including students with SCD in accountability is an important equity issue, the potential pitfalls or unintended consequences of doing so require thoughtful and well considered development and implementation processes.

REFERENCES

Blackorby, J., Taylor, C., & Wei, X. (2016). *Using Growth Models to Measure Child/Student Outcomes for State Systemic Improvement Plans: A Guide for States*. IDEA Data Center. https://files.eric.ed.gov/fulltext/ED584165.pdf

Quenemoen, R. F., & Thurlow, M. L. (2019). *Students with Disabilities in Educational Policy, Practice, and Professional Judgment: What Should We Expect?* (NCEO Report 413). University of Minnesota, National Center on Educational Outcomes. https://nceo.info/Resources/publications/OnlinePubs/report413/default.html

CHAPTER

1

FUNDAMENTAL CONCEPTS

INTRODUCTION

This chapter provides background information about students with significant cognitive disabilities (SCD), expectations for academic achievement and alternate assessments. The chapter also summarizes federal legislation on alternate assessment, reporting and accountability. A basic understanding of these topics is necessary before states can begin considering whether or how to calculate alternate assessment growth measures or include alternate assessment growth in their accountability growth indicator.

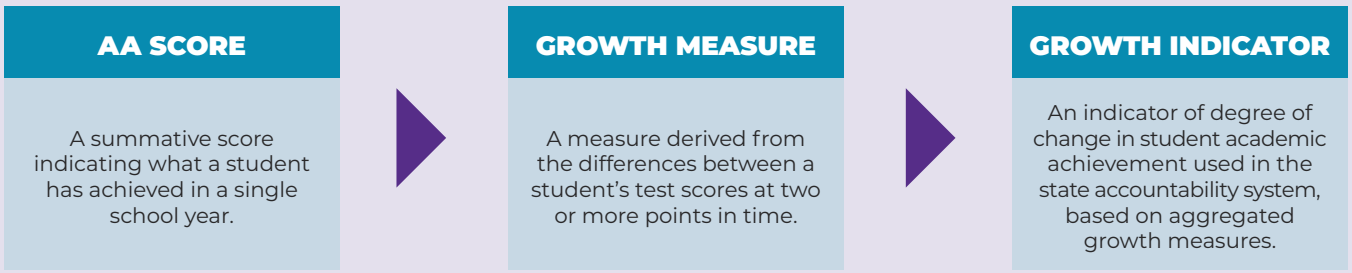
Before 2000, students with the greatest SCD were excluded from statewide assessment systems. Their K-12 curricular priorities often focused on functional and life skills that adults in their lives believed they would need after high school to have a good quality of life as they participated in daily activities in their homes and communities. The few who were employed tended to be based in sheltered workshops completing repetitive tasks for sub-minimum wage. Between 2000 and 2020, reauthorizations of key legislation have led to dramatically different educational experiences and expectations for this population.

Academic expectations for these students have shifted, and they are now learning academic concepts more aligned to their peers. Students with SCD take state-required alternate assessments in academic content, based on alternate academic achievement standards (AA-AAAS); and alternate English language proficiency assessments (AA-AELPS). By the time they leave high school, they will be better prepared academically to pursue postsecondary education and competitive integrated employment, which includes work at minimum wage or higher, with wages and benefits similar to those without disabilities who perform comparable work.

One way that states ensure that students are prepared for postsecondary opportunities is to monitor their academic achievement using statewide assessments. State accountability systems include measures of student status (i.e., proficiency) and growth (i.e., change over time). However, states have experienced challenges in developing reliable and valid growth measures for alternate assessments and have raised questions about how to include those measures in growth indicators for state accountability systems.

Growth is defined as change over time for individual students or groups of students. **Growth measures** are a way of quantifying or otherwise describing the amount of change in achievement over time, for individual students or groups of students. Growth measures are produced by applying **growth models** (e.g., categorical gain, student growth percentile) to assessment scores. **Growth indicators** aggregate information from growth measures for use in accountability systems, so states can monitor their progress toward achieving long-term goals (see Figure 1-1).

FIGURE 1-1: RELATIONSHIP AMONG ASSESSMENT SCORE, GROWTH MEASURE & GROWTH INDICATOR



Given the short history of alternate assessments and the rapid changes in alternate assessment design in the past 20 years, there is very limited literature on growth measures based on alternate academic achievement standards (AA-AAAS) and few examples in practice. In 2016, 14 of the 19 state education agencies (SEA) CCSSO surveyed did not include growth for alternate assessments in their accountability models (Domaleski & Hall, 2016). As described in Chapter 2 of this handbook, seven states responding to a 2021 survey indicated that they included AA-AAAS growth in their accountability academic growth indicator.

WHO ARE THE STUDENTS WITH SCD?

For nearly two decades students with significant cognitive disabilities (SCD) were defined by their eligibility to take statewide alternate assessments, based on alternate academic achievement standards. Only in recent years have states begun defining the population of students eligible to take alternate assessments (Thurlow et al., 2019). Each student's individualized education program (IEP) team uses participation guidelines to decide whether a student with a significant cognitive disability should participate in the general or alternate large-scale assessment.

"Significant cognitive disability" is not a category recognized under Individuals with Disabilities Education Act (IDEA). States, national organizations and researchers have made efforts over years to describe the population and provide criteria to help IEP teams decide whether a student meets the definition of having "significant cognitive disabilities." Under No Child Left Behind (NCLB)-era accountability regulations, no more than 1% of students could be included as scoring Proficient or above on alternate assessments. This regulation was broadly interpreted and allowed for states to test more than 1% using an alternate assessment while counting only 1% as Proficient.

In 2015, the Every Student Succeeds Act (ESSA) referred to the population as "students with *the most* significant cognitive disabilities" [emphasis added] and changed the threshold: now no more than 1% of students are allowed to **participate** in alternate assessments. In many states, this change has prompted efforts and forced them to carefully monitor districts and reduce the number of students participating in the alternate assessment. As a result, states have narrowed their participation criteria, changed their guidance to IEP teams, and provided technical assistance to districts on use of the guidelines.

Given that there is no federal definition of a "student with a significant cognitive disability" many states have defined the term for their districts. The National Center on Educational Outcomes (NCEO) conducted a scan of the states' guidance documents in 2018-2019 (Thurlow et al., 2019). Among the 35 states that defined students with SCD, common characteristics of the definitions included:

- Significant cognitive deficits (35 states)
- Poor adaptive skills (34 states)
- Pervasive needs across setting or time (23 states)
- Based on holistic criteria, not just an IQ score (19 states)
- Receiving extensive, individualized direct instruction (16 states)

These definitions tend to be closely related to the states' alternate assessment participation criteria (Thurlow et al.).

Education agencies from all 50 states and the District of Columbia have alternate assessment participation criteria; 46 of those states include guidance on factors that should **not** lead to a decision that a student should participate in the alternate assessment. In the same study, Thurlow et al. (2019) found that, among the 46 states describing these factors in their participation criteria, the most common were:

- disability label, placement or service (45 states)
- social, cultural, linguistic or environmental factors (45 states)
- excessive absences (44 states)

Researchers have used teacher surveys such as the Learner Characteristics Inventory or the First Contact Survey to describe the population of students who take alternate assessments, in terms of their disability-related characteristics which impact learning and assessment. A small percentage of students with SCD also have sensory and/or physical disabilities. For example, of the more than 90,000 students in 19 states taking Dynamic Learning Maps (DLM) alternate assessments in 2018-2019, approximately 5% were identified as blind or with low vision; approximately 4% were identified as deaf or hard of hearing; and approximately 35% were identified as limited hand mobility and/or required assistance to perform tasks with their hands (Burnes & Clark, 2021).

Students with SCD also communicate in a variety of ways, and many have complex communication support needs that impact the design of their instruction. Based on the 2018-2019 DLM data set, approximately 25% of students with SCD do not use speech to communicate (Burnes & Clark, 2021). About 25% use an augmentative or alternative communication (AAC) device in combination with or instead of speech or sign language. Nearly half of the students with SCD (47%) use some form of symbols to communicate, and about 5% use gestures, vocalizations, and behaviors to communicate (Burnes & Clark, 2021). Students who use speech to communicate tend to be placed in more inclusive settings and have more sophisticated academic and communication skills (Erickson & Geist, 2016).

While more information was available about this population in 2020 than in 2000, there is still room for improvement in understanding the characteristics which impact their learning. For example, students with SCD who are also English learners may be under-identified (Karvonen & Clark, 2019), and students with known or suspected dual sensory loss tend to have fewer or less complex academic skills than students without dual sensory loss (Karvonen et al., 2021). There is also the potential for sensory impairments such as Cortical Visual Impairment to be under-identified, which can impact both instruction and assessment (Karvonen et al., 2021).

What does academic instruction for students with SCD look like?

Common alternate assessment participation criteria include instruction based on alternate or extended content or curriculum standards and extensive individualized instruction or supports (Thurlow et al., 2019). In this situation, what are the academic expectations and what does instruction look like?

Academic Content Standards

Students with SCD are instructed using alternate or extended content standards, which are based on a state's grade-level academic content standards but reduced in depth, breadth and complexity to provide expectations for what students with SCD should learn that year. Some states do not adopt extended content standards but instead develop curriculum frameworks based on general academic content standards to guide instruction. Regardless of whether a state labels the expectations as extended content standards, those expectations are designed to provide access to the academic concepts with an appropriate level of rigor. The table in Figure 1-2 shows examples of those extensions.

FIGURE 1-2: EXAMPLE STATE CONTENT STANDARDS & EXTENDED STANDARDS

STATE CONTENT STANDARD	EXTENDED STANDARD FOR STUDENTS WITH SCD
SOL 9.7 The student will self- and peer-edit writing for capitalization, punctuation, spelling, sentence structure, paragraphing, and Standard English. a) Use parallel structure across sentences & paragraphs. b) Use appositives, main clauses, & subordinate clauses. c) Use commas and semicolons to distinguish & divide main & subordinate clauses. d) Distinguish between active & passive voice. e) Use a variety of sentence structures to infuse sentence variety in writing.	HSE-WE 1 The student will: a) Use standard English rules by using correct punctuation when writing. b) Spell most single-syllable words correctly and apply knowledge of word chunks in spelling longer words.
6.G.A.3: Draw polygons in the coordinate plane given coordinates for the vertices; use coordinates to find the length of a side joining points with the same first coordinate or the same second coordinate. Apply these techniques in the context of solving real-world and mathematical problems.	6.GM.1c7: Use coordinate points to draw polygons.
5-PS1-2: Measure & graph quantities to provide evidence that, regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved.	EE.5-PS1-2: Measure and compare weights of substances before and after heating, cooling, or mixing substances to show that weight of matter is conserved.

Sources: ELA = VAAP Aligned Standards of Learning, Math = NCSC Core Content Connectors, Science = DLM Essential Elements

States may provide additional guidance on content progressions leading up to the grade-level performance expectation in the extended standard, to help teachers design instruction for the full range of students with SCD. For example, Washington offers Access Point Frameworks with three levels of complexity for each extended standard. The Dynamic Learning Maps (DLM) Consortium provides "mini-maps" for each content standard to show skills at five complexity levels, leading up to and extending past the grade-level expectation.

States design their grade-level expectations for students with SCD so that expectations increase across grades, with the goal that students are better prepared for life after high school. Expectations for what students with SCD should learn by the end of high school vary across states. Most research on academic learning for this population is limited to skill acquisition within a single grade. More research is needed to support common expectations for what students should know and be able to do by the end of high school.

Instruction

Special educators consider their students' characteristics when designing instruction. For example, teachers adapt materials based on students' sensory and physical characteristics and the ways they receive information. Teachers also must make sure that students with SCD have ways to demonstrate what they know and can do during instruction. For instance, students who do not use speech to communicate may need to answer questions using a communication board, sign, gesture or other means. For the approximately 10% of students with SCD who have not yet developed skills in symbolic communication (Kearns et al., 2011), academic instruction should include methods for building the student's communicative competence (Kleinert et al., 2019).

Although common approaches are 1:1 or small-group instruction, more than one-third of all students with SCD reportedly demonstrate sustained attention to computer-directed instruction (Burnes & Clark, 2021). Nearly all students with SCD who take alternate assessments can access a computer independently or with support from assistive devices or other people (Burnes & Clark, 2021). Special educators, who are given a great deal of autonomy to design appropriately adapted academic instruction, may integrate academic instruction with other curricular goals. For example, a student with SCD might learn to apply addition and subtraction skills while making purchases during community-based instruction.

While states design extended standards to show vertical articulation (i.e., the increase in expectations as students progress through upper grades), there is very little research on how students with SCD learn academics that build over time. Less is known about expected rates of academic learning for this population, which in part is due to the lack of understanding about the cognition of students with SCD (Kleinert et al., 2009).

The current models of evidence-based academic instruction for students with SCD (Browder et al., 2020a, 2020b) are grounded in a long history of behavioral approaches to teaching discrete skills. These practices do not promote students' conceptual learning of the more comprehensive academics needed to reach grade-level academic expectations. Some research has evaluated more comprehensive instructional approaches targeting multiple student outcomes (e.g., Allor et al., 2010), and resources are emerging to help the field shift its instructional practices (e.g., Browder et al., 2020a; Erickson & Koppenhaver, 2020). However, more research is needed in this area to support effective instruction which ultimately will help students with SCD meet higher expectations.

One approach to helping students with SCD attain higher academic expectations is to provide inclusive instruction. Unfortunately, fewer than 5% of students with SCD are fully included with their peers in general education classrooms (Burnes & Clark, 2021); students with SCD are more often served in segregated settings (Agran et al., 2020). One study that analyzed student IEPs found similar inclusion rates for academic and nonacademic subjects (fewer than 30% of students); it also found that 30% of students with SCD were explicitly excluded from academic instruction in the general classroom (Kurth et al., 2019). Students with SCD who do not use speech to communicate are more likely than their peers to be served in separate schools (Erickson & Geist, 2016).

In more segregated instructional settings, responsibility for designing and delivering academic instruction falls to special educators, many of whom do not have certification in academic subjects. Students with SCD may be left out of the schoolwide tiered support models (e.g., multi-tiered systems of support) which are designed for early identification of students at risk for academic failure and include tiers of increasingly intensive interventions (Thurlow et al., 2020). Receiving academic instruction based on individualized targets, alongside their age-appropriate peers, has academic, social and behavioral benefits for students with SCD (Morningstar et al., 2016). Several state- and federally-funded initiatives are in place to improve inclusive academic instruction for students with SCD. If successful, these initiatives will provide more opportunities for teachers to adapt to having higher expectations for students with SCD and for students with SCD to meet those expectations.

What does the alternate assessment for students with SCD look like?

Under IDEA, alternate assessments were first required in 2000-2001. Early alternate assessments were based on functional curricula (e.g., communication, vocational skills, community integration) and included some functional academic skills. In response to the The No Child Left Behind Act of 2001, states shifted to alternate assessments based on alternate academic achievement standards; for the first time, these assessments included science. Between 2001 and 2010, alternate assessments took the form of portfolios, checklists or structured performance tasks administered 1:1. There was a tension between assessment designs which emphasized standardization of content and administration procedures (as is more typical of large-scale assessments) and assessment designs which emphasized flexibility in defining the assessment targets and administration procedures — i.e., to be consistent with what students experienced during instruction (Gong & Marion, 2006).

States struggled to align alternate assessments to state content standards and develop scoring models that incorporated non-student/academic factors (e.g., participation in activities with nondisabled peers) into their assessment results. For example, a student with SCD might be Proficient on an alternate assessment, not because of correct responses but because of program quality indicators. States also struggled to define standards of acceptable technical adequacy for alternate assessments. They also found it challenging to devise methods for evaluating technical quality which would account for the assessment characteristics or small population size.

As states adopted more challenging college- and career-ready content standards for students with SCD, the U.S. Department of Education (ED) used the Race to the Top initiative to fund two assessment consortia on developing next-generation academic assessments. ED also funded two projects to develop new alternate assessments: the National Center and States Collaborative (NCSC, now called the Multistate Alternate Assessment or MSAA) and Dynamic Learning Maps (DLM). These new, computer-based assessments aligned to extensions of the new content standards. They also used Universal Design for Learning principles to balance standardization with flexibility. The resulting scores are based solely on students' correct responses to items, not program quality indicators.

Both assessment systems are based on theories of action which describe expectations for students with SCD to meet higher expectations over time and to leave school better prepared for post-secondary opportunities (Forte et al., 2016; Karvonen et al., 2020). The new assessments also brought new approaches for evaluating technical quality. This type of evidence has met criteria set forth in ED peer review requirements (U.S. Department of Education, 2018). All of these changes made it possible to trust the inferences of what students know and can do, based on their alternate assessment results.

States use a variety of scoring methods to calculate AA-AAAS scores. These forms typically include raw scores (converted to scale scores), item response theory-based scale scores, diagnostic classification model mastery profiles, and performance levels.

- Raw scores may be converted to a scale score via standard scores (i.e., a z-score distribution with a mean of 0 and standard deviation of 1). A state may set the scale mean at any value and use a conversion table to transform all possible raw score values into scale score values.
- Item response theory-based models generate scaled scores in the form of numeric estimates of a student's ability in the underlying latent trait (typically a unidimensional academic domain, like math).
- Currently used only by the DLM Consortium, diagnostic classification models (DCM; Rupp, Templin, & Henson, 2010) are used to generate mastery profiles that indicate the knowledge, skills and understandings that students with SCD demonstrate on the assessment. The academic content is assumed to be multidimensional, and each student receives a profile of specific skills mastered.
- All types of alternate assessments also report results based on alternate academic achievement or performance levels, which are categorical descriptors of student performance. Performance levels are usually derived by applying cuts to score ranges (or in the case of DCM, mastery profile ranges) for assessments scored using one of the previously described scoring methods.

A state that adopts alternate academic achievement standards is responsible for ensuring that those achievement standards:

1. are aligned with the challenging state academic content standards;
2. promote access to the general education curriculum;
3. reflect professional judgment as to the highest possible standards achievable by such students;
4. are designated in the IEP as the academic achievement standards that will be used for the student; and
5. are aligned to ensure that a student who meets the alternate academic achievement standards is on track to pursue postsecondary education or employment.

Source: 34 CFR 200.1(d)

Requirements 1, 3 and 5 are especially relevant to the potential use of achievement standards in growth measures for students with SCD.

INCLUDING STUDENTS WITH SCD IN STATEWIDE ASSESSMENT & ACCOUNTABILITY

Why does it matter if we include students with SCD in assessment and accountability?

Fundamentally, inclusion of individuals with disabilities is a civil rights issue. The 1997 IDEA requirement to include students with disabilities in the general curriculum was meant to shift from a curriculum based almost exclusively on functional life skills to one addressing knowledge and skills needed for positive post-school outcomes. Including students with SCD in the statewide assessments used for accountability is intended to ensure that these students' test results are considered when states and districts use accountability results to drive decisions about educational programming and resource allocation. Inclusive assessment systems ensure equitable treatment of students with SCD in reporting and accountability (Thurlow et al., 2001).

What is the federal guidance on calculating & reporting growth measured by alternate assessments?

This chapter describes guidance based on the Elementary and Secondary Education Act (ESEA). It is important that state education leaders consider how any state assessment or accountability requirements layer onto these federal requirements.

Federal guidance on this topic has changed several times since ESEA was reauthorized in 2015 as ESSA.

- Requirements for alternate assessments are documented in the regulations (34 CFR 200), most recently updated in July 2019 and in peer review guidance for state assessment systems.
- ESSA accountability regulations finalized in 2016 were rescinded through the Congressional Review Act in 2017 and removed from regulations in 2018. Current accountability guidance is based on ESSA legislation, as well as the 2017 versions of the State Consolidated Plan workbook template and peer review criteria for state plans.

Requirements for Alternate Assessment

- Whenever a state provides a general assessment, it must also provide an alternate assessment. At a minimum, this includes annual assessments in Grades 3-8 and once in high school for ELA and math, plus at least three grades in science (once each in Grades 3-5, 6-9, and 10-12). (34 CFR 200.5(a))
- The number of students with SCD taking alternate assessments cannot exceed 1% of all students who are assessed statewide in each subject. (34 CFR 200.6(c)(2))
- States are responsible for reporting the number and percentage of students with SCD who take alternate assessments (34 CFR 200.6(c)(5)).
- States must report the performance of students with SCD relative to alternate academic achievement standards (34 CFR 200.6(c)(1)(ii))
  - There is no requirement that a state's general and alternate academic assessments must conform to reporting results at the same grain size (i.e., overall score vs. sub-scores).
- To help ensure that assessment results can be used to improve student instruction, states have the discretion to report valid and reliable measures of growth for students with SCD at all alternate academic achievement levels (34 CFR 200.6(c)(1)(iii))

Requirements for Accountability

- States must establish the indicators they will use to differentiate schools annually and monitor their interim progress toward long-term goals for improving academic achievement. 1111(c)(4)(B)
  - Each indicator must be reported for all students and for each subgroup, including students with disabilities. 1111(c)(4)(B)
  - Academic achievement indicators include:
    - For all public schools in the state, based on the long-term goals established under subparagraph (A), academic achievement:
      - as measured by proficiency on the annual assessments required under subsection (b)(2)(B)(v)(I); and
      - at the state's discretion, for each public high school in the state, student growth, as measured by such annual assessments.
    - For public elementary schools and secondary schools that are not high schools in the state:
      - a measure of student growth, if determined appropriate by the state; or
      - another valid and reliable statewide academic indicator that allows for meaningful differentiation in school performance.
- Annual meaningful differentiation of schools must be based on all indicators, for all students and each subgroup of students; and must differentiate schools in which students in any subgroup are consistently underperforming (1111(c)(4)(C))

The table in Figure 1-3 summarizes the requirements and options for assessment and accountability.

FIGURE 1-3: SUMMARY OF FEDERAL REQUIREMENTS FOR ALTERNATE ASSESSMENT SCORE REPORTING & USE OF ALTERNATE ASSESSMENT RESULTS IN ACCOUNTABILITY PLANS

	ALTERNATE ASSESSMENT	STATE ACCOUNTABILITY PLANS
STATES MUST	• Provide assessments in the same grades and subjects as the general assessments. • Limit participation to 1% of all students who are assessed in tested grades. • Report participation. • Report performance.	• Include students with disabilities as a subgroup for all indicators. • Establish long-term goals for the state to improve achievement; goals must include alternate assessment proficiency. • For elementary and middle schools, report a measure [indicator] of student growth or another valid and reliable indicator that supports meaningful differentiation of schools.
STATES MAY	• Report valid and reliable measures of growth to inform instruction.	• For high schools, report a measure of student growth based on assessment scores.

To meet the goals of fully including students with SCD in assessment and accountability, it makes sense to strive for systems in which alternate assessment results are reported and included as part of the subgroup of students with disabilities. If a state reports measures of growth to inform instruction for students in the general assessment, it should consider the same for students taking the alternate assessment. If a state plan includes an indicator based on growth for students over time, it makes sense to include alternate assessment results in calculations for all students and for students with disabilities. However, if states cannot produce student growth measures or growth indicators that are reliable and valid, they risk unintended consequences.

REFERENCES

Agran, M., Jackson, L., Kurth, J. A., Ryndak, D., Burnette, K., Jameson, M., Zagona, A., Fitzpatrick, H., & Wehmeyer, M. (2020). Why Aren't Students with Severe Disabilities Being Placed in General Education Classrooms: Examining the Relations Among Classroom Placement, Learner Outcomes, and Other Factors. *Research and Practice for Persons with Severe Disabilities, 45*(1), 4–13. https://doi.org/10.1177/1540796919878134

Allor, J. H., Mathes, P. G., Roberts, J. K., Jones, F. G., and Champlin, T. M. (2010). Teaching Students with Moderate Intellectual Disabilities to Read: An Experimental Examination of a Comprehensive Reading Intervention. *Education and Training in Autism and Developmental Disabilities, 45*(1), 3-22. http://www.jstor.org/stable/23880147

Browder, D. M., Spooner, F., & Courtade, G. R. (with B. Jimenez) (2020a). Science and Engineering Practices. In *Teaching Students with Moderate and Severe Disabilities* (2nd ed., pp. 255-275). The Guilford Press. https://www.guilford.com/books/Teaching-Students-with-Moderate-and-Severe-Disabilities/Browder-Spooner-Courtade-Contributors/9781462542383

Browder, D. M., Spooner, F., & Courtade, G. R. (with P. Mims) (2020b). Using Instruction That Works: Evidence-Based Practices. In *Teaching Students with Moderate and Severe Disabilities* (2nd ed., pp. 114-140). The Guilford Press. https://www.guilford.com/books/Teaching-Students-with-Moderate-and-Severe-Disabilities/Browder-Spooner-Courtade-Contributors/9781462542383

Burnes, J. J., & Clark, A. K. (2021). *Characteristics of Students Who Take Dynamic Learning Maps® Alternate Assessments* (Technical Report No. 20-01). University of Kansas, Accessible Teaching, Learning, and Assessment Systems. https://dynamiclearningmaps.org/sites/default/files/documents/publication/Characteristics_of_Students_Who_Take_DLM_AAs.pdf

Domaleski, C. & Hall, E. (2016). *Guidance for Estimating and Evaluating Academic Growth*. National Center and State Collaborative. http://www.ncscpartners.org/Media/Default/PDFs/Resources/EstimatingandEvaluatingStudentGrowth.pdf

Erickson, K. A., & Geist, L. A. (2016). The profiles of students with significant cognitive disabilities and complex communication needs. *Augmentative and Alternative Communication*, *32*(3), 187-197. http://dx.doi.org/10.1080/07434618.2016.1213312

Erickson, K. A., & Koppenhaver, D. A. (2020). *Comprehensive Literacy for All: Teaching Students with Significant Disabilities to Read and Write*. Paul H. Brookes Publishing. https://products.brookespublishing.com/Comprehensive-Literacy-for-All-P1165.aspx

Forte, E., Quenemoen, R. F., & Thurlow, M. L. (2016). *NCSC's Theory of Action and Validity Evaluation Approach* (NCSC Brief No. 9). University of Minnesota, National Center and State Collaborative. http://www.ncscpartners.org/Media/Default/PDFs/Resources/NCSCBrief9.pdf

Gong, B., & Marion, S. (2006). *Dealing with Flexibility in Assessments for Students with Significant Cognitive Disabilities* (Synthesis Report 60). University of Minnesota, National Center on Educational Outcomes. https://files.eric.ed.gov/fulltext/ED495895.pdf

Karvonen, M., Burnes, J. J., Clark, A. K., & Kavitsky, L. (2020). *Aligned Academic Achievement Standards to Support Pursuit of Postsecondary Opportunities: Instructionally Embedded Model* (Technical Report No. 20-02). University of Kansas; Accessible Teaching, Learning, and Assessment Systems. https://tinyurl.com/dlm-pso-ie

Karvonen, M., & Clark, A. K. (2019). Students with the Most Significant Cognitive Disabilities Who Are Also English Learners. *Research and Practice for Persons with Severe Disabilities, 44*(2), 71–86. https://doi.org/10.1177/1540796919835169

Karvonen, M., Beitling, B., Erickson, K., Morgan, S., & Bull, R. (2021). *Students with Significant Cognitive Disabilities and Dual Sensory Loss*. Co-published by Accessible Teaching, Learning, and Assessment Systems and the National Center on Deaf-Blindness. University of Kansas, Accessible Teaching, Learning, and Assessment Systems (ATLAS), National Center on Deaf-Blindness (NCDB). https://www.nationaldb.org/media/doc/NCDB_Student_Characteristics_Report_a.pdf

Kearns, J. F., Towles-Reeves, E., Kleinert, H. L., Kleinert, J. O., & Thomas, M. K. (2011). Characteristics of and implications for students participating in alternate assessments based on alternate academic achievement standards. *The Journal of Special Education, 45*(1), 3–14. https://doi.org/10.1177/0022466909344223

Kleinert, H. L., Browder, D. M., & Towles-Reeves, E. A. (2009). Models of Cognition for Students with Significant Cognitive Disabilities: Implications for Assessment. *Review of Educational Research, 79*(1), 301-326. https://doi.org/10.3102/0034654308326160

Kleinert, J., Kearns, J., Liu, K. K., Thurlow, M. L., & Lazarus, S. S. (2019). *Communication Competence in the Inclusive Setting: A Review of the Literature* (TIES Center Report 103). University of Minnesota, The TIES Center. https://ici-s.umn.edu/files/K-gxmnGDc_/ties-center-report-103

Kurth, J. A., Ruppar, A. L., Toews, S. G., McCabe, K. M., McQueston, J. A., & Johnston, R. (2019). Considerations in Placement Decisions for Students with Extensive Support Needs: An Analysis of LRE Statements. *Research and Practice for Persons with Severe Disabilities*, *44*(1), 3–19. https://journals.sagepub.com/doi/10.1177/1540796918825479

Morningstar, M. E., Allcock, H. C., White, J. M., Taub, D., Kurth, J. A., Gonsier-Gerdin, J., Ryndak, D. L., Sauer, J., & Jorgensen, C. M. (2016). Inclusive Education National Research Advocacy Agenda: A Call to Action. *Research and Practice for Persons with Severe Disabilities*, *41*(3), 209–215. https://journals.sagepub.com/doi/10.1177/1540796916650975

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. The Guildford Press. https://psycnet.apa.org/record/2010-13188-000

Thurlow, M. L., Ghere, G., Lazarus, S. S., & Liu, K. K. (2020). *MTSS for All: Including Students with the Most Significant Cognitive Disabilities*. University of Minnesota, National Center on Educational Outcomes/TIES Center. https://nceo.umn.edu/docs/OnlinePubs/NCEOBriefMTSS.pdf

Thurlow, M. L., Lazarus, S. S., Albus, D. A., Larson, E. D., & Liu, K. K. (2019). *2018-19 Participation Guidelines and Definitions for Alternate Assessments Based on Alternate Academic Achievement Standards* (NCEO Report 415). University of Minnesota, National Center on Educational Outcomes. https://nceo.info/Resources/publications/OnlinePubs/report415/default.html

Thurlow, M., Quenemoen, R., Thompson, S., & Lehr, C. (2001). *Principles and Characteristics of Inclusive Assessment and Accountability Systems* (NCEO Synthesis Report 40). University of Minnesota, National Center on Educational Outcomes. https://nceo.info/Resources/publications/OnlinePubs/Synthesis40.html

U.S. Department of Education. (2018). *A State's Guide to the U.S. Department of Education's Assessment Peer Review Process*. https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf

CHAPTER

2

FOUNDATIONS: GROWTH MODELS, MEASURES & INDICATORS

INTRODUCTION

Chapter 1 provided background information on using alternate academic achievement standards (AA-AAAS) in growth models. It described the students with significant cognitive disabilities (SCD) who take the AA-AAAS and their education, and it reviewed both the educational rationales and legal requirements for including these students in assessment and accountability. As noted in Chapter 1, there is very little literature about the rates of academic learning for students with SCD; and there is a need for research into these students' academic growth and the conditions that facilitate their learning.

The goal of Chapter 2 is to orient readers to growth models, growth measures and growth indicators from a conceptual perspective. This chapter does not go into detail about how to construct growth models, measures or indicators; instead, it highlights the technical and conceptual considerations and describes states' practices, as of 2020-2021. Chapter 4 and Chapter 5 will address specific considerations and applications and provide examples related to selecting growth models, developing growth measures and using growth indicators for alternate assessments.

Important Distinctions in Interpretations of Test Score Data

There are three key distinctions in the uses and interpretations of test score data that should be made before considering growth models, measures and indicators (D'Brot, 2017):

STATUS (aka proficiency level)	*IMPROVEMENT*	*GROWTH*
The academic performance of a student or collection of students at a single point in time.	The change in performance at the aggregate level over time within or across grades, without following the same student or collection of students.	The change in academic performance of the same student or same collection of students over two or more points in time.

Growth characterizes the academic performance of individual students or cohorts of students over time (Castellano & Ho, 2013), and ***academic growth*** is broadly defined as change in the amount and/or depth of content knowledge, skills and understandings (KSU) over time. For accountability purposes, academic growth is defined by student test scores. Figure 2-1 depicts the relationship among growth models, measures and indicators.

FIGURE 2-1: RELATIONSHIP AMONG GROWTH MODELS, MEASURES & INDICATORS*

*Test scores from more than two time periods may enter into a growth model. The minimum is two scores.

To foster understanding of this chapter, readers are encouraged to consider the following contextual definitions:

Growth model: A growth *model* is a procedure used to quantify or otherwise characterize the academic performance of students from two or more points in time. The application of a growth model to test data results in a growth measure.

Growth measure: A growth *measure* delineates the degree of change in a student's KSUs over time. For school accountability purposes, growth measures for individual students are summarized at the school, district and/or state levels. For example, one possible summary is that 45% of students made expected growth; another possible summary is that the growth score for the school is the mean student growth percentile of 57. Growth measures also may be reported at the student level — for example, on a student score report.

Growth indicator: A growth *indicator* is a value used in the educational accountability system to indicate the degree to which students in a school, district or state grew academically. If the state uses multiple growth models (for example, one for the general assessment and one for the alternate assessment), the indicator may be based on a combination of summary growth measures. The indicator may take different forms, depending on the nature of the accountability system — for example, some states use index systems, transforming the growth summary into an index score.

WHAT ARE GROWTH MODELS?[1]

Growth models are procedures for estimating student academic growth based on test scores. The models commonly used in state accountability systems are (Castellano & Ho, 2013; D'Brot, 2017; Domaleski & Hall, 2016):

- categorical gains
- gain scores
- residual gain scores
- student growth percentiles
- value-added models
- growth-to-target

Growth-to-target models are used in addition to an underlying growth model — for example, a residual gains model may impose a target on the gain scores to define how much growth is sufficient. The choice of a growth model is constrained by the nature of the test scores (e.g., whether the scores are vertically scaled across grades so that scale scores can be directly compared from grade to grade) and the intended use of the growth measure, although other considerations such as sample size and sample variability are also important (Castellano & Ho, 2013; Domaleski & Hall, 2016; Nehler et al., 2019).

Because score use and interpretation are the primary considerations in selecting a growth model once data limitations are addressed, each of the following models is accompanied by a brief description of its primary interpretation.

[1] For an in-depth discussion of growth models, see *A Practitioner's Guide to Growth Models* (Castellano and Ho, 2013).

Categorical Gain: Degree of Student Growth Related to Achievement Levels

Categorical gains are summaries of changes in student performance measured at the ordinal level over time (Domaleski & Hall, 2016). For example, consider an assessment with four achievement levels. A student performing at Level 2 in third grade and Level 3 in fourth grade could have their growth summarized as +1 using categorical gains. For more fine-grained measures of growth, the state could subdivide the achievement levels in a meaningful way to address growth within a level — for example, from low Partially-Proficient to high Partially-Proficient.

Categorical gains can be extended to cohorts of students by summarizing the proportion of students with positive transitions or by taking the average transition value for the cohort of students (Nehler et al., 2019). Figure 2-2 illustrates a basic categorical gains model. Note that, in this example, maintaining Proficient status across years counts as meeting the growth criterion. In this case, meeting grade-level expectations from year to year is considered sufficient growth in KSUs. Students scoring below the Proficient level, however, are required to gain at least one achievement level the following year.

FIGURE 2-2: A SIMPLE CATEGORICAL GAINS MODEL

Did the student meet the growth criterion?

PRIOR YEAR ACHIEVEMENT LEVEL	CURRENT YEAR ACHIEVEMENT LEVEL: Below Proficient	Approaching Proficient	Proficient	Advanced
Below Proficient	No	**Yes**	**Yes**	**Yes**
Approaching Proficient	No	No	**Yes**	**Yes**
Proficient	No	No	**Yes**	**Yes**
Advanced	No	No	No	**Yes**

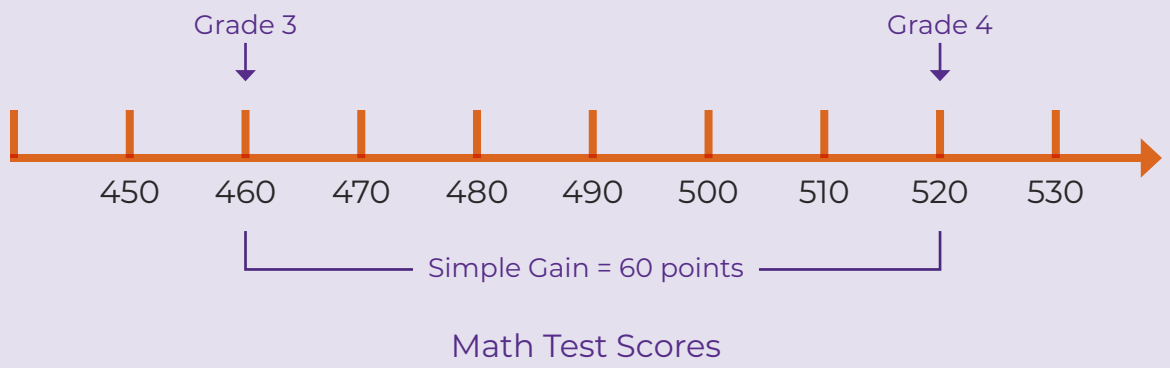
Assessments which are scored using diagnostic classification modeling (DCM) do not produce interval scores; therefore, student achievement levels do not reflect performance in the same ways as they do on other assessments, and typical categorical gain scores would not be appropriate. Transition DCMs (Madison & Bradshaw, 2018a; Madison & Bradshaw, 2018b) have not yet been applied in statewide assessments, and additional research is needed to extend transition DCMs to yield information about gains in overall achievement.

Gain Scores: Magnitude of Student Growth in Terms of Scale Scores

Gain scores are the difference between student test scores at Time One and Time Two. Simple gain scores quantify academic performance by calculating the magnitude and direction of the numeric change in test scores on a vertically scaled assessment (Castellano & Ho, 2013). For example, on a test that is vertically scaled across grades, a student with a score of 460 on the Grade 3 math assessment and a score of 520 on the Grade 4 math assessment would have a simple gain score of 60 in Grade 4. (See Figure 2-3 for an illustration of this model.)

One caveat to keep in mind when interpreting simple gain scores is that vertical scaling does not produce equal interval scales across grades — i.e., the same score gain between pairs of grades may not mean the same degree of learning took place. Using the example in Figure 2-3, a change of 60 points between Grade 4 and Grade 5 may not be comparable to the change of 60 points between Grade 3 and Grade 4 being illustrated.

FIGURE 2-3: SIMPLE GAIN SCORE MODEL



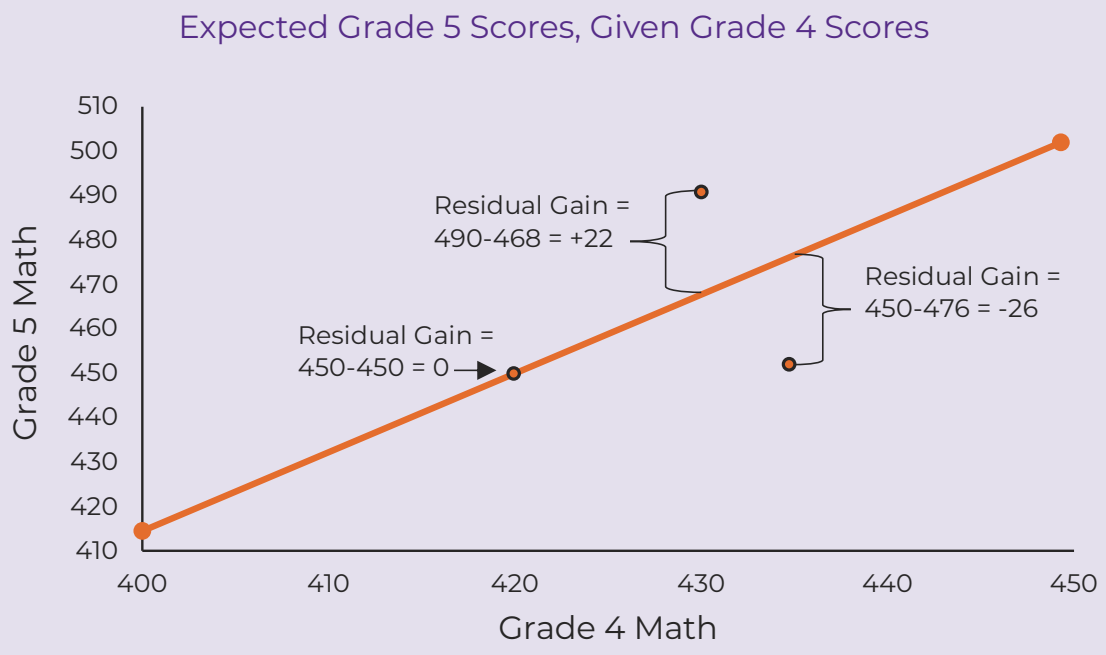
Residual Gain Scores: Degree of Student Growth as Compared to Expected Growth, With Expected Growth Based on Prior Score

Residual gain scores use linear regression to estimate whether a student's score differed from what was expected. That is, was a student's score greater than, less than, or equivalent to what was expected, given on one or more previous year's test scores (Castellano & Ho, 2013). The difference between their performance and the expectation is the *residual*. This residual can be used as a measure of the degree of student growth. For example, a residual score of 15 means the student's current test score is 15 points higher than expected, given the student's prior year(s) performance; a residual gain score of 0 means the student met expectations; and a negative residual gain score means the student scored lower than expected given the student's performance in the prior year or years.

Figure 2-4 illustrates a residual gain model. The solid line is the linear regression of Grade 5 scores on Grade 4 scores and expresses the expected value of a student's Grade 5 score, given their Grade 4 score. For example, a student who scored 420 in Grade 4 is expected to score 450 in Grade 5; if this student scored 450 in Grade 5, their residual gain score is 0. If a student who scored 435 in Grade 4 earned a score of 450 in Grade 5, they would not meet their expected score of 476; this student's residual gain score would be -26. A student who earns higher than their expected score has a positive residual gain. For example, a student whose expected score is 468 and who earns a Grade 5 score of 490 has a residual gain score of +22.

The residual gain score can be mapped to achievement standards (Blackorby, Taylor, & Wei, 2016) to determine whether the student is making enough growth to meet a pre-defined target in the future (see growth-to-target section below). Residual gain scores can also be presented as a measure of growth for cohorts of students. Residual gains are more often used as a part of other growth models (e.g., value-added) than as a stand-alone model (Data Quality Campaign, 2019).

FIGURE 2-4: RESIDUAL GAIN SCORE MODEL



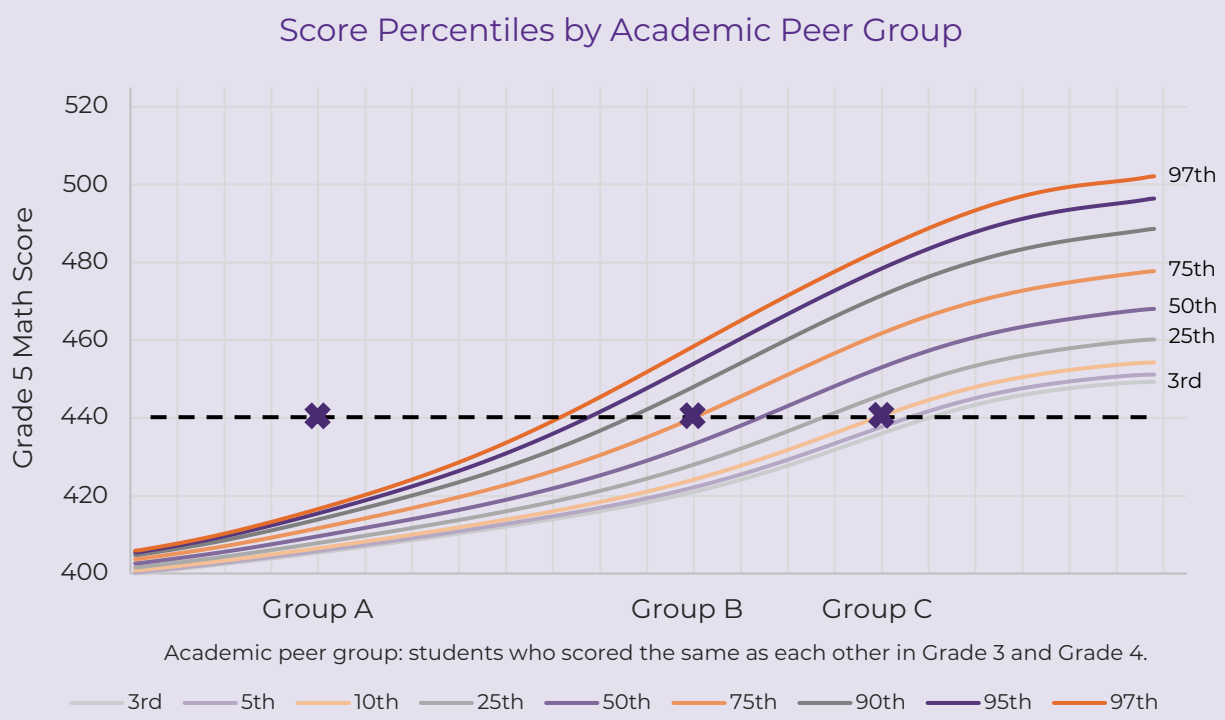
Student Growth Percentiles: Degree of Student Growth as Compared to Expected Growth, With Expected Growth Based on Students With Similar Academic Scores Over Time

Student growth percentiles (SGP) are regression-based models for student growth (Betebenner, 2009). The SGP model compares a student's current test score to the test scores the student is expected to earn, based on how the student's academic peer group has performed. A student's academic peer group consists of students who scored the same way on all relevant prior-year assessments (e.g., for Grade 6 students, scores on Grade 3, Grade 4 and Grade 5 assessments would be used to define the academic peer group). A student's SGP thus is reported in comparison to the student's peer group. For example, a student with an SGP of 40 grew academically more than 40% of students in the same academic peer group.

Figure 2-5 illustrates SGPs. In this figure, Grade 5 students' academic peers are based on their math scores in Grade 3 and Grade 4. The same Grade 5 math score will have a different SGP value, depending on the student's academic peers. For example, a student in Group A is unlikely to earn a score of 440; based on these data, less than 3% (and probably less than 1%) of students with the achievement history of students in Group A received a Grade 5 math score of 440. A student in Group B who received a Grade 5 score of 440 would receive an SGP of 75. A student in Group C who received a score of 440 would receive an SGP of 10, indicating that the student scored lower than 90% of students with the same academic history.

Many applications of SGP models add a growth-to-target component (see section below) to determine whether a student's expected scores are sufficient for the student to meet a specified target (e.g., the student will be Proficient in Grade 8). SGP models also can be extended to cohorts of students. A cohort of students can be compared to other cohorts with similar previous test scores, and the residual value of the predicted and observed test scores would reflect how well the cohort performed relative to these similar cohorts.

FIGURE 2-5: ILLUSTRATION OF STUDENT GROWTH PERCENTILES



Value-Added: Degree of Student Growth as Compared to Expected Growth, With Expected Growth Based on Multiple Predictive Variables

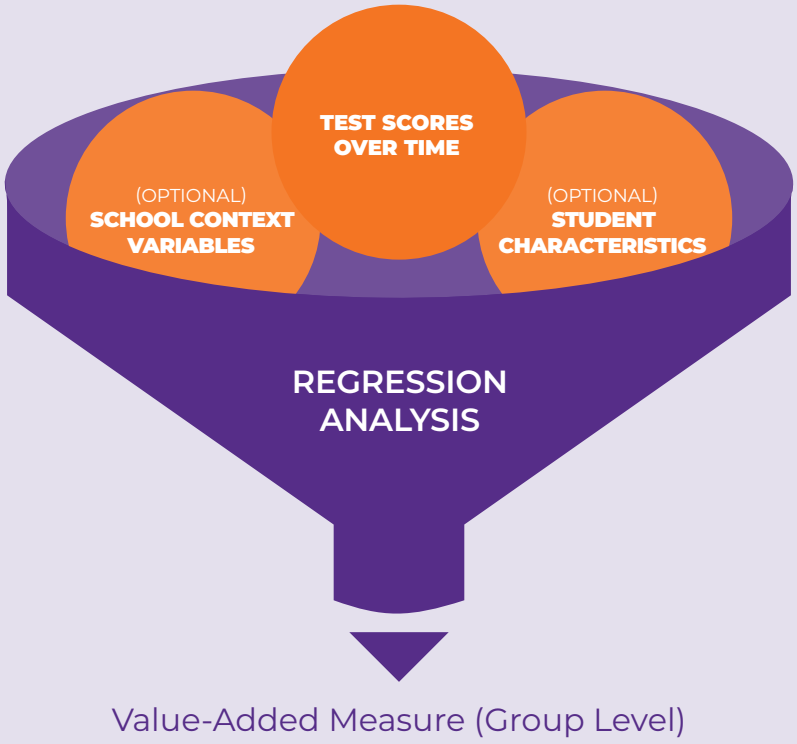
Value-added models are regression-based models of student growth that are designed for making inferences at a group level (e.g., classroom, school, district; Domaleski & Hall, 2016). Predictor variables are used to estimate the degree to which the observed change in score is different from the expected change. Predictor variables in the value-added model are selected from student characteristics or contexts which have been shown to be related to test scores but are not necessarily academic in nature, such as income. The argument for using non-academic predictor variables is to take into account variables outside the influence of the school or teacher; however, using demographic variables in the prediction model can be controversial, so the variables should be selected with care.

Value-added models must meet the assumptions underlying the use of regression models, including that test scores are reported on an interval scale. In addition, value-added models require large enough sample sizes to produce stable estimates.

Figure 2-6 illustrates value-added modeling. The nature of the inputs (i.e., test scores, context variables, and student characteristics) vary depending upon the philosophy underlying the use of the model, and the analyses also vary across models. For example, some states may include a measure of school community socio-economic status as an input. The output is not directly interpretable in terms of score gains, and individual student results are not reported. Often, the result is transformed into a more interpretable measure — for example, a score of 50 indicates that a school achieved expected growth.

A major characteristic of value-added models that distinguishes them from other growth models is that their primary purpose is to isolate and attribute changes in student test scores to a particular entity (e.g., school). More information about value-added models can be found in National Research Council and National Academy of Education (2010), Lockwood et al., (2007), Lockwood and McCaffrey (2007), or Ballou, Sanders, and Wright (2004).

FIGURE 2-6: VALUE-ADDED MODEL



Growth-to-Target: Degree of Student Growth Relative to Growth Needed to Reach a Specified Target in the Future

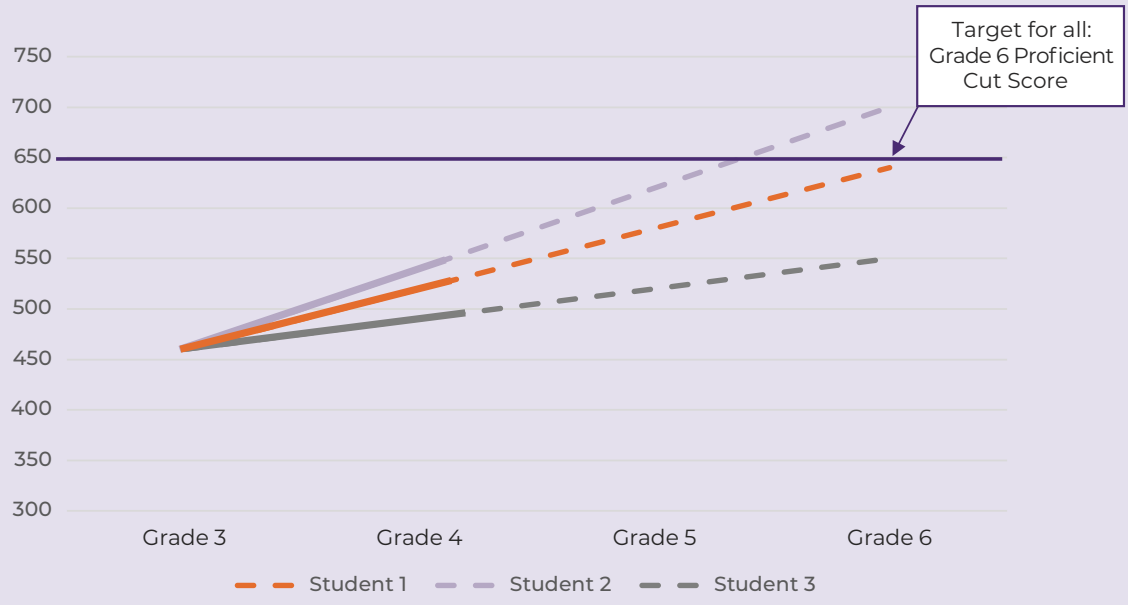
Growth-to-target models are used to evaluate a student's observed growth score against the score needed to reach a specified score target within a defined timeframe. For example, the target may be that all students will earn a score in the Proficient range by Grade 8. A student's growth score is evaluated to determine whether the student will meet the defined target if the student's score grows at the same rate every year. If the student's score trajectory is such that they will reach or exceed the target, the student has made sufficient growth. Otherwise, the student has not met the growth criterion.

These models are applied to growth scores that result from the application of another growth model such as those described earlier in this chapter. The underlying principles of a growth-to-target model (i.e., estimating whether a student is "on track" to reach a particular target in the future) is used in conjunction with other growth models by adding a target to the other growth model.

Figure 2-7 illustrates how a growth-to-target model would work using a simple gain score model. Suppose the target specified by the state is that students will be Proficient in math in Grade 6, and a score of 650-710 defines Grade 6 proficiency. For each possible Grade 3 score, the average number of points needed to reach the target (a score of Proficient [650] by Grade 6) is determined. For example, students scoring 460 in Grade 3 would need to gain an average of 64 points per year (score of **460** in Grade 3 **+** 3 years **x** score gain of 64 **=** score of **652** in Grade 6). Students scoring 500 in Grade 3 would need to gain an average of 50 points per year.

All three students illustrated in Figure 2-7 have the same score of 460 in Grade 3; Student 1 gains 60 points in Grade 4, Student 2 gains 80 points in Grade 4, and Student 3 gains 30 points in Grade 4. If Student 1 continues to gain 60 points each year, they will reach a score of 640 in Grade 6, which is below the Grade 6 Proficient cut score. Therefore, the student has not met the growth criterion in Grade 4. If Student 2 continues to gain 80 points each year, they will reach a score of 700 in Grade 6, which is above the Grade 6 Proficient cut score; Student 2 has met the growth criterion in Grade 4. If Student 3 continues to gain 30 points annually, they will score at 550 in Grade 6, below the Proficient cut score.

FIGURE 2-7: GROWTH-TO-TARGET MODEL BASED ON SIMPLE GAIN SCORES, LINEAR GROWTH



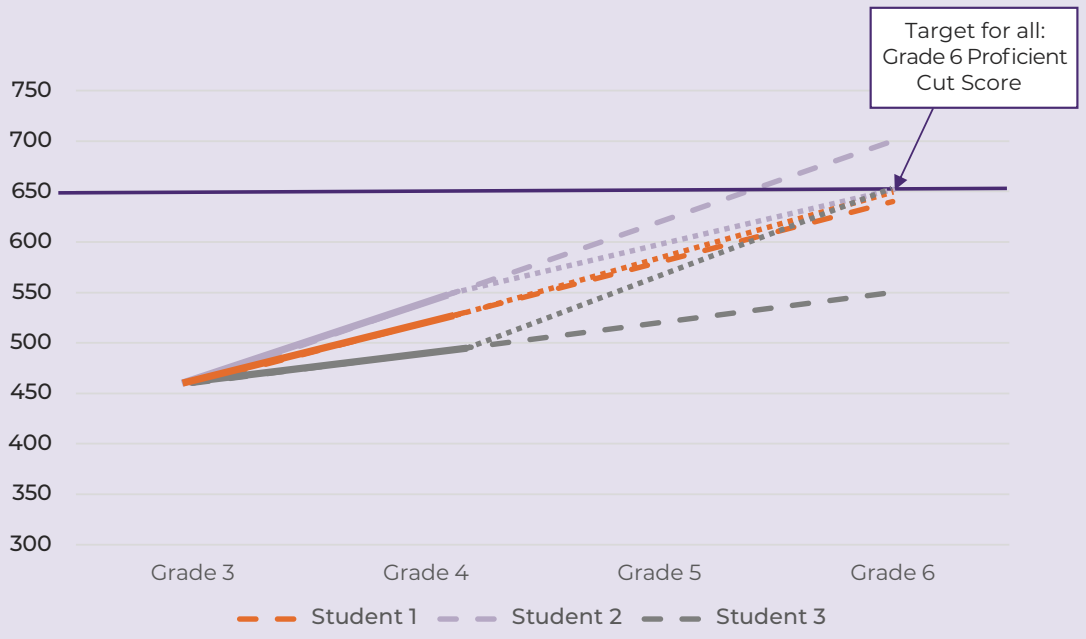
Solid line = Actual scores.
Dashed line = Projected scores, if there is no change in the student's growth rate after Grade 4.

Once a student's gain score has been determined in a given year, the rate at which that student will need to grow in the following years may need to be recalculated to keep them on track for achieving the long-term target. Figure 2-8 illustrates the rates at which the students from Figure-2-7 will need to grow over the next year in order to meet the growth target. Student 1 (with a score of 520 in Grade 4) will need to gain 65 points each year to meet the target; Student 2 (with a score of 540 in Grade 4) will need to gain 55 points each year; and Student 3 (with a score of 490 in Grade 4) will need to gain 80 points each year to score at the Proficient level in Grade 6.

Note that growth-to-target results are not *predictions* of student proficiency in the future based on past performance or other variables. Instead, the results simply indicate what would happen if the student learned at the same rate between the current year and the target year. In reality, a particular student may learn at different rates across years. For example, Student 3 might have an excellent year in math in Grade 5, gaining enough score points to meet growth-to-target. Student 1 might have an intervention in Grade 5 that puts them back on track to meet the Proficient standard in Grade 6. Conversely, Student 2 might struggle with Grade 5 math and not gain enough points to meet growth-to-target.

More sophisticated uses of the growth-to-target principle can be found when it is incorporated as part of other growth models, for example, student growth percentiles (Betebenner, 2009). In the application of growth-to-target illustrated above, the assumption is that students will gain the same number of points each year and that the vertical scaling results in approximately equal intervals between scores over the full scale. Other applications of growth-to-target may not require vertical scales or may be more complex — for example, the student's predicted growth trajectory may be based on research about rates of learning for different grade levels in a particular subject area.

FIGURE 2-8: GROWTH-TO-TARGET MODEL BASED ON SIMPLE GAIN SCORES, LINEAR GROWTH



Solid line = Actual scores.
Dashed line = Projected scores if there is no change in the student's growth rate after Grade 4.
Dotted line = Growth rate needed to meet target in Grade 6, based on Grade 4 score.

DATA REQUIREMENTS FOR GROWTH MODELS[2]

States should select a model based on the desired uses and interpretations of growth measures. However, a state's options for calculating growth measures are restricted based on the type of scores generated for the AA-AAAS. For example, a state could not apply a gain score model to an AA-AAAS that has separate scales for each grade, because the gain score model requires a vertical scale. Another factor that affects the choice of model is the sample size needed to calculate the growth measure.

One factor that can affect the interpretation of any growth measure for an alternate assessment is the stability of the population of the students taking the AA-AAAS. Currently, states are modifying their criteria for AA-AAAS participation to meet the Every Student Succeeds Act (ESSA) requirement that no more 1% of a school's student population take the alternate assessment. In addition, students with SCD may switch between taking the alternate and general assessment from year to year. One study estimated that 25% of students with SCD switch, and those who do are on the higher end of the performance continuum (Saven et al., 2016).

Finally, the technical and construct-related qualities of the AA-AAAS itself can affect whether a growth measure can be appropriately derived and, if so, which model(s) can be used. For example, when students start at different points on the score scale, the nature of the score distributions can affect the degree of growth. If a section of the scale near a cut point is sparsely populated, measuring growth between the lower proficiency level and the next proficiency level may be more difficult than it would be at other sections of the scale.

Categorical Gain Models

Categorical Gain Models require using achievement level scores. The achievement levels need to be vertically articulated across grades, so that each achievement level implies the same degree of KSUs relative to the expectations in the extended standards for that grade level. For example, the student who is Proficient in Grade 3 has attained the same level of KSUs (relative to the Grade 3 extended standards) as the student who is Proficient in Grade 4 has (relative to the Grade 4 extended standards). This type of articulation is often supported by cross-grade policy definitions in each content area and empirically supported during standard-setting.

The remaining growth models — **gain scores, residual gains, value-added, SGP** and **growth-to-target** — require a numeric score. Simple gain scores require that scores be on a vertical scale. The other types of models can use vertically scaled scores but do not require them. The scales used to report test scores must be meaningful, in terms of characterizing a student's level of attainment on the assessed constructs. An increase in scale score should reflect an increase in construct-related KSUs; the scores also should be reliable and support valid inferences.

Sample size and other statistical requirements for calculating different growth models vary by model. For the *categorical gains model*, there are no requirements for minimum sample size, and calculations do not depend on sophisticated statistical analyses. Similarly, for calculating *gain scores*, there are no requirements in terms of sample size.

Sample size may affect the reliability of *residual gain scores* and how the residual gain score can be used as part of a more complex model (for example, value-added models). To calculate residual gain scores, the assumptions of linear regression must be met, since regression is used to make a prediction of expected score based on the current test score (Castellano & Ho 2013).

As with residual gain score models, the data used in *value-added models* must meet the assumptions for linear regression. Because value-added models rely on regression models that control for a number of predictor variables, larger sample sizes are typically required. It is recommended that at least 1,000 students are used when calculating SGP (D. Betebenner, personal communication, January 16, 2021); the size of the sample needed depends on the number of prior achievement scores used and other factors (Culbertson, 2016).

[2]For more information about technical characteristics of growth models, including reliability, bias and measurement error, see Castellano & McCaffrey (2020); Monroe & Cai (2015); McCaffrey et al., (2015); Shang et al., (2015); Wyse & Seo (2014).

The following table (Figure 2-9) summarizes some salient characteristics of the growth models discussed.

FIGURE 2-9: SELECTED CHARACTERISTICS OF GROWTH MODELS*

	CATEGORICAL GAIN	SIMPLE GAIN SCORE	RESIDUAL GAIN SCORE	STUDENT GROWTH PERCENTILES	VALUE ADDED	GROWTH-TO-TARGET
CLAIM	Degree of student growth related to performance levels	Magnitude of student growth expressed in scale scores	Degree of student growth as compared to expected growth, with expected growth based on prior score	Degree of student growth as compared to expected growth, with expected growth based on students with similar academic scores over time	Degree of student growth as compared to expected growth, with expected growth based on multiple predictive variables	Degree of student growth relative to growth needed to reach a specified target in the future
DATA REQUIREMENTS	Vertically articulated cut scores across grades	Vertically scaled scores	Meaningful scores in terms of characterizing level of attainment of targeted KSUs	Meaningful scores in terms of characterizing level of attainment of targeted KSUs; large sample sizes	Sufficient longitudinal data to provide stable results; meaningful predictors	Meaningful scores in terms of characterizing level of attainment of targeted KSUs
EASE OF INTERPRETATION	Tied directly to performance levels, easily interpreted by families and educators	Intuitive, familiar measure that ties into how many people think about "growth"	Can be explained to users in familiar terms related to test scores	Can be explained to users in familiar terms using analogies	Difficult to explain, not expressed in relation to test scores	Can be explained to users in familiar terms, depending on model that growth-to-target is based on
PRIMARY CAUTION	Can only reflect coarse levels of growth	May be misinterpreted; vertical scales are not equal interval across grades	Assumptions of linear regression must be met; normative rather than criterion referenced	The model is normative rather than criterion referenced	Provides group-level interpretations; not related to learning theory – heavily data based	Not a stand-alone model; added to other models

*Sources: Castellano & Ho (2013); Domaleski & Hall (2016); Data Quality Campaign (2019)

WHAT ARE GROWTH MEASURES?

The application of a growth model to student test data results in **growth measures**. Measures of student growth are used to describe changes in academic performance over time, as demonstrated through test scores. Growth can be measured from year to year — for example, demonstrating a change in academic performance from Grade 3 to Grade 4 — or within year — for example, demonstrating a change in academic performance from the fall to spring semester during an academic year (Nehler et al., 2019). For individual-based growth measures, the unit of growth is the change between Score 1 and Score 2 or between a combination of scores and the current score; for cohort-based growth measures, the unit of growth is a summary (e.g., the mean) of the change.

"Change" is operationally defined differently by different growth models, as noted above. Moreover, the growth model applied to the general assessment may be different from the model applied to the alternate assessment. In this case, the operational definition of "change" will be differently for each assessment; however, a common principle underlying change, such as growth toward readiness for postsecondary opportunities, may underly both operational definitions.

Growth measures are used to develop academic growth indicators for state accountability systems. In some cases, these indicators are simply a summary of student growth scores, such as the mean or median student growth percentile. In other cases, the indicators are based on a judgment of the adequacy of the growth measure, such as the proportion of students meeting a growth criterion. Growth measures also may be used for other purposes — such as to inform curriculum changes at the school level; to be used as a State Identified Measurable Result (SIMR) in the State's Systemic Improvement Plans; or to track and substantiate the progress of individual students (see Chapter 6 for discussion of these uses). Each of these uses of the growth measure should be supported and validated, preferably by using an approach which is consistent with a theory of action (see Chapter 3).

The Effects of Technical Quality and Other Characteristics of the AA-AAAS on the Development of Growth Measures

A state's AA-AAAS must have sufficient evidence of technical quality as a precondition for calculating growth. As Koretz (1992) pointed out, the reliability and validity of the components in an indicator system affect the quality of the system, and these characteristics become more important as the importance of the systems in which they are used increases.

Fortunately, the technical quality of AA-AAAS systems has improved substantially since they were first required by the Elementary and Secondary Education Act (Karvonen et al., 2017; Sir, 2017). The stringency of peer review guidelines for state AA-AAAS systems is similar to that for general assessment systems (U.S. Department of Education, June 2018); in fact, many of the criteria for technical quality are identical, with differences based primarily on the context in which the assessments are administered and the flexibility needed to appropriately assess students with SCD.

For example, the types of evidence needed to satisfy Critical Element (CE) 3.1 (*Overall Validity, Including Validity Based on Content*) are the same for the general assessment and the AA-AAAS. These include test blueprints and the results of independent evaluations of alignment between the assessments and the targeted academic content. For CE 3.2 (*Validity Based on Cognitive Processes*), eligible evidence includes the results of cognitive labs and empirical evidence for the relationship of the assessment to other assessments requiring similar levels of cognitive complexity. States must demonstrate adequate reliability of scores for both the AA-AAAS and general assessment systems (U.S. Department of Education, June 2018).

Several of the critical elements have an obvious connection to prerequisites for growth calculations. For example, an AA-AAAS must be linked to the state's academic content standards at the intended depth and cognitive complexity for each grade (CE 2.1) and provide a precise estimate of student performance across the full performance continuum (CE 4.3). The AA-AAAS standards are expected to be challenging and aligned, so students are on track to pursue postsecondary opportunities (CE 6.3). Note that, if an AA-AAAS system has met peer review requirements for technical quality, it has met a necessary *but not sufficient* condition for supporting the development of growth measures.

A state's positive peer-review outcome does not guarantee that the state can use its AA-AAAS scores to produce growth measures. For example, peer review guidance does not require assessment scores to be vertically scaled, but a well-developed vertical scale is essential for calculating growth using simple gain scores. Normative growth measures with targets, such as SGPs, require explicitly indicating how much growth should be expected for students with SCD — who comprise a much more heterogeneous group than the other 99% of students who take statewide assessments; moreover, little is known about how students with SCD learn academics over time. Thus, the AA-AAAS scores may need to be further evaluated to ensure that results are equitable and that student achievement is the driver of student scores.

As of 2021, most states' AA-AAAS scoring models are based on scale scores or diagnostic classification models (DCM). It will be important for states to consider issues such as model fit, potential ceiling or floor effects and assumptions inherent within the chosen scale about variability in student achievement over time. Because DCM-based scoring is founded on patterns of mastery of separate attributes (skills) within a subject, results do not aggregate into interval-level data that could then be treated like a raw or scale score for calculating growth. These considerations will be further discussed in Chapter 4.

WHAT ARE GROWTH INDICATORS?

Growth indicators are summaries and transformations of growth measures, typically expressed at the school, district and state levels. Under ESSA, most states incorporate an academic growth indicator in their school accountability systems as their "other academic indicator" for elementary and middle schools.

As noted in Chapter 1, state accountability systems use their accountability indicators to differentiate schools and to monitor progress toward long-term goals. The state's academic growth indicator will typically be based on a transformation of its growth measure, which is based on a particular growth model. For example, State A uses individual student growth percentiles as its measure of English language arts (ELA) growth and defines the ELA growth indicator as follows:

> A school receives 1 point for each student whose SGP is between 1 and 25; 2 points for students with an SGP between 26 and 50; 4 points for students with an SPG between 51 and 75; and 5 points for an SGP higher than 75. The total number of points earned by the school is divided by the total number of students reported. The resulting number is the school's ELA academic growth indicator score.[3]

Growth indicators may be categorical or continuous. Depending on the intended uses of the growth indicator, growth measures can be aggregated into growth indicators based on individual students or based on cohorts of students. In the example above, the growth indicator was aggregated using individual student growth measures.

If the same growth model cannot be used for the general assessment and AA-AAAS — for example, if the general assessment uses the SGP model and the AA-AAAS uses categorical gains — a growth indicator can be designed to incorporate growth measures resulting from both models. (This topic is further addressed in Chapter 5 and Chapter 6.) Whether or not the same growth measures are used for both types of assessment, when a growth indicator is developed that incorporates both general assessment and alternate assessment growth measures, an equitable growth indicator requires that each student's growth measure is given equal weight, regardless of the test taken.

The intended and unintended consequences of using a particular growth model, measure or indicator need to be carefully considered, as discussed in the subsequent chapters. It is critical that the growth measures and indicators resulting from applications of the growth model provide an unbiased picture of changes in student achievement.

Which models are states using for their AA-AAAS growth measures, and how are they incorporating them into their growth indicators?

CCSSO recently surveyed the states that are members of the Assessment, Standards & Education for Students with Disabilities Collaborative (ASES), and Accountability Systems and Reporting Collaborative (ASR) and received 25 responses.

- Seven (7) states reported that their accountability systems included growth on the alternate assessment in their accountability systems.
  - Four (4) of these seven states reported using a categorical gain model with their AA-AAAS scores;
  - Two (2) states use a growth-to-target model based on student growth percentiles; and
  - One (1) state uses percentile rank residuals to model growth;
- Three (3) states used a different model for their general and AA-AAAS growth measures, but all seven (7) states converted those measures so they could be combined into a single academic growth indicator.

Appendix C summarizes the responses from all states that responded to the survey.

[3]Most growth indicator calculations are more complex than this illustration.

SUMMARY

This chapter described and differentiated growth measures, growth models and growth indicators and discussed the relationships among the three concepts. While a state's underlying philosophy, theory of action and proposed uses of a growth measure will be key to the development of the measure, understanding the limitations and requirements of growth models will help the state to set parameters around the type of measure it can use and how to incorporate the AA-AAAS measure into its indicator system.

Chapter 3 addresses how the components of a theory of action relate to defining and developing growth measures and indicators based on alternate assessments. Chapter 4 and Chapter 5 will use the foundational information covered in the first three chapters to discuss applications in the development of growth measures and indicators. Chapter 6 outlines a process for considering whether and how to use alternate assessment growth in state accountability systems.

REFERENCES

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for Student Background in Value-Added Assessment of Teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37-65. https://eric.ed.gov/?q=Ballou%2c+Sanders&id=EJ727499

Betebenner, D. (2009). Norm- and Criterion-Referenced Student Growth. *Educational Measurement: Issues and Practice*, *28*(4), 42–51. https://onlinelibrary.wiley.com/doi/10.1111/j.1745-3992.2009.00161.x

Blackorby, J., Taylor, C., & Wei, X. (2016). *Using Growth Models to Measure Child/Student Outcomes for State Systemic Improvement Plans: A Guide for States*. IDEA Data Center. https://files.eric.ed.gov/fulltext/ED584165.pdf

Castellano, K. E. & Ho, A. D. (2013). Contrasting OLS and Quantile Regression Approaches to Student "Growth" Percentiles. *Journal of Educational and Behavioral Statistics, 38*(2), 190-215. https://doi.org/10.3102/1076998611435413

Castellano, K. E. & McCaffrey, D. F. (2020). Comparing the Accuracy of Student Growth Measures. *Journal of Educational Measurement*, *57*(1), 71-91. https://doi.org/10.1111/jedm.12242

Culbertson, M. J. (2016). *Precision of student growth percentiles with small sample sizes*. Regional Educational Laboratory Central. https://files.eric.ed.gov/fulltext/ED568943.pdf

Data Quality Campaign. (2019). *Growth Data: It Matters, and It's Complicated*. https://dataqualitycampaign.org/wp-content/uploads/2019/04/DQC-Growth-Data-Resources.pdf

D'Brot, J. (2017). *Considerations for Including Growth in ESSA State Accountability Systems*. Council of Chief State School Officers. https://ccsso.org/sites/default/files/2017-10/CCSSOGrowthInESSAAccountabilitySystems1242017.pdf

Domaleski, C. & Hall, E. (2016). *Guidance for Estimating and Evaluating Academic Growth*. National Center and State Collaborative. http://www.ncscpartners.org/Media/Default/PDFs/Resources/EstimatingandEvaluatingStudentGrowth.pdf

Karvonen, M., Wakeman, S., & Kingston, N. (2017). Alternate Assessment. In M. L. Wehmeyer & K. A. Shogren (Eds.). *Handbook of Research-Based Practices for Educating Students with Intellectual Disability* (pp. 102–118). Routledge. https://www.routledge.com/Handbook-of-Research-Based-Practices-for-Educating-Students-with-Intellectual/Wehmeyer-Shogren/p/book/9781138832107

Koretz, D. (1992). *Evaluating and Validating Indicators of Mathematics and Science Education*. Rand. https://files.eric.ed.gov/fulltext/ED353127.pdf

Lockwood, J. R., & McCaffrey, D. F. (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics, 1,* 223-252. https://doi.org/10.1214/07-EJS057

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Le, V., & Martinez, F. (2007). The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures. *Journal of Educational Measurement, 44*(1), 47-67. https://doi.org/10.1111/j.1745-3984.2007.00026.x

Madison, M. J. & Bradshaw, L. P. (2018a). Assessing Growth in a Diagnostic Classification Model Framework. *Psychometrika, 83*(4), 963-990. https://doi.org/10.1007/s11336-018-9638-5

Madison, M. J. & Bradshaw, L. (2018b). Evaluating Intervention Effects in a Diagnostic Classification Model Framework. *Journal of Educational Measurement, 55*(1), 32-51. https://doi.org/10.1111/jedm.12162

McCaffrey, D. F., Castellano, K. E., & Lockwood, J. R. (2015). The Impact of Measurement Error on the Accuracy of Individual and Aggregate SGP. *Educational Measurement: Issues and Practice, 34*(1), 15-21. https://eric.ed.gov/?id=EJ1054132

Monroe, S. & Cai, L. (2015). Examining the Reliability of Student Growth Percentiles Using Multidimensional IRT. *Educational Measurement: Issues and Practice*, *34*(4), 21–30. https://doi.org/10.1111/emip.12092

National Research Council & National Academy of Education. (2010). *Getting Value Out of Value-Added: Report of a Workshop*. http://rlhoover.people.ysu.edu/ClassConnections/core-readings/Getting%20Value%20Out%20of%20Value-Added.pdf

Nehler, C., Clark, A., & Karvonen, M. (2019). *White Paper: Considerations for Measuring Academic Growth on Dynamic Learning Maps® (DLM®) Alternate Assessments*. University of Kansas, Center for Accessible Teaching, Learning, and Assessment Systems. https://dynamiclearningmaps.org/sites/default/files/documents/publication/Growth_White_Paper.pdf

Saven, J. L., Anderson, D., Nese, J. F. T., Farley, D., & Tindal, G. (2016). Patterns of Statewide Test Participation for Students with SCD. *The Journal of Special Education*, *49*(4), 209–220. https://doi.org/10.1177/0022466915582213

Shang, Y., VanIwaarden, A., & Betebenner, D. W. (2015). Covariate Measurement Error Correction for Student Growth Percentiles Using the SIMEX Method. *Educational Measurement: Issues and Practice, 34*(1), 4-14. https://doi.org/10.1111/emip.12058

Sir, D. N. (2017). *Comparing Two Alternate Assessments: Dynamic Learning Maps and Multi-State Alternate Assessment.* (Publication No. 10685721) [Doctoral dissertation, Seton Hall University]. ProQuest Dissertations and Theses Global. https://scholarship.shu.edu/cgi/viewcontent.cgi?article=3481&context=dissertations

U.S. Department of Education. (2018). *A State's Guide to the U.S. Department of Education's Assessment Peer Review Process*. https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf

CHAPTER

3

USING A THEORY OF ACTION

INTRODUCTION

Chapter 1 provided background information on using alternate academic achievement standards (AA-AAAS) in growth models. It described the students with significant cognitive disabilities (SCD) who take the AA-AAAS and their education, and it reviewed both the educational rationales and legal requirements for including these students in assessment and accountability. As noted in Chapter 1, there is very little literature about the rates of academic learning for students with SCD; there is a need for research into these students' academic growth and the conditions that facilitate their learning. Chapter 2 described fundamental concepts related to growth measures, growth models and growth indicators.

Chapter 3 will describe how a theory of action can support the development of growth measures and indicators. This chapter is intended to provide an overview of how using a theory of action can assist developers in documenting and establishing the validity of inferences drawn from, and the appropriate use of, any policy-related measure, not just growth measures based on AA-AAAS. Although central concepts are illustrated by using examples from alternate assessments and growth measures, this chapter does not build a theory of action for an alternate assessment growth measure or for incorporating one into a state accountability system's growth indicator.

This chapter will support the reader in understanding how the principles behind a theory of action can be used to determine whether and how to include AA-AAAS in a state academic growth indicator. It also will highlight how the theories of action for the assessment system and for the accountability system are intricately linked. Chapter 4 and Chapter 5 will focus on applications of the concepts discussed in this and previous chapters.

HOW IS VALIDITY RELATED TO THEORIES OF ACTION?

The validity of inferences drawn from a measure and the validity of uses made of the measure should be established and evaluated during the measure's design, development and implementation. This principle applies to accountability systems, which are used to make decisions about schools based both on individual measures and on combinations of measures (see American Educational Research Association, American Psychological Association, & National Council Measurement in Education, 2014, especially Chapter 13; Yarbrough, et al., 2010). Three critical steps in determining how to develop a growth measure based on AA-AAAS scores and introduce it into the state's accountability system are to evaluate:

- how the growth measure would relate to the form and purposes of the state's AA-AAAS system;
- how the growth measure would fit in with the purposes and uses of the state's accountability system; and
- the degree to which the use of the growth measure provides equitable and fair results.

To accomplish this, it is helpful for states to use processes stemming from theories of action.[4] A theory of action allows policymakers and developers to use a logical chain of reasoning to effectively construct and evaluate validity arguments (Haertel, 1999) about what a policy is intended to accomplish. Theories of action are useful tools for both designing and subsequently evaluating the validity of inferences made and actions taken that are based on educational assessment and accountability systems.

In general, theories of action specify the goals for policy and policy initiatives and the mechanisms for reaching those goals. In addition, they provide information on how the different parts of a system are intended to work together to produce the desired goals (Marion, Lyons, and D'Brot, 2016). Using a theory-of-action approach to develop and evaluate growth measures facilitates discussions and decisions around expectations, uses and assumptions underlying the measure.

Marion et al. (2016, pp. 3-4) specify the components states need to consider when initially developing a theory of action for an accountability system:

- the goals of the system;
- the purposes and intended uses of system results;
- the intended outcomes of the system;
- mediating outcomes needed to achieve the ultimate outcomes; and
- the underlying assumptions that must be met for the system to function.

While not all states have explicit or finely-developed theories of action for their assessment and accountability policies, some or all of the components of a theory of action underlie all state assessment and accountability systems.

For example, all states are required to have explicit statements on the purposes of their assessment systems as well as explicit statements on the intended uses and interpretations of assessment results (U.S. Department of Education, 2018). To support the valid uses and interpretations of test results, states design assessments in particular ways, connect the assessments to learning standards, and evaluate the degree to which their design and development efforts are successful; success is evaluated by documenting the processes used in test design and development, analyzing psychometric characteristics of the assessments and conducting validity studies.

[4] It is beyond the scope of this document to go into depth about theories of action or validity theory. For more information about developing theories of action and incorporating them into a validity argument, see the references cited in this chapter, in particular D'Brot, LeFloch, English, and Jacques (2020) for information specific to accountability systems.

Similarly, in their state plans submitted to ED states are expected to describe the long-term goals for their schools' academic achievement, graduation rate and English language proficiency levels. States also must specify how their accountability system indicators are related to those long-term goals. Since one of the required uses of a state's accountability system is to identify which schools need various levels of support, each state must lay out its methodology and rationale for the procedure it uses to meaningfully differentiate schools (U. S. Department of Education, 2017).

Figure 3-1 shows examples of how the components of a theory of action are already included in some aspects of state assessment and accountability systems. In developing systems that meet federal peer review requirements, state departments of education must consider these key components — which lays the foundation for documenting and evaluating validity.

FIGURE 3-1: EXAMPLES OF CORRESPONDENCE BETWEEN REQUIRED STATE SYSTEMS COMPONENTS & A THEORY-OF-ACTION (TOA) **FRAMEWORK**

TOA COMPONENT	ASSESSMENT OR ACCOUNTABILITY SYSTEM COMPONENT
Goal(s)	Often, the broad goals for a state *accountability* system are specified in state legislation or regulation. For example, the Code of Virginia (8VAC20-131-370) states that the goals of the state accountability and accreditation system are: 1. Building on strengths in schools and addressing specific areas needing improvement; 2. Driving continuous improvement in school achievement for all schools; 3. Identifying areas for technical assistance and the use of school improvement resources; and 4. Providing a comprehensive picture of school quality information to the public.
Purposes & Uses	Element 2.1 of the U.S. Department of Education's *assessment* peer review guidance (2018) requires that: "The State's test design and test development process … includes: [s]tatement(s) of the purposes of the assessments and the intended interpretations and uses of results (p. 36)." One of the required uses of the state *accountability* system is to identify schools needing various levels of support (for example, see the criteria under A.4.vi: Identification of Schools (U. S. Department of Education, 2017).
Intended Outcomes	Criteria A.4.iii.a.1, A.4.iii.b1 and 2, and A.4.iii.c.1 require states to identify and describe the long-term outcomes (called "goals" in the guidance) for academic achievement, graduation rate, and English language proficiency in their state *accountability* plans (U. S. Department of Education, 2017).
Mediating Outcomes	According to Section A.4.iv of the state plan peer preview guidance (U. S. Department of Education, 2017), states must specify how their *accountability* system indicators are related to their long-term outcomes (called "goals" in the guidance). In doing this, states are attending to the conditions that are necessary to achieve their intended outcomes.
Underlying Assumptions	For Critical Element 2.2, the *assessment* peer review guidance (U. S. Department of Education, 2018) suggests that evidence could include "a description of the process the State uses to ensure that the item types (e.g., multiple choice, constructed response, performance tasks, and technology-enhanced items) are tailored for assessing (a) the **academic content standards**; or (b) the **English language proficiency** (ELP) **standards** (p.39, emphasis in the original)." The relationship between item types and the content targeted for assessment forms the basis of an underlying assumption in the theory of action; the assumption is the assessment format(s) adequately address the constructs underlying the content standards.

Attending to components of a theory of action can provide or uncover evidence that supports the validity of uses of an accountability system and expose areas that need further consideration or study, even if a formal theory of action is not used. Theories of action support an argument-based approach to validity (Kane, 2006) and can provide a logical, natural language-based flow to considering validity issues (for example, see Clark & Karvonen 2020). However, the right knowledge and expertise are needed to develop and evaluate a theory of action. For growth measures derived from AA-AAAS scores to be used in the state accountability system, the development team should include experts in (a) educating students with SCD, (b) constructing accessible large-scale measures of achievement, and (c) designing, and developing educational accountability systems.

HOW CAN A THEORY OF ACTION BE USEFUL IN DEVELOPING AA-AAAS GROWTH MEASURES?

In this sub-section, each growth measure is discussed in the context of its use in a school accountability indicator system. If the measure is developed to have additional uses, the state will need to address considerations related to these additional uses, such as the utility and fairness of growth measures reported at the individual student level. The following sub-section begins with describing components of a theory of action for the growth measure and ends with an example of a growth measure based on these components.

As discussed earlier, there are several ways to define growth using any test scores. When defining the growth measure for an AA-AAAS (as with a general assessment), states need to consider both the nature of the test scores and the characteristics of the tested population. Similarly, there are multiple ways to structure a theory of action (see, for example, Bennett, 2010; Karvonen & Clark, 2020).

The general framework found in Marion, Lyons, Pace and Williams (2016) will be used here to illustrate how a theory-of-action approach can help states to define an AA-AAAS growth measure. Note that developing a growth measure based on AA-AAAS will have many components in common with developing the AA-AAS itself. For illustrative purposes, we use examples of components that overlap with those that would relate to a theory of action for an AA-AAAS.

Goals

Describing the ultimate goals for having growth measures based on the AA-AAAS is the first step in developing a growth measure. While the ultimate goal may be to ensure that academic growth of the students with SCD "counts" in the state accountability system, there are likely other goals for developing a growth measure. Some examples of these goals might be to:

- increase the achievement level of students with SCD;
- ensure that students with SCD are better prepared for postsecondary education;
- encourage attention to achievement growth for students with SCD;
- promote and make visible equity in educating students with SCD; or
- support students with SCD in making progress in the general curriculum.

Purposes and Intended Uses of Results

Clearly and explicitly describing the purpose(s) and use(s) of the growth measure will lay the foundation for supporting valid inferences based on that measure. For example, while the driving force for developing the measure may be to include all students (not just those taking the general assessment) in the accountability system's academic growth indicator, there may be other purposes for developing the measure.

There may be a single intended use of the growth measure — for example, including the measure in the overall academic growth indicator for each school — or there may be more than one intended use example (e.g., if the AA-AAAS growth measure is also provided as a stand-alone piece of information at the school and/or student level).

For example, the state's federal State Systemic Improvement Plan (Schiller, et al., 2020) requires the determination of a state-identified, measurable result for the purpose of improving the education of students with disabilities; a state may wish to use the AA-AAAS growth measure for this purpose. States also may wish to include the growth measure on individual student score reports for the alternate assessment.

In cases of purposes and uses of the growth measure that go beyond serving as part of an accountability indicator, it is important for states to develop a theory of action around the components that are specific to those uses. It is also valuable for states to consider these additional uses when developing the growth measure for inclusion of that measure in the state accountability system.

Intended Outcomes

The intended outcomes for a growth measure using AA-AAAS scores will be based on the state's goals for using the measure, but they will be more specific and measurable — that is, the state department of education will be able to evaluate the degree to which the intended outcomes have been met. For example, the intended outcomes may be to increase:

- the graduation rate for students with SCD;
- resources devoted to educating students with SCD;
- attention to within-year growth for students with SCD;
- the number of students with SCD who participate in postsecondary employment or schooling;
- and/or enhance the validity of inferences drawn from the state's accountability growth indicator.

As developers consider intended outcomes of the growth measure, it is critical to think about potential unintended outcomes and inappropriate uses of the measure and to warn users of these unintended outcomes. For example, the aggregated growth measure may be a reliable measure of students' growth at the school level when incorporated into a growth indicator, but it may not be reliable enough to make high-stakes decisions at the teacher level. If this is the case, and if there is likely to be a desire at the LEA level to use the AA-AAAS growth measure for teacher-evaluation purposes, the state department of education should notify districts that the measure has not been developed or evaluated for this purpose.

Mediating Outcomes

The mediating outcomes delineate conditions and actions that are needed for the state to meet the intended outcomes and overall goals. As noted above, for the AA-AAAS growth measure, many of these mediating outcomes will overlap with mediating outcomes related to the AA-AAAS itself. For example, for the intended outcome: "*Increase the graduation rate of students with SCD*," a mediating outcome may be: "*Teachers maintain student interest and engagement in learning*," or perhaps: "*Students receive transition services that align with their coursework and interests*." For the intended outcome: "*Increase the number of students who are participating in postsecondary employment or schooling*," one mediating outcome may be: "*Schools offer classes covering the knowledge, skills and understandings (KSU) necessary for postsecondary employment*."

Underlying Assumptions

Marion, Lyons, Pace and Williams (2016) recommend creating a relatively high-level theory of action before "zooming in" on the details of key sections. For instance, the examples cited above are relatively high-level; therefore, prior to defining the underlying assumptions, a state would create a logic chain specifying how the intended and mediating outcomes are expected to occur and tying those outcomes to the uses and goals. Once the general "how" is determined, the state could then define the underlying assumptions — i.e., the conditions that must be in place and actions that must be taken for the AA-AAAS growth measure to work as the state intended.

For example, a goal of both a growth measure based on AA-AAAS scores and the AA-AAAS system itself might be to prepare students with SCD for competitive, integrated postsecondary employment. An intended outcome could be to increase the proportion of students with SCD who pursue competitive integrated employment after high school graduation. In support of this intended outcome, some mediating outcomes[5] might be that:

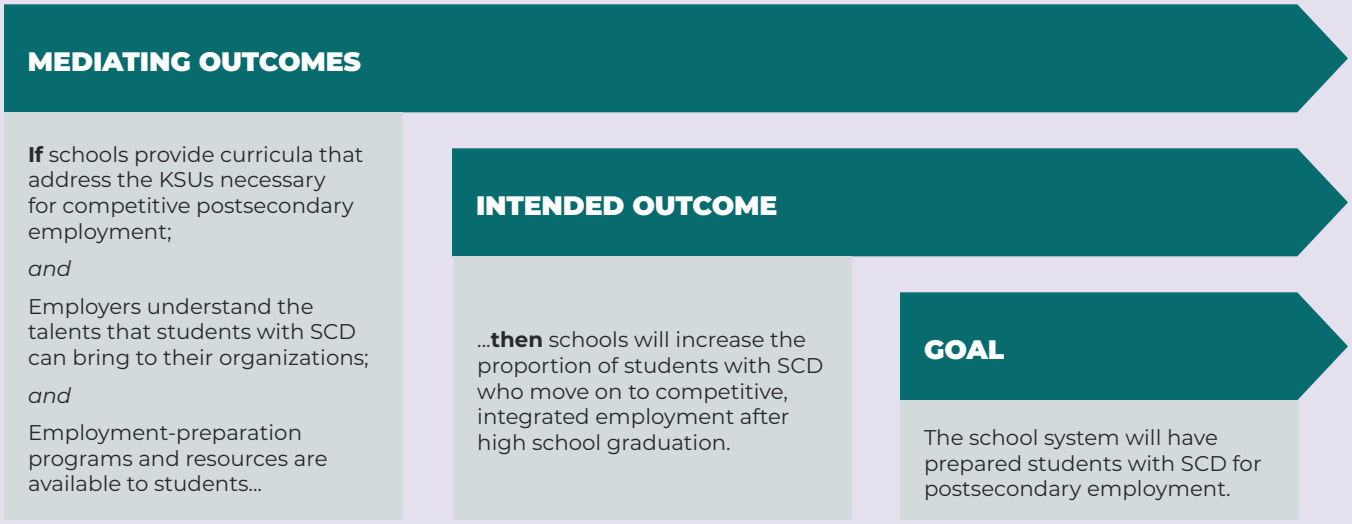
- schools provide curricula that address the KSUs students with SCD need for competitive postsecondary employment;
- employers understand the talents that students with SCD can bring to their organizations; and
- employment-preparation programs and resources are available to students with SCD.

[5] This is not intended to be a complete list of mediating outcomes.

However, the use of the growth measure will differ from the purposes and uses specified for the AA-AAAS score. For example, the growth measure may be intended to be used in the overall academic growth indicator for accountability purposes and to help teachers gauge student progress over time. Figure 3-2 shows the components of a high-level theory of action for this scenario.

The high-level theory of action also can illuminate and help the state guard against unintended outcomes. For example, in this scenario, an unintended outcome might be that IEP teams use a student's lack of academic growth as a reason to place the student in a segregated occupational setting, design a transition plan that presumes no postsecondary employment, or exclude the student from certain transition services. Thus, it would be important for the state to put in place procedures and safeguards to minimize the possibility of such outcomes.

FIGURE 3-2: SAMPLE HIGH-LEVEL LOGIC CHAIN



Next, the underlying assumptions — i.e., definitions of conditions that must be in place for the intended and mediating outcomes to hold — can be identified. For example, for the mediating outcome: "Schools provide curricula that address the KSUs necessary for competitive postsecondary employment," some underlying assumptions might be:

- The KSUs necessary for postsecondary employment can be identified.
- The state content standards address the KSUs needed for postsecondary employment.
- The state content standards contribute to the KSUs needed for postsecondary employment by building on skills across grade levels until the necessary KSUs are achieved; that is, they are vertically aligned.
- The state's AA-AAAS system measures the KSUs necessary for postsecondary employment at each grade level.
- The state's AA-AAAS growth measure reflects the vertical alignment and across-grade increase in KSUs needed for competitive postsecondary employment.
- Stakeholders accurately interpret the AA-AAS growth measure for individual students and groups of students.

Note that the sample of assumptions varies, in keeping with the need to address characteristics of the state's academic content standards, the assessment systems, and the growth measure. Even in this small sample, it is apparent that assumptions are not independent — there are obvious relationships between some assumptions.

Most of the underlying assumptions cited above are related to the AA-AAAS itself and are conditions underlying test design and development. The following are examples of a few assumptions which are specific to the growth measure, followed by propositions that can be evaluated by using evidence:

Assumption:
The state's AA-AAAS growth measure reflects the vertical alignment and across-grade increase in KSUs needed for competitive postsecondary employment.

Sample Proposition:
The content standards measured by the AA-AAAS are vertically aligned across grades. (This relates to test design and development.)

Assumption:
The growth measure can be appropriately aggregated into the accountability academic growth index.

Sample Proposition:
The meaning of the AA-AAAS growth indicator is generally equivalent to the meaning of the general assessment growth indicator, so that the two indicators can be meaningfully combined.

Sample Proposition:
Students taking the general assessment and students taking the AA-AAAS are weighted (or counted) equally within the indicator.

Assumption:
At the aggregate level, the growth measure reflects an increase or decrease in the attainment of targeted KSUs from one grade to the next.

Sample Proposition:
The scores that are the basis of the growth measure can be meaningfully compared from year to year. For example, a group of students with SCD that scores Proficient in Grade 3 and Proficient in Grade 4 has met grade-level academic expectations in both grades. A group of students with SCD that scores Proficient in Grade 3 and Partially Proficient in Grade 4 has not achieved the standards as well in Grade 4 as that group did in Grade 3.

Assumption:
At the student level, the growth measure reflects an increase or decrease in the attainment of targeted KSUs from one grade to the next.

Sample Proposition:
The scores a student earns in Year 1 and Year 2 reflect a difference in the complexity or amount of KSUs the student has attained in Year 1 and Year 2.

Sample Proposition:
The changes in student KSUs reflected by the growth measure are unbiased and fair.

Examples of how assumptions and their associated propositions can be evaluated will be addressed in Chapter 4 and Chapter 5 of this handbook.

SUMMARY

This chapter described the components of a theory of action and discussed how a theory of action can be used to help develop, document and evaluate a growth measure based on AA-AAAS scores. A theory of action can bring critical questions to the surface during development, which can help states attend to issues that might otherwise be missed. They also can use these questions to evaluate the validity of the growth measure once it has been developed.

More information on using theories of action and examples of well-developed theories of action can be found in the literature, including Schiller, E., Hayes, S., and Nagle, K. (2020), Bennett, Kane, & Bridgeman (February 2011), Gholson & Guzman-Orth (2019), and Marion, Lyons, & Williams (2016).

REFERENCES

American Educational Research Association, American Psychological Association, & National Council Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302356.pdf

Bennett, R. E. (2010). Cognitively Based Assessment of, for, and as Learning (CBAL): A Preliminary Theory of Action for Summative and Formative Assessment. *Measurement*, *8*, 70-91. https://doi.org/10.1080/15366367.2010.508686

Bennett, R. E., Kane, M., & Bridgeman, B. (2011, Feb. 10-11). Theory of Action and Validity Argument in the Context of Through-Course Summative Assessment [Paper presentation]. *Invitational Research Symposium on Through-Course Summative Assessments*, Atlanta, GA. https://www.ets.org/Media/Research/pdf/TCSA_Symposium_Final_Summary.pdf

Clark, A. K., & Karvonen, M. (2020). Constructing and Evaluating a Validation Argument for a Next-Generation Alternate Assessment. *Educational Assessment*, *25*(1), 47-64. https://doi.org/10.1080/10627197.2019.1702463

D'Brot, J., LeFloch, K., English, D., Jacques, C. (2020). *Evaluating State Accountability Systems*. American Institutes for Research. https://oese.ed.gov/resources/oese-technical-assistance-centers/state-support-network/resources/evaluating-state-accountability-systems-esea/

Gholson, M. L., & Guzman-Orth, D. (2019). *Developing an Alternate English Language Proficiency Assessment System: A Theory of Action* (Research Report No. RR-19-25). ETS. https://doi.org/10.1002/ets2.12262

Haertel, E. H. (1999), Validity Arguments for High-Stakes Testing: In Search of the Evidence. *Educational Measurement: Issues and Practice*, 18: 5-9. https://doi.org/10.1111/j.1745-3992.1999.tb00276.x

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17–64). National Council on Measurement in Education and the American Council on Education. https://www.worldcat.org/title/educational-measurement/oclc/69104278

Marion, S., Lyons, S., and D'Brot, J. (2016). *Developing a Theory of Action to Support High-Quality Accountability System Design*. Center for Assessment. https://www.nciea.org/sites/default/files/publications/Theories%20of%20Action%20for%20ESSA_040506_2.pdf

Marion, S. F., Lyons, S., Pace, L., & Williams, M. (2016). A Theory of Action to Guide the Design and Evaluation of State Innovative Assessment and Accountability System Pilots. www.innovativeassessments.org. https://www.nciea.org/sites/default/files/publications/Marion_IADA%20Theories%20of%20Action.pdf

Schiller, E., Hayes, S., & Nagle, K. (2020). *Using a Theory of Action to Develop Performance Indicators to Measure Progress Toward a SiMR*. IDEA Data Center. https://ideadata.org/sites/default/files/media/documents/2020-02/1-IDC-SSIP-WhitePaper_1.pdf

U.S. Department of Education. (2018). *A State's Guide to the U.S. Department of Education's Assessment Peer Review Process*. https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf

U.S. Department of Education. (2017). *State Plan Peer Review Criteria: The Elementary and Secondary Education Act of 1965, as Amended by the Every Student Succeeds Act*. https://www2.ed.gov/admins/lead/account/stateplan17/essastateplanpeerreviewcriteria.pdf

Yarbrough, D. B., Shula, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users* (3rd ed.). SAGE Publications, Inc. https://us.sagepub.com/en-us/nam/the-program-evaluation-standards/book230597

CHAPTER

4

APPLICATIONS OF GROWTH MEASURES BASED ON AA-AAAS SCORES

APPLICATIONS OF GROWTH MEASURES BASED ON AA-AAAS SCORES

INTRODUCTION

Chapter 1 reviewed background information relevant to using alternate academic achievement standards (AA-AAAS) in growth models. The students who take the AA-AAAS and their education were described and reviewed both educational rationales and the legal requirements for including these students in assessment and accountability. Chapter 2 sought to orient readers to growth models, growth measures and growth indicators from a conceptual perspective. Chapter 3 described how a theory of action can support the development of growth measures and indicators. It also provided an overview of how using a theory of action can assist developers in documenting and establishing the validity of inferences made from and appropriate use of any policy-related measure, not just growth measures based on AA-AAAS.

Chapter 4 introduces the reader to how growth models can be applied to produce growth measures for contemporary AA-AAAS, including:

- key questions to ask when considering the policy and conceptual definitions of growth on extended standards and the alternate assessment;
- considerations for the AA-AAAS design, administration and scoring; and
- examining the technical characteristics of growth models when applied to the AA-AAAS.

This section lays the groundwork for incorporating growth on the AA-AAAS as an indicator, which is covered in in Chapter 5. Additionally, Chapter 4 builds on the prior chapters, as a strong theory of action is necessary in order to evaluate whether (a) a growth model can produce growth measures appropriately and (b) aggregations of the measure lead to sound interpretations of growth indicators (the latter of which is addressed in Chapter 5). Chapter 4, also raises several concepts, including intended policy goals, general growth inferences and technical considerations for growth.

WHAT ARE GROWTH MODELS, MEASURES AND INDICATORS?

For the purposes of this chapter, it is helpful to revisit the relationship among growth models, measures and indicators, which were defined in Chapter 2 (see page 22). As a reminder, a growth **model** is the procedure by which scores are calculated to arrive at a growth **measure**. Growth measures can be computed at the student level and may be presented alongside the student's Individual Score Report (ISR) to assist in interpretation. These measures then usually are aggregated in the educational accountability system as a growth **indicator** to reflect the degree to which students in a school, district or state have progressed academically.

If a state uses multiple growth models (for example, one for the general assessment and one for the alternate assessment), the indicator may be based on a combination of summary growth measures. The indicator may take different forms, depending on the nature of the accountability system; for example, some states use index systems and transform the growth summary into an index score.

WHAT MUST WE CONSIDER BEFORE APPLYING GROWTH MODELS TO AA-AAAS SCORES?

While a majority of states have identified a growth model that is applied to the general assessment, most have not applied any growth model to their alternative assessments. Keep in mind that identification alone is not enough: under Elementary and Secondary Education Act (ESEA) assessment and accountability systems, it is imperative to understand why a particular model is selected to produce a growth measure.

One common error associated with the selection and use of any growth model is attempting to adapt a methodology whose interpretations and technical requirements cannot be supported by the underlying data and their characteristics. Depending on the design of the alternate assessment, it may not be feasible to apply the same growth model to the alternate assessment that is used for the general assessment. To determine feasibility, it is important to consider:

1. the claims a state wants to make for growth on the alternate assessment;
2. the claims a state wants to make for growth on the indicator (assuming the measure can be aggregated across the general and alternate assessments); and
3. how to best evaluate those claims using the evidence.

There is often overlap with the priorities and criteria associated with the growth model for general assessments— but before assuming that the same model can be applied to the alternate assessments, it is important to consider several questions.

Under ESEA assessment and accountability systems, it is imperative to understand the reasons why a particular model is selected to produce a growth measure. D'Brot, LeFloch and English (2020) offer a comprehensive evaluation process for Every Student Succeeds Act (ESSA) accountability systems, indicators and measures. Their generalizable, three-part reflection can be applied to growth measures:

1. What policy goals is the state trying to meet, and what mechanisms are in place to support them (theory of action)?
2. What evidence is available that the measures and indicators are operating as intended (technical quality of design)?
3. What behaviors are incentivized based on the design of the measure or indicator (impact of design)?

This chapter will address these three questions in relation to the selection of growth models and implementation of growth measures in assessment and accountability systems.

HOW SHOULD GROWTH MODELS BE SELECTED?

In order to select and apply growth models to the AA-AAAS, a state must first define *student growth* for students with the most significant cognitive disabilities (SCD). Is the state interested in describing the academic progress of an individual student/cohort of students with SCD over time, or does the state need an evaluative or predictive judgment that relates to the content standards as well? The former is likely more straightforward, as the latter will require more technical information, as well as a stronger communications plan to support interpretation and some acknowledgement that results are a projection and not a guarantee of future performance.

Growth measures provide more detail about individual students than snapshot performance measures. Thus, the application of growth widens the view of student performance on an assessment. Depending on the grain size of change examined in each model, the interpretation of gains can vary significantly. For example, growth models which rely on scale scores communicate changes in scale-score increments. Growth models that rely on categorical descriptions (e.g., transition tables using performance levels) communicate changes only in terms of the number of categories available.

This section provides a starting point from which to address the following topics when considering possible growth models:

- What questions should states consider when selecting a growth model?
- What aspects of the AA-AAAS are important to consider?
- What inferences are states trying to make when using a growth model?
- What technical characteristics for a growth model should be considered?

This chapter does not provide a prescriptive, step-by-step guide for selecting a growth model for assessment and accountability systems; rather, it presents a series of questions that can help inform both the policy and technical constraints states may face as they identify a growth model to support students with SCDs, their educators and students' families.

WHAT QUESTIONS SHOULD STATES CONSIDER WHEN SELECTING A GROWTH MODEL?

Before considering whether a growth measure can support a school-, district- or state-wide growth indicator, it is important for states to consider a series of questions regarding both the policy and technical issues stemming from the use of growth models for AA-AAAS. The following table provides sample questions, which will be addressed in greater detail in the rest of this chapter.

FIGURE 4-1: CONSIDERATIONS WHEN SELECTING A GROWTH MODEL (adapted from D'Brot 2017)

POLICY CONSIDERATIONS	TECHNICAL CONSIDERATIONS
P1 - What policy goals or signals does the state want to send in relation to student growth on the alternate assessment? P2 - In what ways should growth on the AA-AAAS inform conversations about individual student progress? P3 - How can the application of a growth model on the alternate assessment improve outcomes by promoting progress (rather than making a summative determination or punitive judgments) for both lower- and higher- performing students? P4 - How can the growth model be applied equitably when characterizing student growth across critical subgroups? How does the state develop the measure, and how can the state communicate about its meaning in way(s) that minimize unintended consequences?	T1 - What are the technical characteristics of the AA-AAAS? What scores are produced? How many students take the AA-AAAS? T2 - Are growth-model calculation requirements supported by the technical characteristics of the AA-AAAS? (This may include issues associated with the stability of the population, not just the sample size of students.) T3 - What are the technical characteristics of the growth measure? How might they inform inclusion in the growth indicator (discussed in Chapter 5)?[6]

The selection of a growth model is constrained by the technical characteristics of the AA-AAAS and growth-model requirements, but when determining where to start, it is helpful to consider the ramifications for policies on promotion, grade-level placement, and who qualifies for alternate assessment. States might ask the following questions to spur conversation around the AA-AAAS and identify appropriate, potential growth models.

[6] The topic of the growth indicator will be discussed in more detail in Chapter 5.

How are expectations for students with SCD defined conceptually?

As noted previously, "significant cognitive disability" is not a disability category under the Individuals with Disabilities Education Act (IDEA), but states administer the AA-AAAS to give students an opportunity to demonstrate the knowledge, skills and understandings (KSUs) associated with extended standards. States' grade-level academic expectations for students with SCD may be expressed in alternate or extended-content standards. Achievement-level descriptors (ALD)[7] for each grade's alternate assessment provide content-based descriptions of what students are expected to achieve. The KSUs described in ALD can be one way of understanding the state's grade-level expectations for students with SCD from grade to grade.

Interpreting data on student achievement based on grade-level expectations also requires an understanding of who is being assessed, as well as an assessment system that includes appropriate accessibility supports or accommodations. While students with SCD are not federally defined, the common characteristics shown in Chapter 1 underscore the need for a clearly specified set of expectations for students with SCD, as well as well articulated definitions of assessment participation criteria. This includes a shared understanding of the conditions under which students will be assessed, including options for students to flexibly provide responses to item presentations, provision and monitoring of needed accommodations and other Universal Design features.

Monitoring assessment accommodations may be challenging, but a clear understanding of what accommodations students are eligible for (and which ones they receive) can help states identify whether students universally are given access to the types of accommodations during assessments which they receive instructionally. When reviewing student performance on the AA-AAAS, it is important to keep in mind that the defensibility of scores depends upon a state's confidence in the standardized access to and provision of needed accommodations — that is, the state is defensibly confident that a given score indicates the student's achievement relative to expectations and is not related to other factors, such as missing or mismatched accommodations.

Beyond gaining a conceptual understanding, states can determine how standards-based expectations and performance on the AA-AAAS are framed, which will affect how they are used as part of a growth measure. For example, does the AA-AAAS result in raw scores, scale scores, performance levels and/or performance profiles? Depending on assessment results, test designers then can review the growth-model types (see Chapter 2) and determine whether certain models could be appropriately applied to the AA-AAAS results.

[7] Also called performance-level descriptors (PLD).

How is academic growth for students with SCD defined conceptually?

Like the conceptual definition of performance expectations for students with SCD, the conceptual definition of growth for students with SCD is based on the extended standards. However, a conceptual understanding of growth is dependent on the way in which the extended standards are vertically articulated across grades and grade spans, if specified. While it is expected that the KSUs will build over time, designers should understand the degree to which the complexity of the extended standards also increases over time. This can help a state understand the connection between the extended standards, the KSUs reflected in those standards, and how the extensions and KSUs build upon one another — which can help inform an evaluation of students' year-over-year growth.

Additionally, students who take the alternate assessment may demonstrate more variability in their growth over time, compared to students who take the general assessment, and this variability may skew the distribution of growth based on a prior score. Understanding the characteristics of students' score histories and what they represent may also affect how the state defines its *referent group*, in order to interpret growth, which is raised in the next paragraph.

A conceptual understanding of growth for students with SCD is important when interpreting student progress from grade to grade. Furthermore, focusing on student growth on the extended standards can help educators establish and articulate that students with SCD can progress toward and move beyond high expectations in instruction and assessment. This may be supported by establishing conceptual growth expectations that can be applied to students across the performance continuum on the AA-AAAS. This is an important assumption, which also can reinforce the idea that — while students who remain on track with content standards across the years demonstrate acquisition of new content each year — students with more significant impairments (who may score at the lower end of the performance continuum) still should be expected to learn. By improving their understanding of growth on the alternate standards, states can improve their understanding of how to interpret changes in AA-AAAS scores over time.

As part of developing a conceptual definition of growth, states should consider the desired referent point for growth. That is, should student growth be compared to other, similar students? Should growth be compared to an individual's historical performance? Or should growth be compared to target expectations on extended standards at some point in time? Depending upon the answer to these questions (which are explored more fully below), states can cross-reference the growth model types listed in Chapter 2 to determine the most appropriate model.

> "[F]ocusing on student growth on the extended standards can help educators establish and articulate that students with SCD can progress toward and move beyond high expectations in instruction and assessment."

WHAT ASPECTS OF THE AA-AAAS ARE IMPORTANT TO CONSIDER?

Before considering the most appropriate growth model to apply to an AA-AAAS, states first must understand the design and administration of the AA-AAAS and the types of scores it produces. With regard to its design, the alternate assessment must appropriately reflect the intended reduction of both the depth and breadth of the extended standards. (See Peer Review Critical Element, 2.1, Tindal, et al., 2016.) Assuming the assessment appropriately covers the extended standards, other aspects of assessment design should be confirmed, including AA-AAAS claims, specification of achievement level descriptors (ALDs) and whether ALDs are vertically articulated. How recently were content standards changed? If so, when can the state reasonably expect that content standards will be implemented with fidelity, and at what point will the assessment reflect the latest content standards?

Another design consideration revolves around how alternate assessment results are calculated. For example, are scores on the AA-AAAS a product of a single testing window during the spring, or do they reflect multiple administrations over the course of the year? While both approaches likely could support a spring-to-spring analysis of growth for the purposes of accountability, interpreting a within-year series of administrations will greatly depend on the test design, blueprint and alignment to the enacted curriculum throughout the year. For example, assessments that are embedded in the curriculum should map tightly to the content addressed. Ideally, students who have been exposed to the content as part of a curricular sequence should then demonstrate proficiency on the delivered content. Over the course of the year, students should perform equally as well on each subsequent, embedded assessment — if they have met expectations with the content instructed during the span of that lesson, unit or course portion.

In addition to design considerations, AA-AAAS administration conditions will affect determinations regarding which growth models are appropriate. Administration characteristics will have a direct impact on the technical considerations which will be raised in the next chapter, but it is important to understand administration conditions that might influence score interpretations. For example, administration constraints could include the number of grades being tested, the total population being tested, the accessibility supports that are available and/or allowed during test administration.

Perhaps most importantly, identifying and documenting administration practices can help states compare the current year's assessment administration to previous years' administration. By ensuring that the AA-AAAS is administered with fidelity consistently across years, states can detect changes in performance over time which are a function of student communication and performance, not idiosyncrasies in test administration.

As an example of how the structure of an assessment can influence growth-model selection, Nehler, Clark and Karvonen (2019) examined assumptions regarding distributions, scaling, sample sizes and matched records for students participating in the Dynamic Learning Maps® (DLM) alternate assessment. Each of these factors was found to influence the ability to reliably and validly measure students' growth. For example, there were no vertical scales or interval-level scales when comparing student mastery of KSUs across years. Similarly, sample sizes varied drastically and could be prohibitively small after splitting the sample across model, grade and subject. Ultimately, the team concluded that additional work would be needed to facilitate reliable and valid measurements of student growth for the DLM assessments. It is important for states to conduct these types of evaluations when considering the types of inferences desired.

WHAT INFERENCES ARE YOU TRYING TO MAKE WHEN USING A GROWTH MODEL?

The application of growth models to educational assessment has been predominantly in the realm of school accountability. ESSA requires that states select an academic achievement indicator that may include a measure of student growth (see ESSA, 2015). While accountability-focused metrics are typically used for attributing student performance to a particular school, it is helpful to specify the intended purpose of a given model. Because the focus here is on selecting a growth model, the first decision will be whether to make descriptions or support predictions using growth data (see Castellano & Ho, 2016; D'Brot, 2017). Growth measures can be interpreted in one (or both) of the following ways:

Descriptive: Interpretation is based on previously observed student performance and is typically used to describe growth, retrospectively, for students or groups of students.

Predictive: Interpretation is based on possible future student performance and is typically used to describe prospective growth for students or groups of students. Some projections also include quantifications of growth-to-target or growth-to-standard (i.e., the distance to proficiency).

These interpretations can help narrow the selection of possible growth approaches. Further, some models can satisfy both descriptive and predictive interpretations. However, one should be careful not to assume that any predictive interpretations are set in stone or guaranteed. That is, projections of student performance are simply informed by prior test scores and either projected forward or compared to some future target. In many cases, states will try to break the projection by improving student learning or by accelerating the acquisition of student KSUs.

WHAT TECHNICAL CHARACTERISTICS FOR THE GROWTH MODEL SHOULD BE CONSIDERED?

The technical characteristics of the alternate assessment will have a strong influence on selecting a growth model. While this concept will be explored further in Chapter 5, the characteristics of any individual component (in this case the AA-AAAS) will have an impact on the reliability of the growth measure, and subsequently on the growth indicator. When selecting a growth model, one must consider the practical and technical characteristics of the AA-AAAS (particularly the test score properties) to determine whether certain growth models will be supported. More specifically, the statistical foundation upon which a growth model is built must be first supported by the test score properties of the alternate assessment.

As noted in the Chapter 2, minimally, states would need at least two measures of achievement from two different points in time to calculate a growth measure. The extent to which growth can be interpreted will be in part determined by the design of the test and the distance between administrations; however, at least two points for a student or cohort of students are needed in order to calculate the growth measure. Additionally, the AA-AAAS must demonstrate sufficient evidence of reliability, validity and fairness (see AERA, APA, & NCME, 2014) to support the requirements of the Individuals with Disabilities Education Act (IDEA) and ESSA. The following table, extracted from a table initially presented in Chapter 2, can be used to lay out growth-model options related to the characteristics of the data. Several considerations are made for the models presented below.

FIGURE 4-2: SELECTED CHARACTERISTICS OF GROWTH MODELS*

	CATEGORICAL GAIN	SIMPLE GAIN SCORE	RESIDUAL GAIN SCORE	STUDENT GROWTH PERCENTILES	VALUE ADDED	GROWTH-TO-TARGET
INTENDED CLAIM	Degree of student growth related to performance levels	Magnitude of student growth expressed in scale scores	Degree of student growth as compared to expected growth, with expected growth based on prior score	Degree of student growth as compared to expected growth, with expected growth based on students with similar academic scores over time	Degree of student growth as compared to expected growth, with expected growth based on multiple predictive variables	Degree of student growth relative to growth needed to reach a specified target in the future
DATA REQUIREMENTS	Vertically articulated cut scores across grades	Vertically scaled scores	Meaningful scores in terms of characterizing level of attainment of targeted KSUs	Meaningful scores in terms of characterizing level of attainment of targeted KSUs; large sample sizes	Sufficient longitudinal data to provide stable results; meaningful predictors	Meaningful scores in terms of characterizing level of attainment of targeted KSUs

*Sources: Castellano & Ho (2013a); Domaleski & Hall (2016); Data Quality Campaign (January 2019)

For the purposes of considering the technical characteristics of the AA-AAAS, growth-model types will be grouped into the following categories: (a) categorical gains, (b) simple gain scores and (c) regression-based analyses, each of which having unique data constraints. These technical characteristics are discussed in further detail below.

Categorical Gains

A **categorical-gain** growth model is a descriptive growth model that attempts to present growth as the degree to which students have changed across performance levels. In many current categorical-gain models, categories are split into sub-categories by dividing score ranges into smaller segments. These sub-category models are a common approach in state accountability systems and typically are referred to as *transition* or *value* tables. In transition/value tables, students who make progress across performance levels or sub-levels will typically net schools or groups an increasing number of points, based on prior-year performance.

The following example of a value table reflects how a growth indicator may be designed based on aggregated growth measures (which is discussed in Chapter 5). See Table 4-3 for an example of a transition/value table using categorical gains. Please note that this example transition/value table would be used to aggregate student scores into a school-level, district-level, or state-level score.

FIGURE 4-3: AN EXAMPLE TRANSITION/VALUE TABLE USING SUB-CATEGORIES, OR PERFORMANCE LEVELS (adapted from Hill, et al. 2005)

		CURRENT YEAR							
		L1 Low	L1 High	L2 Low	L2 High	L3 Low	L3 High	L4 Low	L4 High
PRIOR YEAR	L1* Low	0	120	160	185	200	200	200	200
	L1 High	0	90	130	150	195	200	200	200
	L2 Low	0	50	95	130	165	175	195	195
	L2 High	0	30	55	95	130	160	185	195
	L3 Low	0	0	30	80	100	130	150	175
	L3 High	0	0	0	30	70	105	135	160
	L4 Low	0	0	0	0	40	75	115	145
	L4 High	0	0	0	0	25	50	95	125

* "**L**" refers to the performance level of an assessment (e.g., Performance Level 1).

As is shown in this value table, there must be enough variation within performance levels to support splitting them into two sub-categories or performance levels. While there may be sufficient precision around the middle of the test scale, there may be increasing amounts of error as one approaches the edges of each test scale (see Briggs & Betebenner, 2009). As the standard error of measure increases (typically around the tails of the scale), it becomes more likely that student performance estimates will "bounce" around split performance levels, demonstrating more volatility than appropriate to yield defensible interpretations of student growth.

Adaptive tests can reduce the amount of measurement error at the tails of the scale, but require many more resources to implement well (e.g., larger item pool, an appropriately designed adaptive engine, administration systems capable of delivering an adaptive test and sample sizes large enough to test the needed number of items). It is important to note that the assessment administration rules associated with a minimum attempted items requirement also will have an impact on score distributions, as scores bases on the AA-AAAS may be positively skewed if the threshold of items required for a valid attempt is not reached.

When considering the use of performance-level value tables or categorical gains, it may be likely that students taking the AA-AAAS will stay in the same categorical level across years, even when students' performance may increase or decrease (Tindal et al., 2014b). This insensitivity to student-level growth also may be replicated or even exacerbated when categorical gains are aggregated to the school level. In practical terms, this will require an evaluation of whether students are actually moving across sub-categories or the sub-categories are sensitive enough to detect change.

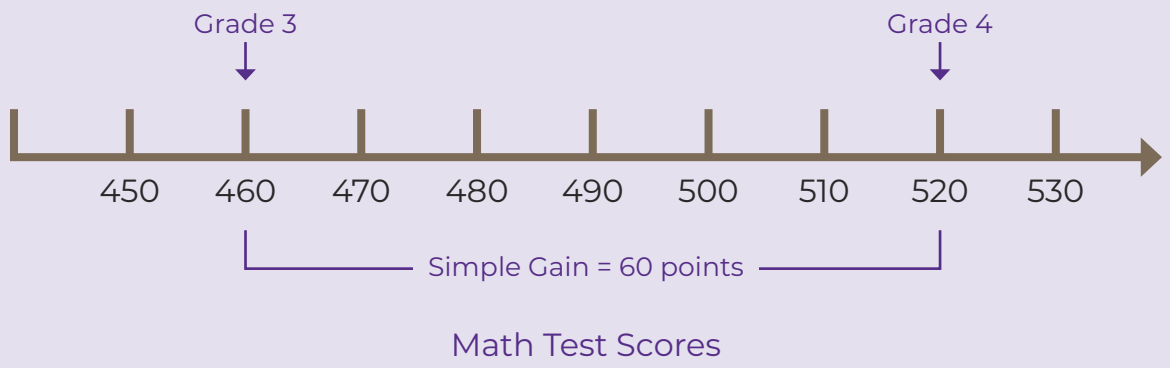
That is, the persistence of student placement in sub-categories will decrease the amount of information that growth on the AA-AAAS contributes to models of school quality (e.g., growth indicators or school differentiation scores). An alternative option may be to decrease the size of sub-categories by increasing the number of sub-categories (e.g., from two sub-categories to three sub-categories within a performance level, if empirically supported). However, decreasing the size of the sub-categories may be problematic, because the score range may approach or exceed the conditional standard error of measure, which makes interpretations about changes across sub-levels less defensible.

The potential insensitivity of categorical gains based on the AA-AAAS can mask gains that are observed instructionally, potentially raising incoherence with educator experiences and reports of student growth. However, it is worth noting that categories of performance are easier to understand and categorical gains are conceptually less demanding, making them potentially easier to promote publicly.

Simple Gain Scores

Simple gain scores are used when attempting to quantify the amount of growth a student has demonstrated based on a comparison of scores from an assessment conducted at two different points in time. Typically, this type of growth model requires the use of a vertical scale in order to make interpretations. A simple gain score can be either descriptive or predictive (to a degree), and individual differences can be aggregated to show average gains at a student group, grade, school, district and state level. It is possible to calculate a student's "distance to standard," but the defensibility of such a projection can be questionable, given the single-student observation. Predictions can be calculated by extrapolating gains into the future, but they rely on the same, vertical-scale assumptions as the description of student growth. The figure below presents a sample of the simple gain score.

FIGURE 4-4: SIMPLE GAIN SCORE MODEL



As shown in the figure above, this type of growth model appears to be relatively intuitive upon first glance. However, there are many assumptions associated with the use of simple gain scores. For example, vertical scale scores arguably do not have equal interval properties to support cross-grade interpretations of gain scores, and different content or scale characteristics may differ significantly. This problem is exacerbated if there are too few items at certain parts of the scale, which will limit the amount of information that can be used to support consistent inferences about student performance. That is, to defensibly and consistently interpret vertical scale score gains, vertical scales need to demonstrate equal interval properties in order to functionally compare changes at different points of the scale (Betebenner & Linn, 2009). This characteristic of vertical scales is a difficult assumption to attain (Ballou, 2008).

Furthermore, simple gain scores are subject to floor and ceiling effects. For example, if the highest or lowest obtainable scale scores on a given scale are not consistent, students who are performing at the extremes may have inconsistent gains over time. Additionally, students with particularly high prior scores may demonstrate negative scale-score gains despite their consistently high performance, corresponding to the increasing grade-level expectations or content. Conversely, students with particularly low prior scores may demonstrate artificially large and positive scale-score gains, primarily because they can only move up the scale. These scale dependencies, coupled with the difficulty associated with interpreting a scale score in general, can make understanding scale-score gain difficult.

The difficulty of this interpretation exists for both the general assessment and the AA-AAAS, but for the AA-AAAS this difficulty may be exacerbated by the way in which the alternate content standards are designed or how the assessment is administered, scored or reported. For example, as grade levels increase, extended standards may reflect an increasing distance from the general standards as the student transitions from elementary to high school. A state may be able to improve user understanding of scale scores and simple gains by providing sample items or concrete descriptors of performance as they relate to student scale scores on the assessment.

Regression-Based Analyses

Regression-based analyses include those conditional analyses that express scores in terms of expectations based on past scores (e.g., residual gains, student growth percentiles) and models that use entire student-score histories as an outcome — i.e., an outcome that compares observed scores to differences from expected growth using multiple predictive variables (multivariate and value-added models). Generally, conditional models are those which, based on prior scores, establish expectations for the student's current score. The observed score then is compared to the expected score, conditioned on past scores.

Residual gains, student growth percentiles (SGP), multivariate and value-added models simply express the comparisons between current scores and the prior scores that were used to set up conditional expectations for current scores in different ways. In this way, regression-based models can be either descriptive or predictive, depending on whether projections are made using prior observations. For example, residual gains and value-added models can yield statistical expectations for a student, given past performance, that can be compared to a target cut score; SGPs can be used to calculate adequate growth percentiles to meet a future standard, based on some time bound (e.g., three years).

These types of growth models tend to demonstrate greater precision in their estimates and higher amounts of variability in the range of growth estimates that students might demonstrate using coarser-grained models. However, their primary limitation is that they provide normative expressions of student growth and are thus dependent on the performance of the group against which they are normed, which will require the state to define the normative group.

When considering students who take the AA-AAAS, the norming group is constrained and will reduce the availability of scores to inform growth estimates. Additionally, the general trends of performance over time can impact how norm-based growth is interpreted. For example, if all students are demonstrating an exceptional amount of growth on an assessment, then the lowest amounts of conditional growth are actually sufficient to meet growth to standard. Conversely, if the entire set of students is declining, then those students who are declining the least are estimated to have the highest growth, which may actually be indicative of losing content over time. This limitation is addressed in the following section, which discusses growth-to-standard.

The increased precision associated with regression-based analyses can be informative when examining student growth, but the data limitations begin to escalate around sample sizes and score types. There must be a sufficient sample of students who take the AA-AAAS to calculate growth measures using regression-based models. States can leverage technical advisory committees (TAC) to understand how limited sample sizes might affect the calculation of growth measures (see Faul, Lang, & Buchner, 2007) to explore the limits of minimum sample sizes to calculate growth measures. However, there is a practical difference between the analytical precision required for state-level analyses and what is needed for school-level aggregations — which is an important distinction to make when examining the reliability of growth analyses year over year.

While the precision of an aggregate growth estimate for a school might improve as the sample size increases, the standard error of estimates is actually constant, regardless of the aggregated sample size (see Betebenner, DePascale, Marion, Domaleski and Martineau, 2016). Therefore, states should be mindful of the total population of students with SCD needed to calculate regression-based analyses. Specifically, examinations of variability suggest that states should require a minimum sample size of 1,000 for SGPs per grade, whereas other regression-based analyses require considerations of the nesting variable counts and desired power (Betebenner, p.c.; Castellano & Ho, 2013b). However, sample size alone may be an insufficient consideration, as states also need to examine whether students are distributed across the score scale, in order to ensure that different score histories are available to inform the model.

In addition to sample size, number of predictors will improve the estimation of regression-based analyses. As the number of predictors increases (e.g., prior scores or predictive variables) the precision, accuracy and validity of regression-based analyses improve. The primary challenges here are that improving the accuracy of estimates requires an increasingly large number of continuous data points, reflecting student histories and identifying which predictors are appropriate to use for students who take the AA-AAAS. A practical challenge with missing data revolves around the potential "missingness" of students who may not be included in the AA-AAAS.

For example, as many as 25% of students may switch between the alternate and general assessment from year to year (see Saven, Anderson, Nese, Farley, & Tindal, 2016), which poses a continuity challenge. In cases where states are seeing little variation in their student population, there may be variation regarding who is retained in growth analyses, due to statutory requirements. ESSA imposes a 1% statewide cap on students who take the alternate assessment, and consistently identifying students who take the assessment year over year can have an impact on the stability of growth estimates over time.

While regression-based models have similar data constraints, the inferences made based on residual gain scores, value-added scores and SGPs vary greatly. The key to defensibly using these growth models is to check that the underlying statistical assumptions are met. Readers are encouraged to review Castellano and Ho's (2013a) review of growth models to understand the nuances of calculating and interpreting regression-based growth models.

Considerations for Growth Models in General

For all analyses, a consideration of sample sizes and the volatility of students who are identified as having SCD will be critical. For example, if a given school or district has major changes year over year with regard to the sample, whether in composition, size or characteristics, then the interpretations for growth will vary. Furthermore, any major year-over-year changes in student composition will have implications for aggregating growth measures into a single growth indicator for a school/district and the consistency of growth indicators over time. Readers are encouraged to refer to Chapter 5 to consider how growth measures might influence growth indicators within an accountability system.

SUMMARY

Chapter 4 has introduced the reader to how growth models can be used to produce growth measures using contemporary AA-AAAS, including:

- key questions to ask when considering the policy and conceptual definitions of growth on extended standards and the alternate assessment;
- considerations for the AA-AAAS; and
- examination of the technical characteristics of growth models when applied to the AA-AAAS.

This chapter also has laid the groundwork for evaluating growth as an indicator (which will be covered in Chapter 5) and built upon the prior chapters, as a strong theory of action is required to evaluate (a) whether a growth model is being applied to growth measures appropriately and (b) whether measure aggregations lead to sound interpretation of growth indicators (the latter of which is addressed in Chapter 5).

Chapter 4 covered several concepts, including policy goals, general growth inferences and technical considerations for growth. This chapter also supported the reader in understanding how the selection and application of growth measures must be compared to the theory of action, and it offered suggestions for considering the underlying data when selecting growth models based on AA-AAAS scores.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. https://www.apa.org/science/programs/testing/standards

Ballou, D. (2009). Test Scaling and Value-Added Measurement. Education Finance and Policy. 4(4): 351–383. https://direct.mit.edu/edfp/article-abstract/4/4/351/10099/Test-Scaling-and-Value-Added-Measurement?redirectedFrom=fulltext

Berk, R. A. (2004). *Regression Analysis: A Constructive Critique*. SAGE Publications, Inc. https://methods.sagepub.com/book/regression-analysis

Betebenner, D. W., DePascale, C., Marion, S., Domaleski, C., & Martineau, J. (2016). Precision, Interpretability, and Utility of SGPs: A response to *Why we should abandon student growth percentiles* by Sireci, Wells, and Keller. The National Center for the Improvement of Educational Assessment: Dover, NH. https://www.nciea.org/sites/default/files/publications/CFA-Sireci_Response.pdf

Betebenner, D. W. & Linn, R. L., (2010). *Growth in Student Achievement: Issues of Measurement, Longitudinal Data Analysis, and Accountability*. Paper presented at the Exploratory Seminar: Measurement Challenges within the Race to the Top Agenda, December, 2009. Educational Testing Service. https://www.semanticscholar.org/paper/Growth-in-Student-Achievement-%3A-Issues-of-%2C-Data-%2C-Betebenner-Linn/4c3f7752cdce6f7b3b99c800bb78692bab439e91

Briggs, D. C., & Betebenner, D. W. (2009, April). *Is Growth in Student Achievement Scale Dependent?* Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA. https://www.semanticscholar.org/paper/Title%3A-Is-Growth-in-Student-Achievement-Scale-Briggs-Betebenner/36001318cc5c95a9cae8283cb39db9a5441a0df8#paper-header

Castellano, K. E., & Ho, A. D. (2013). A Practitioner's Guide to Growth Models. Council of Chief State School Officers. https://scholar.harvard.edu/andrewho/publications/practitioners-guide-growth-models

Castellano, K. E. & Ho, A. D. (2013b). Contrasting OLS and Quantile Regression Approaches to Student "Growth" Percentiles. *Journal of Educational and Behavioral Statistics, 38,* 190-215. https://doi.org/10.3102/1076998611435413

D'Brot, J. (2017). Considerations for Including Growth in ESSA State Accountability Systems. The Council of Chief State School Officers. https://ccsso.org/sites/default/files/2017-10/CCSSOGrowthInESSAAccountabilitySystems1242017.pdf

D'Brot, J., LeFloch, K., English, D., Jacques, C. (2020). *State Support Network: Evaluating State Accountability Systems*. American Institutes for Research. https://oese.ed.gov/resources/oese-technical-assistance-centers/state-support-network/resources/evaluating-state-accountability-systems-esea/

Domaleski, C., & Hall, E. (2016). *Guidance for Estimating and Evaluating Academic Growth*. http://www.ncscpartners.org/Media/Default/PDFs/Resources/EstimatingandEvaluatingStudentGrowth.pdf

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/bill/114th-congress/senate-bill/1177

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. https://www.researchgate.net/publication/259703298_GPower_3_A_flexible_statistical_power_analysis_program_for_the_social_behavior_and_biomedical_sciences

Hill, R., Gong. B., Marion, S., DePascale, C., Dunn, J., & Simpson, M.A. (2005). *Using Value Tables to Explicitly Value Student Growth*. The National Center for the Improvement of Educational Assessment. https://www.nciea.org/publications/MARCES_RH07.pdf

Nehler, C., Clark, A., & Karvonen, M. (2019). *White Paper: Considerations for Measuring Academic Growth on Dynamic Learning Maps® (DLM®) Alternate Assessments*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems (ATLAS). https://dynamiclearningmaps.org/sites/default/files/documents/publication/Growth_White_Paper.pdf

Saven, J. L., Anderson, D., Nese, J. F. T., Farley, D., & Tindal, G. (2016). Patterns of Statewide Test Participation for Students with Significant Cognitive Disabilities. *The Journal of Special Education*, 49(4), 209–220. https://journals.sagepub.com/doi/abs/10.1177/0022466915582213

Tindal, G., Stevens, J., Nese, J. F. T., Schulte, A., Elliott, S. N., Kurz, Al., & Buckley, J. (2014a, April). *Critical Issues in Studying Growth on State Tests for Students with Disabilities*. Council for Exceptional Children (CEC). https://www.ncaase.com/docs/NCAASE_CEC_fini.pdf

Tindal, G., Nese, J. F. T., Farley, D., & Saven, J. (2014b, April). *Documenting Growth for Students with Significant Cognitive Disabilities*. National Council on Measurement in Education (NCME). https://www.ncaase.com/docs/TindalNeseFarleySavenNCME_AltAssessV3.pdf

CHAPTER

5

INCLUDING GROWTH MEASURES FROM THE AA-AAAS IN GROWTH INDICATORS

INCLUDING GROWTH MEASURES FROM THE AA-AAAS IN GROWTH INDICATORS

INTRODUCTION

Chapter 1 covered background information relevant to using alternate academic achievement standards (AA-AAAS) in growth models; described the students who take the AA-AAAS and their education; and reviewed both educational rationales and the legal requirements for including these students in assessment and accountability. Chapter 2 oriented readers to growth models, growth measures and growth indicators from a conceptual perspective. Chapter 3 described how a theory of action can support the development of growth measures and indicators; it also provided an overview of how using a theory of action can help test developers document and establish the validity of inferences drawn from, and appropriate use of, any policy-related measure — not just growth measures based on the AA-AAAS.

Chapter 4 introduced the reader to how growth models can be applied to produce growth measures for contemporary AA-AAAS, which lays the groundwork for incorporating growth on the AA-AAAS as an indicator in accountability systems.

Chapter 5 will focus on considerations associated with:

- requirements under the Every Student Succeeds Act (ESSA);
- questions to consider when designing an indicator for accountability systems;
- considerations related to the technical quality of growth measures and indicators; and
- the potential utility and impact of growth indicators for the AA-AAAS.

As in previous chapters, this chapter will build on the assumption that a strong theory of action is necessary to determine whether a growth model can be appropriately incorporated into an accountability system. This theory of action then can enable practitioners to evaluate the incorporation, interpretation and use of growth measures based on the AA-AAAS against pre-defined criteria.

The following figure outlines how the questions posed throughout this chapter are organized, based on whether they address the state's policy goals; the technical quality of the growth model, measures and indicator; and the impact and utility of data-reporting on the growth indicator.

FIGURE 5-1: CONCEPTS RAISED IN THIS CHAPTER RELATED TO A STATE'S GROWTH INDICATOR BASED ON ALTERNATE ASSESSMENT

POLICY GOALS

Alignment with the overall system theory of action
Fair representation of the construct

TECHNICAL QUALITY

Conceptualization of reliability in accountability
Sufficient internal technical quality for the assessment
Technical quality of combined measures
Contribution to the state's system of differentiation

IMPACT AND USE

Utilty of data as representations of progress for students with SCDs
Interpreting data based on the growth indicator
Connection to other data to inform continuous improvement
Coherent links between accountability data and continuous improvement data

Before addressing these topics and their associated questions, it is important to understand the requirements for growth under ESSA and also review information on growth models, measures and indicators.

REQUIREMENTS FOR GROWTH UNDER ESSA

The Every Student Succeeds Act (ESSA) required state accountability systems to be operational in the 2017-2018 school year in order to start making school determinations in the 2018-2019 school year. In light of the COVID-19 pandemic, all states requested waivers from accountability using spring 2020 assessments, and many requested an addendum to their waivers to eliminate differentiation requirements using spring 2021 data. As a result, many states today are considering how to best restart, revise or phase in accountability systems that use growth for school years 2020-2021 and 2021-2022. As a reminder, states are required to establish and measure the following annual indicators for all students, and separately for each subgroup (see D'Brot, 2017):

1. Proficiency in reading/English language arts and mathematics based on performance on state assessments, with the option to measure academic growth for high school students in these subjects over time;
2. For public elementary schools and secondary schools that are not high schools, student growth or another valid and reliable indicator that allows for meaningful differentiation in student performance;
3. For high schools, graduation rates based on the four-year adjusted cohort rate, with the option to include extended-year adjusted cohort rates;
4. For English learners (ELs), progress in achieving English language proficiency (ELP), based on the state's ELP assessment, within a state-determined timeline; and
5. At least one indicator of school quality or student success that (a) allows for statewide differentiation of school performance and (b) is valid, reliable and comparable, which may vary by elementary, middle and high schools.

These requirements for accountability can be grouped into those that include reporting on student subgroups as specified in ESSA and those with school-differentiation requirements based on a state-determined set of goals and school-differentiation methodologies. The remainder of this chapter focuses on the second requirement, student growth, as the other academic indicator.

Status, Improvement and Growth

For the purposes of this chapter, one must revisit the relationship among growth **models**, **measures** and **indicators**, which were defined in Chapter 2. (See page 22). Before learning how these growth *models* and growth *measures* inform growth *indicators*, it is important to recall how **growth** differs from **status** and **improvement**. Status is akin to proficiency, and it refers to the performance of a student or collection of students at a single point in time. Improvement refers to the change in performance at the aggregate level over time within or across grades, without following the same student or cohort of students. Growth refers to the change in academic performance of the same student or same cohort of students over two or more points in time.

When considering growth, there are three distinctions to consider. As a reminder, a growth *model* is the procedure through which scores are calculated to arrive at a growth *measure*. Growth measures also may be presented alongside a student's individual score report (ISR) to assist in interpretation. These measures then usually are aggregated in the educational accountability system as a growth *indicator* to reflect the degree to which students in a school, district or state grew academically. If a state uses multiple growth models (for example, one for the general assessment and one for the alternate assessment), the indicator may be based on a combination of summary growth measures. The indicator may take different forms, depending on the nature of the accountability system; for example, some states use index systems, and the growth summary is transformed into an index score.

For example, a state with an alternate assessment based on categories of performance may use a categorical gains model; if the general assessment is based on scale scores, the state may use a conditional model (e.g., student growth percentiles). States can combine these two growth measures in different ways. For example, adjustments can be made to student-level estimates pre-aggregation or to group estimates post-aggregation. The latter may be accomplished by, for example, producing a weighted composite of estimates from both the alternate and general assessments. In the case of the former, a state may create an adjusted metric — such as by transforming student growth percentile (SGP) estimates to performance bands (e.g., Low, Typical and High) to correspond to the number of performance-level changes on the alternate assessment's categorical gains model.

As noted previously, a state could apply business rules to develop overall growth indicators for each assessment, then create a weighted composite based on the number of growth scores from students taking the alternate and the general assessment, respectively.

The following section offers questions for states to consider when designing an indicator for accountability systems, as well as considerations for the reliability of measures and indicators and the potential utility and impact of growth indicators for the AA-AAAS.

QUESTIONS TO CONSIDER WHEN DESIGNING A GROWTH INDICATOR FOR ACCOUNTABILITY SYSTEMS

Under the Elementary and Secondary Education Act (ESEA), as amended by ESSA, states have the opportunity to embed an additional academic indicator within their systems of annual meaningful differentiation (AMD). While many states have elected to operationalize this as a growth indicator, it is important to address this overarching question in light of the state's theory of action for systems of AMD:

> What is the rationale behind the design, development and implementation of this growth measure for the other academic indicator, and how does a growth indicator conform to this rationale?

In order to answer this question, states could engage in the following four-step process to establish general criteria for making design, development and implementation decisions. While these steps are relevant for the other academic indicator required under ESSA, they are framed specifically for a growth indicator, as presented in the table below (See D'Brot, LeFloch, English, & Jacques, 2019 for a comprehensive discussion of these steps):

FIGURE 5-2: STEPS FOR DESIGN, DEVELOPMENT AND IMPLEMENTATION OF A GROWTH INDICATOR

KEY STEPS FOR DESIGN, DEVELOPMENT & IMPLEMENTATION	EXPLANATION OF KEY STEPS WITH A FOCUS ON SCORES FROM THE AA-AAAS
1. Identify the specific policy objective for the growth indicator.	What policy objective does the growth indicator serve? The policy goals should dictate how the growth indicator is selected and calculated (e.g., descriptive vs. predictive growth scores). If the objective is intended to balance the status focus of achievement in the overall system of AMD, it will be important to determine whether the way growth is calculated is contributing to the system as intended. Furthermore, making this determination can help states identify whether sufficient data are available to meet this policy objective. When considering scores based on the AA-AAAS, how are growth results being presented to students, educators and the public? In what ways should growth on the alternate assessment be treated similarly or differently from the growth indicator, based on the general assessment?
2. Identify the necessary policy mechanisms or levers used to meet this objective.	What measures are used to calculate the growth indicator? Are there any challenges with calculating, including or aggregating measures for this indicator? Does combining growth across the general and alternate assessments require making transformations or adjustments? How are states helping stakeholders understand how to interpret and use data from this indicator, given these adjustments, combinations aggregations or potential differences between the general and alternate assessments?
3. Determine the intended behaviors that this indicator is attempting to elicit.	What behaviors are states trying to incentivize through the way in which the growth indicator is operationalized for the alternate assessment, when compared to the general assessment? Are they trying to focus attention on a specific aspect of school performance or growth for students with SCDs? How will indicator results be communicated to parents, educators and the community? Are interpretations similar to or different than growth indicators using the general and alternate assessments?
4. Determine the expected results of the indicator to establish a comparison point.	It will be important to understand what expected results might look like from a given growth model. For example, a value table should yield students moving up, moving down and staying in the same performance categories or sub-categories. If students with SCDs are not demonstrating progress over time as expected, is it a function of the model, the assessment or the standards? These outcome-related issues may highlight challenges with calculating, including or aggregating measures for the indicator. If growth data are functioning as expected, it is more appropriate to then consider whether users understand how to interpret and use data from this indicator for students with SCDs, school- and district-wide.

In light of insights gained after following these steps, states are encouraged to address the following, specific questions as they consider how to combine growth measures into a single growth indicator for state systems of AMD.

How well does the growth indicator align with the overall system theory of action and its policy objectives?

There are several considerations to keep in mind when examining the design decisions that go into creating a growth model for elementary and secondary schools — e.g., what the overall policy objectives are; whether the growth indicator actually reflects student progress as intended; whether relevant design decisions adhere to ESEA requirements, etc. As one explores the construct of growth and its interactions with other indicators, it is significant to identify how the growth indicator is calculated, specifically using scores from the AA-AAAS.

It is important to understand whether or not the growth indicator based on the AA-AAAS fairly represents the construct as intended for students with SCDs. The growth indicator provides states with an opportunity to expand their conceptualizations of academic performance to include a focus on progress (e.g., individual student growth, cohort-based growth, or snapshots of improvement by grade over time). Another essential consideration is the type of information the growth indicator is intended to represent, and whether the results from the growth indicator based on the AA-AAAS actually reflect student progress as intended.

While growth indicators will vary in their complexity, they should supplement the academic achievement indicator and support the state's educational vision of communicating progress for all students, including those with SCDs. Depending on how this indicator is designed, it will be useful to help the public understand how this indicator serves the state's theory of action and how a focus on student growth facilitates progress toward the state's educational objectives.

How well does the growth indicator fairly represent the construct as intended?

When considering whether a measurement tool represents the construct well, it is typical to look to the Standards (2014) for guidance. Because the growth indicator for alternate assessment is based on scores from the AA-AAAS, reviewing and referencing documentation from the state's peer review, which is based on the standards for the alternate assessment, is essential. As part of peer review requirements, states are expected to document and present the technical quality associated with assessment system operations, validity, reliability, fairness and accessibility; inclusion of all students; alignment; and achievement standards (see U.S. Department of Education, 2018). This evidence is a powerful starting point to help determine whether the alternate (and general) assessment can defensibly contribute to a growth indicator.

Additionally, it is critical to consider how fairness may differ when considering the alternate assessment separately, as compared to indicators that comprise scores from the AA-AAAS in whole or part. When referencing measurement, fairness emphasizes that the test must be fair, accessible and appropriate for all individuals in the intended population for the intended use of that test (Standards, 2014). However, when considering fairness for indicators of academic growth, fairness and bias is focused — e.g., with regard to whether schools or group(s) have an equitable opportunity to demonstrate progress.

For indicators of growth based on the AA-AAAS, it is particularly significant to determine whether there are any potential sources of bias. For example, subgroup and school-level aggregates should be examined to determine whether there is sufficient variability in growth aggregates across key conditions (e.g., school size, demographic composition) to determine whether students are making progress.

The variability of growth and its influence on a school's growth indicator will be affected by the size of the total school population. Under most circumstances, there will be fewer students for whom growth can be calculated because of the need for prior scores to calculate a growth measure. When evaluating growth for alternate assessments, the reduction in available data due to lack of prior scores will be amplified by the reduced number of students who are eligible to take the AA-AAAS, especially when there may be fewer than one percent of students in a school who take the AA-AAAS.

To ensure that interpretations are being made in the context of the alternate assessment, it is key to investigate the comparatively small number of growth scores available for students who take the alternate assessment. Specifically, it is important to make sure that growth on the alternate is not being overshadowed by growth based on the general assessment. If so, it may be necessary to lean on the policy-focused goals of growth and reiterate the growth expected in the standards as students matriculate across grades.

A key goal of states' systems of AMD is to differentiate at the school level, so variability of growth scores also should be explored, at least at the school-level of the system. The variability of school-level growth indicators then can be examined across the state to determine whether there are systematic issues with regard to the growth indicator not having enough influence on overall school differentiation. If issues emerge, states then can determine whether the lack of variability is due to the business rules or aggregation decisions that might decrease the variability in the growth model's use, or whether the issues with decreased variability are a function of the model type (e.g., value tables vs. regression-based analyses).

While it is an imperfect comparison, the variability in growth measures based on alternate assessments can be compared to the variability in growth measures based on the general assessment to determine whether differences between the two sets of growth measures exist. If differences do exist, then it will be vital to understand the source of those differences. If, for example, there is substantially less variability in growth measures based on the AA-AAAS, it may raise questions about:

- the capability of the measure to detect change in student performance or school-level aggregates;
- the sensitivity of the scores (i.e., the ability of the model to differentiate changes in performance that may not be detected by rates of proficiency or performance index scores) on which the growth measure is base; or
- the characteristics of the standards upon which the assessment is built.

Despite these questions, growth estimates can be reasonably trustworthy estimates of student performance; nonetheless, to support interpretation and use of growth data, states would do well to rule out these factors.

While it is difficult to predict why scores may be less variable, critically examining growth measures based on the AA-AAAS can help states understand how to interpret the changes in growth scores or the assessment scores and ensure that groups of students with SCD have sufficient opportunity to demonstrate progress on the assessment. However, it is best for states and practitioners to rely on their interpretation of the expected growth in the standards (whether based on the general or alternate standards) to inform their interpretation of growth on the assessment. That is, if growth scores reflect some quantitative amount of progress, they need to be interpreted through the lens of the qualitative progress in content, based on standards.

QUESTIONS TO CONSIDER WHEN ADDRESSING THE TECHNICAL QUALITY AND RELIABILITY OF GROWTH INDICATORS

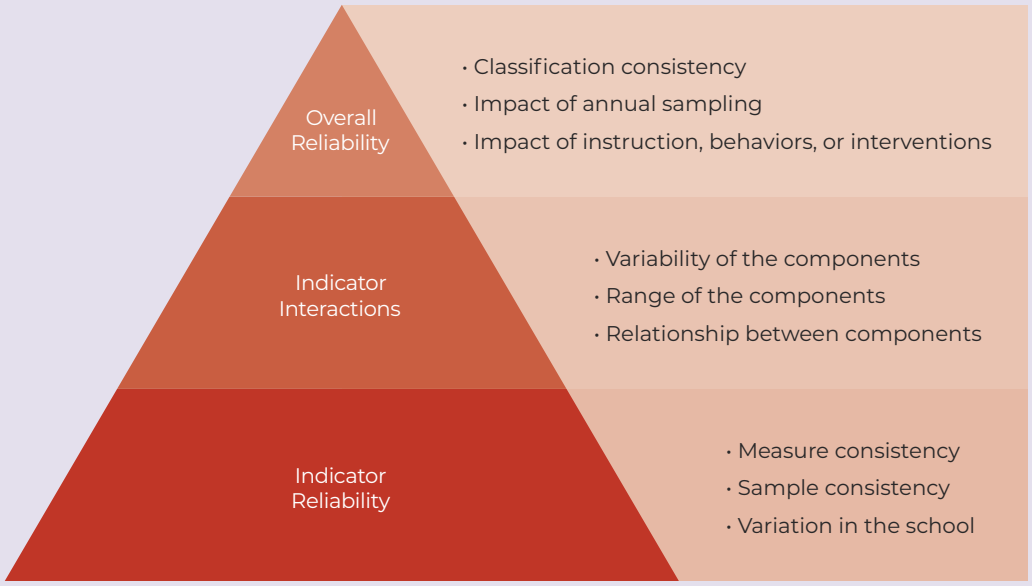
In addition to general questions related to how a growth indicator relates to the overall theory and whether the growth indicator is a fair representation of progress for students with SCD, states also are encouraged to examine the technical quality and reliability of the indicators. Chapter 4 raised several issues to consider when evaluating the reliability of growth measures based on the AA-AAAS. It is important to evaluate the alternate assessment scores and the growth measures upon which it is based; however, the following section focuses specifically on how indicators function in the accountability system.

How should reliability be conceptualized in accountability?

Reliability in accountability systems (which include growth indicators) is not as simple as quantifying consistency in scores over time. Evaluating the reliability of indicator scores and their resulting school designations begins with understanding the impact of measurement and sampling issues on school-level estimates of system indicators, which affect the variability of the indicators and overall accountability-system decisions. Therefore, there are three primary areas of reliability to which we must attend: the reliability of indicators; the interactions among indicators; and the resulting reliability of school scores and determinations. (See D'Brot, Lyons and Landl, 2017.) It is also important to note that these three sources of variability are heavily influenced by measurement and sampling error. (See Hill & Depascale, 2005.)

Examining all three components — indicator/component reliability, interactions among indicators and reliability of school determinations — is a critical step to understanding the utility and impact of an accountability system. The following figure presents how these facets of reliability interact, which will affect whether a growth indicator might have an impact:[8]

FIGURE 5-3: FACTORS INFLUENCING TECHNICAL QUALITY IN ACCOUNTABILITY (D'Brot, Lyons, & Landl, 2017)



The remainder of this section focuses on the technical quality (with a focus on reliability) of indicators and indicator interactions.

[8] Note that an in-depth discussion of evaluating the reliability and technical characteristics of overall scores, which is available in D'Brot, Lyons and Landl, 2017, is beyond the scope of this chapter.

Does the growth indicator demonstrate sufficient internal technical quality for growth measures based on the AA-AAAS?

A common concern for growth indicators in accountability is to first determine whether the measures that comprise the growth indicator demonstrate sufficient internal consistency and reliability. While peer review and assessment development practices based on the Standards (2014) dictate best practice for the technical characteristics of the alternate assessment, it is imperative to examine the reliability of the indicator and subsequently, the reliability of the system of AMD. This perhaps can best be contrasted by considering (a) the definition of measurement error for an assessment (i.e., the representation of the difference between an observed score and a true score; see Brennan, 2006) and (b) the consistency of scores based on aggregations of the assessment scores, with a particular focus on growth.

When considering measurement error, the first priority should be on determining whether the measures that compose the growth indicator demonstrate sufficient internal consistency and reliability; this will likely be a test-based measure that relies on the evidence collected from peer review (see U.S. Department of Education, 2018). Assuming there is sufficient reliability in the scores from the AA-AAAS, one then can understand the idea of consistency in scores on the growth indicator. For a growth indicator, there are at least two areas of focus: (a) the consistency of the scores and (b) the variability in the size and characteristics of the population.

These concepts are regularly used when evaluating assessment data, due to the inclusion of non-assessment data (e.g., climate surveys, attendance data, grades in college, etc.), which are typically part of a state's school quality and student success (SQSS) indicator(s). While not a focus of this chapter, states that opt to combine academic growth and other data into their SQSS indicator should consider additional criteria for non-test based measures, because of:

- different requirements for comparability;
- the potential misalignment with the intended use of the data and its application in accountability;
- resource constraints to collect data;
- fewer administration requirements related to standardization; and
- increasing its susceptibility to sources of error.

For considerations on how to deal with these challenges, please see Marion and Lyons' (2016) examination of data quality criteria (i.e., unit of analysis, level of inference, potential corruptibility and level of data burden).

While common sources of measurement error typically are examined through peer review (e.g., inconsistency of the content and standards assessed, time of test administration, grade spans, equating considerations, differential motivation, administration factors, scoring changes, etc.), population characteristics can affect growth indicator technical quality. For example:

- What is the impact of the population taking the AA-AAAS?
- How well is the population identified?
- How inclusive are identification policies across the state?
- How might accommodation types and disability categories affect comparisons among students with SCDs?
- What other types of student characteristics are represented in the populations?

Excessive amounts of volatility in the school sample or very low variability in growth measures from the AA-AAAS may have an impact in how indicators interact. Once states have collected and examined the evidence supporting the reliability of the measures comprising the growth indicator, they can begin to consider how the technical quality of the indicator might affect interactions among indicators.

Can the measures that make up the growth indicator (i.e., the AA-AAAS and the general assessment) be compared and combined appropriately? How can we understand the technical quality of the indicator when combined?

To interpret the reliability of school accountability results, one must ascertain how and to what degree indicators demonstrate expected characteristics and relationships. Before examining how an overall indicator might influence school determinations, it is important to understand the interaction between growth indicators based on alternate and general assessments.

One key consideration for states that are combining growth indicators across alternate and general assessments is to understand the extent to which growth measures can be compared and combined as a single indicator. A growth indicator overall will be expected to contribute to a state's overall system of meaningful differentiation, so it is key to understand the interplay between the two growth indicators. There likely will be fewer issues if the alternate and general assessment growth models are the same (e.g., both are gain scores or both are value tables). Measures produced from different growth models will need to be compared for their similarity in range, shape and variability; otherwise, the differences in variability from one assessment (e.g., general assessment) can overshadow the amount of variability that is contributed by the other assessment — and this will affect how growth is represented in the system of AMD.

For example, if a continuous growth scale (e.g., student growth percentiles of 1-99) is combined with a categorical growth model (e.g., four performance categories are split into eight sub-categories) without some transformation, the larger scale with increased variability will be over-represented in the system of AMD. This can be compounded by comparing the differences between theoretical ranges and actual ranges of certain scales. Continuing with this example, SGPs may represent percentiles of 1-99 at the student level, but they could cluster in a more restricted range when aggregated in practice.

When combining measures to create an indicator, outliers also can affect interpretations of measures by influencing standardization techniques, comparisons of range or comparisons of averages. However, outliers also can reflect very high-performing or underperforming schools that should be recognized in system development. Due to the types of measures that are typically used for the growth indicator, it is essential to understand how trends in performance may affect the interpretation of the indicator and how it affects the overall differentiation of schools. If data are standardized in any way (e.g., indexes based on scale scores or thetas, etc.), one will need to determine whether there are any restrictions in available data points, or whether strong modal or multi-modal characteristics emerge.

If combined data are reflecting data characteristics that are difficult to interpret, states might consider adjusting their methods to make data more interpretable or usable. Note: standardizing or transforming data that is typically interpreted as student growth can make interpretations more challenging, especially for the public. Excessive transformations can make data less actionable and should be considered in conjunction with reporting design.

How does the growth indicator contribute to the state's system of annual meaningful differentiation?

As noted previously, reliability and error are common concepts that measurement professionals and state education agency (SEA) officials consider when examining assessment results from the alternate assessment. While consistency is a central consideration, it is perhaps more appropriate to consider the risk of misclassifying a school when evaluating differentiation (Hill & DePascale, 2003). It may be necessary to examine the consistency in classification in conjunction with evidence of school capacity, behavior, policy and practice to corroborate differentiation. That is, is consistency in school classification a function of consistency or unchanged behaviors? Is the consistency in school classification a product of a consistent sample of students? Or is the school's consistency due to difficulty in making progress on earning points or changing classifications on indicators and the state's system of AMD?

It is likely that states will examine data-related changes to performance to understand how a growth indicator is functioning. However, capacity or behavioral change can have a longer-term impact on changes to growth indicators. While the behaviors associated with a school designation are not typically elements of the reliability of a school's rating, it is valuable to understand both the behaviors that are expected as a function of potential identification (e.g., identifying a school as Additional Targeted Support and Improvement for students with disabilities) and the behaviors that may be an early indication of why school performance is what it is.

Furthermore, there may be opportunities to understand how improvements in supporting students with SCD may not be reflected in changes to the overall differentiation. If, for example, there is a particularly small set of students who take the alternate assessment, improvements in their instructional experience may not be reflected at the school level. However, if a school's Additional Targeted Support and Improvement (A-TSI) identification is in part due to the performance of students with disabilities, identifying promising practices supporting students with SCD should be applauded and sustained to improve student group performance.

States should be cognizant of excessive volatility in school differentiation (with a particular focus on the volatility in the growth indicator), but some variation is expected in school scores from year to year. Furthermore, year-to-year variation and within-year variability should be expected in growth indicators more than they are in other academic indicators in a system of AMD, because observed changes on an assessment using growth rates will be detectable earlier than they are with proficiency rates.

However, excessive variation can be a result of too much error in the indicator; this can be attributed to both measurement error (i.e., the error associated with the assessments or tools comprising the growth indicator) and sampling error (i.e., the error associated with a different group of students being tested or used for accountability analyses each year; Shavelson & Webb, 1981). In addition to variation based on the measurement tool, there is the natural variance that may emerge due to changes in a school's personnel and/or instructional behaviors.

Sampling error may be the most relevant issue to consider because of the ways in which school composition interacts with school-based-intervention effects on students. Since school- and student-level interventions and instruction are typically idiosyncratic and vary from year to year, it may be difficult to isolate how and by what degree they affect school outcomes. It can be informative to differentiate whether changes in performance are a function of the change in the samples of students or teachers or are actually a product of the instruction and interventions that students with SCD are receiving. Additionally, there may be a greater need to understand how students are being supported within a school to confirm assumptions about changes in growth indicator scores, if the supports for students with SCD are individualized and/or there is instructional consistency (i.e., the same teacher) across grades.

Evaluating whether indicators support meaningful differentiation involves more than determining how much measures contribute to the results of a given indicator. It is also informative to understand whether changes in the indicators lead to changes in overall school performance. For example, progress on a transition table may change somewhat slowly, as compared to scale score-based changes — a phenomenon which may be amplified if students who take the AA-AAAS remain in the same category over time (Tindal et al., 2014). Extending this comparison to value tables and norm-referenced growth, progress on a transition table may have jumps and pauses, due to the nature of student progress in terms of scale scores on an assessment. By contrast, indicators based on norm-referenced growth approaches (e.g., student growth percentiles or value-added modeling) may reflect more restricted ranges in aggregates than those based on percent proficiency, sometimes making it difficult to influence overall school differentiation.

Conducting these types of analyses can help states determine whether the growth indicator is influencing the state's system of AMD as intended. For example, certain student-growth approaches are much less influential than proficiency because of their restricted aggregate ranges (despite their increased volatility year-over-year at the student level); such approaches may require adjustments to policy or empirical weights in the state's system of AMD. It is important to explore whether or not these restrictions are more pronounced due to the characteristics of the growth indicator based on the AA-AAAS.

Additionally, it is critical to recognize the relative sample size of students taking the alternate assessment, as compared to students who take the general assessment. It will require states to make a policy decision, followed by an empirical examination of simulations, regarding how best to aggregate the alternate assessment and general assessment growth measures. For example, if a weighted average is used, the relative proportion of students with SCD (i.e., by definition limited to 1% statewide) will have a minimal impact on the overall growth indicator. This elevates the policy objective or goal, as well as how states may wish to communicate growth in students who take the alternate assessment.

To obtain a single growth indicator, common growth measures across alternate and general assessments could be aggregated before transformation; different growth approaches may require a transformation to make indicator attributions equivalent or to make the alternate portion more influential. It will be important to base this decision on both policy goals and empirical evidence.

QUESTIONS TO CONSIDER WHEN ADDRESSING THE USE AND IMPACT OF THE GROWTH INDICATOR BASED ON THE AA-AAAS

The design and technical characteristics of a growth indicator are key foundations to examine as part of a strong design, development and implementation plan. However, a significant portion of implementation is determining whether growth based on an alternate assessment has an impact on student and educator behavior. The remainder of this section raises questions related to the use and impact of the growth indicator based on the AA-AAAS.

Given that they represent important signals of progress for students with SCDs, how useful are data from the growth indicator to consumers of the state's accountability system?

When designing accountability systems that include growth as a key indicator, it is wise to revisit the theory of action to understand why those data are perceived as informative and useful. Accountability indicators and reporting are often at risk of being relegated to information that serves a punitive or labeling purpose. Consumers of the state's accountability system should be supported in understanding that data reflect valuable markers of school quality and can be used to set future performance targets linked to more real-time data.

Although academic growth indicators often are referenced as more sensitive data points, school principals and classroom educators may feel that they are insufficiently sensitive to reflect actual changes to instructional practice or curriculum implementation. Depending on the growth measure, a growth indicator may function as a more sensitive measure of school progress. However, its interpretation should not be over-extended; it is useful to communicate growth from large-scale assessments as a reflection of a point-in-time measure that is intended to better highlight progress toward grade-level expectations than achievement alone.

Interpreting an accountability indicator and determining its perceived value often can be a valuable exercise when planning well-structured resources and a thoughtful communication strategy. States may want to determine whether there is enough framing around the importance, role and intended uses of growth data based on the AA-AAAS. If there are guiding principles around why the growth indicator was designed and how it reflects growth expectations for students with SCDs, states should ensure that these principles are readily available to the public and included in communication strategies.

Additionally, if there are expectations regarding how to use the growth indicator for planning or how to use data to inform progress toward student goals, states should also integrate that information into their communications plans. Communications also might distinguish between data derived from the aggregated growth indicator and disaggregated growth data about specific students. States likely can remain high-level when describing the overall growth indicator, but it will be important to connect student-focused or subgroup-focused growth data to other resources (e.g., improvement plans, data analysis templates, planning documents) supporting contextually appropriate teaching and learning.

Growth has been of interest in state accountability since the growth model pilot (Hoffer, et al., 2011), and it has increased in visibility with ESEA flexibility and the passage of ESSA. Today, growth measures have received a large degree of attention as a high-stakes outcome in accountability, because they condition outcomes on baseline performance. However, it is important to frame how data from the growth indicator reflects and supports the high-level objectives of the educational system and the state's accountability system's theory of action.

States should consider supplementing or highlighting communications and documentation and/or increasing access to resources that describe how the growth indicator reflects a reconceptualization of point-in-time performance through information that is intended to confirm what educators already know about students at the end of the year. Instead of relying on growth as an informational tool for any particular student, it may be more appropriate to frame it as an information tool for system progress against standards or to help evaluate the implementation of instructional approaches at the school level. However, it does provide states with the opportunity to refocus conversations on individual student growth as a function of it being a high-stakes measure in the larger system.

Are primary users able to interpret reported data based on the growth indicator using scores from the AA-AAAS?

In order for primary users (e.g., educators and family members) to interpret data based on the growth indicator, the presentation of growth data from the AA-AAAS must be clear and without excessive jargon. Educators and instructional leaders will be the primary consumers of student growth data, and they must have a clear understanding of how it relates to other system indicators and performance results and how growth reflects state or local priorities.

The state can engage in certain analyses to identify opportunities or connections that can be the basis of additional communications or resource plans. For example, the state can examine results from the growth indicator across the state and over time to determine which cut scores, thresholds or flags within the indicator reflect meaningful growth or performance for students with SCDs. The state then can identify how growth on the alternate assessment compares to proficiency cuts, growth observations or expected growth targets, where relevant. These connections can be communicated to the public and primary consumers of accountability data.

For example, many administrators may be particularly interested in whether students are meeting proficiency targets, in order to meet some measure of interim progress (as per ESSA) or threshold for exiting school-identification status. Highlighting how growth is an early indicator of growth-toward-proficiency can help schools determine whether certain instructional strategies or supports have more or less impact on students with SCD; it also helps maintain student and educator enthusiasm about improvements in student learning.

States also can identify potential stakeholder groups that would be the primary users of this data. One can consider how each group might interact with growth-indicator data, then determine whether sufficient resources are in place to support that group's interpretation and use of the data. For example, district leaders may need a blend of both state accountability and related school data, whereas school leaders may need reports that highlight school-level data with explanations regarding the data's relevance for improving accountability-based data. For each of the identified stakeholder groups and users of the state's accountability system, there may be common or unique data displays, visualizations or reporting strategies that can be used to examine, interpret and act upon indicator data.

While developing meaningful reports and resources is an iterative and resource-intensive process, it can lead to changing conversations about student learning. By including a growth indicator based on the AA-AAAS, users may think differently about grade-level student growth expectations. Depending on the key stakeholders identified, states might consider reframing, repackaging or creating additional resources that are targeted to high-leverage groups that use accountability data (e.g., district leaders, principals or school-improvement teams) that connect the state summative assessment (if applicable) to other localized sample data (e.g., course grades, engagement measures, progress monitoring data, etc.) with a focus on students with SCD.

These communications resources may be especially valuable for states that combine general and alternate assessment growth measures into the same growth indicator. In this case, it may be more useful to focus on the reason why growth is relevant for each of these populations and how growth models may better reflect the growth expectations of the standards, instead of focusing on the technical differences between the two.

"Educators and instructional leaders will be the primary consumers of student growth data, and they must have a clear understanding of how it relates to other system indicators and performance results and how growth reflects state or local priorities."

Do primary users understand how data from the growth indicator are connected to other coherent and meaningful data to inform continuous improvement practices?

A risk associated with accountability data is that it can be perceived as post-mortem and irrelevant to teaching and learning. While this can sometimes be remedied by using growth data, these measures still can be seen simply as multiple, point-in-time data strung together. States can combat this perception by explaining how growth data are linked with coherent lagging and leading indicators that support continuous improvement. However, this focus on continuous improvement must be viewed in the context of a classroom, school or district's instructional design.

For example, there will be differences in helping educators understand cross-year growth data as a corroboration of within-year instructional growth, depending on whether inclusive models or self-contained models of instruction are the dominant delivery method. States and districts may partner with schools or teachers to help them understand how certain instructional strategies benefit different groups of students with SCD, and how that is reflected in growth indicator performance.

It can pay dividends to help practitioners and educators understand the links among instructional decisions, professional development selection, program implementation, progress indicators and outcome indicators (both for and beyond high-stakes accountability) and how high-stakes assessment data are a reflection or confirmation of what they already know. In other words, states should consider which resources or support structures will help users of state accountability data identify not only what to improve, but also how they might drive improvement using other local academic performance and growth data for students with SCD.

Doing so will require states to identify related outcome data or lagging indicators that can inform growth indicator improvement(s) that could be made available statewide, regionally and/or at the district and school levels. If states are able to identify leading indicators and process data that are linked to improvements on the statewide summative assessment, it can give educators and administrators a stronger measurement infrastructure for tracking student improvements and gains in performance.

These coherent data should be based on evidence or research (e.g., indicators of engagement, indicators of classroom success, rates of attendance or other relevant measures for students with SCDs) and should be compared to the theory of action of the accountability system or something more global (e.g., the state's strategic plan). It also will be informative to test these data links with practitioners for validity and to differentiate the connections by grade span, content area and outcome type, if relevant (e.g., scores, performance levels, standards, learning targets, etc.).

What steps should states take to create coherent links between accountability (i.e., outcome) data and data that are more process- or condition-oriented?

A state's theory of action should provide a rationale for each indicator selected for inclusion in the system. This rationale should clarify the actions and initiatives the state believes will be incentivized by including the indicator and how those actions will serve to improve performance. States should articulate the range of actions it believes will lead to improved academic achievement based on any statewide, systematic, regional or districtwide initiatives or interventions (e.g., endorsed, evidence-based strategies; tiered intervention approaches; curriculum and instruction initiatives).

Connecting evidence across time (i.e., process, leading and lagging data) can be challenging if users of the state's accountability system are not aware of all the available data. It will be valuable to help users understand the ways accountability and related outcome data are connected to process-oriented data that may be associated with program decisions at the school or district level (e.g., intervention selection, curricular programs, progress monitoring or use of district-developed assessment to refine classroom assessment practices). In other words, it is important to connect accountability not just to outcomes/test data but also to processes of instruction and daily progress monitoring.

In conjunction with building state- and district-based school improvement teams, it will be helpful for states to embed analyses of coherent data connections — differentiated by grade span and data type (e.g., academic achievement data, access to opportunity-to-learn data, student and teacher engagement data, course-related data) — into school-improvement planning processes, data reviews and/or reviews of district- or school-led school improvement and support plans. Doing so can help systematize conversations among school personnel that elevate growth-based expectations for students with SCD — i.e., by, basing discussion on a system, versus a one-time score.

SUMMARY

Prior chapters of this handbook explained how a strong theory of action is necessary to guide the design, development, implementation and evaluation of a growth model based on the alternate assessment. Chapter 5 has built on prior chapters to illustrate how incremental decisions and component reliability must be compiled over time to design an AA-AAAS-based growth indicator that demonstrates both policy goal alignment and technical quality. This process is reflected in the three-part analysis of accountability indicators' reliability:

1. Is the growth indicator based on the AA-AAAS itself reliable?
2. How do the general-assessment and alternate-assessment growth measures or indicators (depending on the design) interact as an overall growth indicator?
3. How does the growth indicator influence the overall reliability of school determinations?

In order to explain these concepts of reliability, this chapter presented a series of considerations that states can use to guide the design, selection or evaluation of their growth approach based on the AA-AAAS. After describing ESEA requirements for growth and reviewing the distinctions among status, improvement and growth, this chapter presented a series of questions aligned to the sequential design, development and implementation stages of accountability systems. (See D'Brot, in press.) These included questions related to the policy goals of the growth indicator; the technical quality of the growth measures and indicator; and the impact and use of the growth indicator.

This chapter also showed that (a) the selection and application of growth measures has direct implications for the design and implementation of growth indicators and (b) growth based on the alternate assessment must be compared to growth based on the general assessment simultaneously.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. https://www.apa.org/science/programs/testing/standards

Brennan, R. L. (2006). Perspectives on the Evolution and Future of Educational Measurement. In. R. L. Brennan (Ed.), *Educational Measurement* (pp. 1-16). Praeger Series on Higher Education. https://www.worldcat.org/title/educational-measurement/oclc/69104278

D'Brot, J. (2017). *Considerations for Including Growth in ESSA State Accountability Systems*. The Council of Chief State School Officers. https://ccsso.org/sites/default/files/2017-10/CCSSOGrowthInESSAAccountabilitySystems1242017.pdf

D'Brot, J., Lyons, S., & Landl, E. (2017). *State Systems of Identification and Support Under ESSA: Evaluating Identification Methods and Results in an Accountability System*. The Council of Chief State School Officers. https://ccsso.org/resource-library/state-systems-identification-and-support-under-essa-evaluating-identification

D'Brot, J., LeFloch, K., English, D., Jacques, C. (2020). *State Support Network: Evaluating State Accountability Systems*. American Institutes for Research. https://oese.ed.gov/resources/oese-technical-assistance-centers/state-support-network/resources/evaluating-state-accountability-systems-esea/

Hill, R. & DePascale, C. A. (2003). *Adequate Yearly Progress Under NCLB: Reliability Considerations*. Paper presented at the 2003 Annual Meeting of the National Council on Measurement in Education.

Hill, R. & DePascale, C. A. (2003). Reliability of No Child Left Behind Accountability Designs. *Educational Measurement Issues and Practice, 22*(3), pp. 12-20. https://www.researchgate.net/publication/227662232_Reliability_of_No_Child_Left_Behind_Accountability_Designs

Hoffer, T. B., Hedberg, E.C., Brown, K. L., Halverson, M. L., Reid-Brossard, P., Ho, A. D., & Furgol, K. (2011). *Final Report on the Evaluation of the Growth Model Pilot Project*. U.S. Department of Education https://www2.ed.gov/rschstat/eval/disadv/growth-model-pilot/gmpp-final.pdf

Marion, S., Lyons, S. (2016). *In Search of Unicorns: Considering the "5th Indicator" in ESSA Accountability*. National Center for the Improvement of Education. https://www.nciea.org/sites/default/files/pubs-tmp/Marion%20Lyons_ESSA%20Accountability_5th%20Indicator_111416.pdf

Shavelson, R. J. & Webb, N. M. (1981). Generalizability theory: 1973-1980. *British Journal of Mathematical and Statistical Psychology*, 34(2), 133–166. https://bpspsychub.onlinelibrary.wiley.com/loi/20448317/year/1981

Tindal, G., Nese, J. F. T., Farley, D., & Saven, J. (2014, April). *Documenting Growth for Students with Significant Cognitive Disabilities*. National Council on Measurement in Education (NCME). https://www.ncaase.com/docs/TindalNeseFarleySavenNCME_AltAssessV3.pdf

U.S. Department of Education. (2018). *A State's Guide to the U.S. Department of Education's Assessment Peer Review Process*. https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf

CHAPTER

6

RECOMMENDED PRACTICES FOR MAKING DECISIONS ABOUT ALTERNATE ASSESSMENT GROWTH MODELS, MEASURES & INDICATORS

RECOMMENDED PRACTICES FOR MAKING DECISIONS ABOUT ALTERNATE ASSESSMENT GROWTH MODELS, MEASURES & INDICATORS

INTRODUCTION

Previous chapters in this handbook provided context for the issues involved in making decisions about alternate assessment growth (Chapter 1) and offered a non-technical introduction to growth models, measures and indicators (Chapter 2). Chapter 3 provided an overview of theories of action to set the stage for decisions about growth measures (Chapter 4) and growth indicators (Chapter 5). Chapter 6 integrates the concepts and key considerations from Chapters 2-5 to describe a process for including alternate assessment growth measures in growth indicators for state accountability systems. This chapter also describes several potential extensions and uses of growth measures in other contexts, including the State Systemic Improvement Plans (SSIP) required under the Individuals with Disabilities Education Act (IDEA) and continuous program improvement at state and local levels.

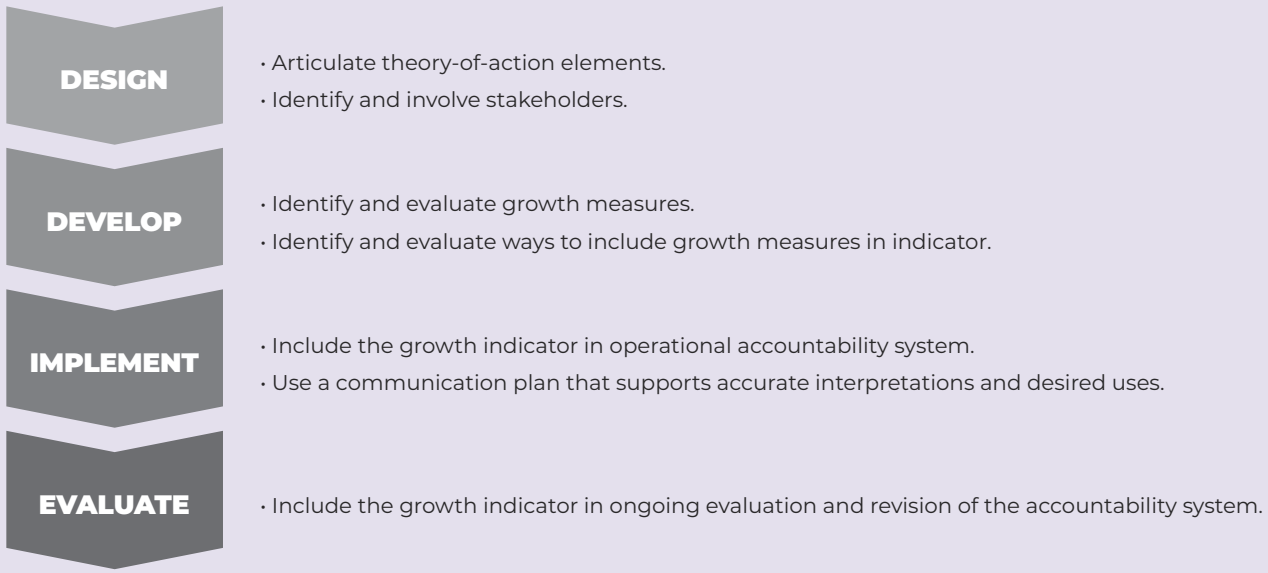
Regardless of the context, decisions about using alternate assessment growth measures should be grounded in the intended purposes and uses of the results and the claims a state wishes to make. Since inclusion in Every Student Succeeds Act (ESSA) accountability is a primary driver for states considering this issue, states should start by defining the intended claims they wish to make about the growth measure and the growth indicator. States also are encouraged to consider whether incorporating alternate assessment growth into ESSA or IDEA accountability plans or using alternate assessment growth data for other purposes will promote equity in educational systems and outcomes for students with the most significant cognitive disabilities (SCD).

A PROCESS FOR INCLUDING ALTERNATE ASSESSMENT GROWTH IN ACCOUNTABILITY GROWTH INDICATORS

This section outlines a process that states can use to evaluate options and make decisions for including alternate assessment growth in accountability systems. The process supports states as they attend to policy goals and mechanisms to achieve those goals; the technical quality of the design; and the ways in which the design of the growth indicator incentivizes behaviors. (See Chapter 4 for additional discussion of these considerations.)

The recommended process is grounded in the framework described in Operational Best Practices for Accountability (D'Brot, 2020). Recommendations align with a subset of these best practices; correspondence to specific best practices are noted in each section. The operational best practices span all phases of accountability system work — including design, development, implementation, evaluation and revision (see Figure 6-1).

FIGURE 6-1: STEPS TO INCLUDE ALTERNATE ASSESSMENT IN A GROWTH INDICATOR ACROSS THE ACCOUNTABILITY SYSTEM LIFE CYCLE



The question of whether and how to include alternate assessment growth measures in an accountability growth indicator is one small segment of an accountability system. The following section is based on an assumption that states already have established accountability plans under ESSA and that adding alternate assessment growth is a modification of the existing plan. States developing new plans are encouraged to explicitly consider students with SCD at all stages of the process.

For example, the state should ensure that students with SCD are included in their accountability system's theory of action (Operational Best Practice 1.4). States also may wish to confirm that the measures used for non-achievement indicators (e.g., ESSA's School Quality or Student Success indicator) include data related to the educational experiences of students with SCD.

DESIGN PHASE

Step 1: Ensure the accountability system theory of action elements are clearly defined and applicable for students who take alternate assessments. *[Operational best practices 1.2 – 1.4]*

As noted in Chapter 3, a state may have a variety of goals for including alternate assessment growth in their growth indicator. Do the ultimate goals correspond with those established for students taking general assessments? If the goal for both groups of students is to make sure students are ready to pursue post-secondary opportunities after graduation, there may be more parallels between general and alternate assessment which can be used to support subsequent decisions when selecting a growth model and defining the growth measure. If a goal for alternate assessment growth is to promote student progress in the general curriculum, the state may define different outcome measures or targets for alternate assessment growth.

The state also may need to consider whether the potential mediating outcomes are defined in ways that include students who take alternate assessments. Are there programs or units not previously identified in the state's theory of action that are responsible for promoting academic growth for students with SCD? For example, if some of the resources being used to support academic instruction flow through a special education office, it will be important to capture that information to support later adjustments in response to annual accountability results. If the theory of action for the state's existing accountability system does not include language describing goals, outcomes or assumptions about the role of alternate assessments in supporting students with SCD, the state should modify its theory of action as needed before taking additional steps toward developing alternate assessment growth measures or indicators. Note that any modifications should be coherent extensions of the existing theory of action.

Step 2: Identify and Involve Stakeholders. *[Operational best practices 2.1 – 2.5]*

Many state systems do not have an explicit focus on student groups beyond those identified for Additional Targeted Support and Intervention and Targeted Support and Intervention purposes; it is important for the state education agency (SEA) to examine its system design to be sure it is adequately representative of diverse student groups. The best way to build broad representation is to engage all the key players with an investment in promoting positive outcomes for students with SCD.

A state with an existing accountability plan may have already identified and engaged such stakeholders. For example, a staff member from the state's special education office serves on a committee that reviews accountability results each year or may have been consulted regarding other indicators; representative(s) of a parents' or disability-rights advocacy group helped to create the accountability theory of action; etc.

It is also important for states to include stakeholders with expertise in various aspects of the alternate academic achievement standards (AA-AAAS) when they are considering whether and how to incorporate alternate assessment growth in a growth indicator. Relevant areas of expertise might include:

- assessment design and technical adequacy
- academic expectations for students with SCD across the K-12 grade span
- teacher preparation
- knowledge of educational programs, including related services that may impact academic progress

Stakeholders can play different roles in the design and use of the alternate assessment growth measure in the growth indicator. For example, those with more technical expertise could collaborate closely during the design and development phase, while those with population expertise might serve in an advisory capacity during the design and evaluation phases. A state's accountability system Technical Advisory Committee should include one or more members with expertise on alternate assessment.

Also, if a state has contracted its accountability-related services, the contracting organization should have enough expertise to implement procedures for including alternate assessment growth once the measure and indicator are defined. Finally, a state should make sure that its communication plan and materials include information about how alternate assessment growth is included in the growth indicator and how the state has defined the goals and intended outcomes for academic growth among students with SCD.

DEVELOPMENT PHASE

As noted in Chapter 1, although including alternate assessment growth in an accountability growth indicator may be a desired characteristic of an effective accountability system, ESSA does not require states to have an alternate growth indicator. The development phase begins with evaluating the suitability of the alternate assessment score as the basis of a growth measure (see Chapter 4). The score must be supported by sufficient evidence of validity and technical quality for the alternate assessment system, including but not limited to the types of evidence produced for U.S. Department of Education assessment system peer review. Depending on the growth models under consideration, AA-AAAS scores may need to possess other characteristics (e.g., vertical scaling; see Chapter 2 and Chapter 4). Suitable alternate assessment scores are a prerequisite for producing and reporting a reliable and valid growth measure.

At this point, the development process moves through two phases: developing growth measures (i.e., choosing and applying a growth model) and including alternate assessment growth measures in the accountability growth indicator. Distinguishing between these phases (also seen in Chapter 4 and Chapter 5) supports states in taking a principled approach to both. That is, a state would identify a defensible growth model for its alternate assessment as a prerequisite to including an alternate assessment growth measure in its growth indicator. Note that states may have other uses for the growth measure itself. (See *Extensions* section, p. 96.)

Step 1: Identify and evaluate potential AA-AAAS growth measures. *[Operational best practices 4.5, 4.6]*

States start this step by engaging stakeholders to explore ways of operationally defining the alternate assessment growth measure, given the desired interpretations and uses of the growth measure. (The state will not necessarily use the same growth model for alternate assessment as for general assessment.)

States need to (a) articulate what growth means for students with SCD and (b) decide which model best supports that priority. By articulating content-related growth expectations for students with SCDs, states can better conceptualize how to measure such growth. It also supports decisions regarding whether to measure these students' growth based on their progress toward a standard (i.e., criterion-referenced growth) or their progress against peers or previous performance (e.g., norm-referenced growth). This conceptual approach elevates policy considerations and desired interpretations, so that statistical considerations are no longer the primary driver.

Note that growth measures that include a criterion-referenced component will likely be more highly correlated to achievement than those which are predominantly norm-referenced. Although norm-referenced growth measures tend to increase the availability of information for school quality-related measures (as required under ESSA), the benefits of using a growth model — in terms of achieving both policy and practical goals — may outweigh the advantages of using a norm-referenced model. Indeed, the growth model can make a unique contribution to the overall differentiation of a state's schools.

States also may need to consider the likely interpretation of growth-measure findings by different stakeholders. For instance, are the basic interpretations of one type of growth model (see Chapter 2) more or less intuitive for educators and parents than another type of model? If the measure will be used for other purposes (i.e., beyond federal or state accountability reporting), does a single alternate assessment growth measure support the multiple, intended uses? What are the potential misinterpretations and/or misuses of growth-measure based data, and how could the state mitigate any potential consequences?

The issue of operational definitions raises the question of which measures are defensible, including considerations of reliability, validity, fairness and feasibility. Options for alternate assessment growth models may be constrained by (a) the psychometric model being used to score and/or scale the assessment or (b) the size and stability of the population from year to year. (See Chapter 2.) Note that to be defensible, a growth measure must be supported by evidence regarding vertical articulation of the underlying academic construct. The state may already collect this type of evidence for peer review purposes.

Once a state has identified potential growth models, based on how well they fit goals and intended uses and how likely they are to produce a defensible growth measure, the next step is to apply the prospective models and evaluate whether the resulting growth measures produce the intended information. If there are multiple, potential growth models, the state might evaluate the absolute quality of information produced by each model and the relative quality of the measures produced by both models. For example, if each prospective growth model produces equivalently-precise growth measures, is one set of measures more interpretable than the other?

Step 2: Identify and evaluate potential ways to include the alternate assessment growth measure in the state's growth indicator for accountability. *[Operational best practices 4.1.b, 4.3, 4.5, 4.6]*

After selecting an alternate assessment growth model and producing a growth measure, it is time to consider the possibilities and constraints for incorporating the growth measure into the accountability growth indicator. These considerations would be similar to those that states use for other indicators. It will be important to involve stakeholder groups in considering the possible interpretations, uses and outcomes of incorporating the alternate assessment growth measure into the growth indicator. For example:

1. What are the implications for students with disabilities sub-group reporting, when students who take alternate assessments are added to the available data for a school?
2. How would stakeholders interpret and act on growth indicator results when alternate assessment growth is included or excluded?
3. How does including alternate assessment growth in a growth indicator contribute to the meaningful differentiation of schools?
4. Would an alternate assessment growth measure have a differential impact on school identification, based on the proportion of students in a school taking the alternate assessment? If so, how might this outcome impact interpretations about the growth indicator?
5. What is the impact of a student's educational placement (e.g., neighborhood school vs separate school vs. homebound) on school results?
6. What are the potential, unintended consequences of including alternate assessment growth in the growth indicator? What are the potential misinterpretations and/or misuses of the growth indicator, and how could the state mitigate those?

If there are multiple, potential ways to combine general assessment and alternate assessment growth into the growth indicator, it will be important for states to pilot each method with stakeholders and evaluate the results, using questions such as those listed above for each of the options.

From a fairness perspective, when a state is calculating an overall growth measure, each student's growth should count equally and have appropriate influence on a school's growth indicator, regardless of whether the student took the general or alternate assessment. The state should be able to provide a rationale for choosing a given method of including alternate assessment growth in the growth indicator; the state also should be able to confirm that the original rationale for including an academic growth indicator still holds once alternate assessment growth is included.

As with other measures, if states want to emphasize the need for accountability systems to inform services for students with disabilities, states may choose to include disaggregated growth on the alternate assessment (or growth for students with disabilities overall) as a component within the growth indicator.

IMPLEMENTATION PHASE

Step 1: Ensure that alternate assessment growth is appropriately accounted for in the state's operational accountability system. *[Operational best practices 6.1-6.4]*

This step includes ensuring that the state's growth indicator includes business rules, procedures, timelines, documentation and data-integrity methods that apply to both (a) data from the general and alternate assessments and (b) the process of combining the two sources into a single indicator.

Step 2: Address the inclusive growth indicator through communication strategies to support public and educator use of the data as intended. *[Operational best practices 7.1, 7.4, 7.5]*

Publicly reported accountability results should include messages that support the intended interpretations. For academic achievement and growth, messages may need to proactively address common misconceptions about academic outcomes for the population of students with SCD. Educators may need additional information to support decision-making for the purpose of improving educational programs that serve students with SCDs, such as by improving teacher capacity to teach the extended content standards.

States also may wish to work with stakeholders who helped design the alternate assessment growth measures, to engage their support in planning communications; doing so will help to ensure that communication strategies are implemented appropriately and will be effective in promoting accurate understanding and interpretation of the findings.

EVALUATION AND REVISION PHASES

Step 1: Include the updated growth indicator in the cycle of monitoring, evaluating and revising the accountability system as needed. *[Operational best practices 8.1 - 8.5]*

Once alternate assessment growth measures are incorporated into accountability growth indicators for operational use, many of the earlier processes can be reapplied as states engage in monitoring and evaluating their accountability system. If earlier decisions on alternate assessment growth measures or the growth indicator were based on preliminary data, do operational accountability results across years support those earlier predictions? Operational data indicate a need to define growth differently, choose a different growth model or set different targets.

As with other indicators, it is important to evaluate operational results in the context of any disruptions that occurred during the time period under evaluation — for example, due to policy changes, disruptions to the trend data (e.g., those caused during the pandemic) or shifts in the underlying population. States may wish to evaluate and interpret results for the alternate assessment growth measure before evaluating the accountability growth indicator. It is also valuable to be sure that alternate assessment growth is included in all monitoring and continuous improvement. Results also could lead to changes in the theory of action.

Two appendices at the end of this chapter provide additional information on including alternate assessment growth in growth indicators.

- Appendix A provides detailed flowcharts to support states in making decisions at each stage of the process. Steps within the flowcharts are anchored in the guiding policy and technical questions from Chapter 4.
- Appendix B contains vignettes describing two hypothetical states and their decision-making process, following the steps depicted in the flowcharts.

EXTENSIONS: USING ALTERNATE ASSESSMENT GROWTH MEASURES IN OTHER CONTEXTS

Results-Driven Accountability under IDEA

ED's Office of Special Education Programs (OSEP) holds states accountable for their implementation of IDEA- and requires states to track improvement in results for students with disabilities. Under OSEP's Results Driven Accountability (RDA) framework, all states must set annual achievement targets on 16 common indicators. (Each state also must develop an SSIP, and evidence about that plan serves as the 17th indicator.) OSEP makes annual determinations about each state, based on the evidence it submits in the State Performance Plan/Annual Performance Report (SPP/APR) and other sources at its discretion (e.g., from site visits). Based on evidence of IDEA compliance and results for children and families, OSEP identifies each state using one of four categories: meets; needs assistance; needs intervention; or needs substantial intervention.

Data about participation in and performance on the AA-AAAS are incorporated in OSEP Indicator 3 (Assessment). Academic growth is not included in any of the standard 16 indicators, nor is it required in the SSIP. However, there is an option to use assessment growth to evaluate progress toward meeting a State Identified Measurable Result (SIMR), which is an outcome measure for the SSIP. Based on an analysis of states' reports on their SSIPs in FY17 (U.S. Department of Education, 2019), most states (58%) targeted improved outcomes in reading. Other states identified SIMRs in both reading and math (25%); math only (12%); or graduation, post-school outcomes, or early childhood outcomes (5% total).

Within the academic subjects, SIMRs might be based on targets set for specific grade(s) and evaluated across cohorts (e.g., increasing the percentage of students who are Proficient in reading by Grade 3), or targets which are set and monitored for groups of students over time (e.g., reducing achievement gaps). States also have the option to focus on subgroups of students with disabilities. In 2018, 35 states included all students with disabilities in their SIMRs. Thirty-five states used state assessment data but only seven states reported using alternate assessment data to evaluate outcomes (Lazarus et al., 2021). The vast majority of SIMRs evaluated change across cohorts of students (i.e., improvement) rather than growth within cohorts.

There are some similarities to expected practices for developing and implementing an ESSA accountability plan and an IDEA SSIP. States are expected to engage stakeholders, rely on a theory of action and evaluate the quality of their data. Expectations for the SSIP may differ from ESSA accountability when deciding how to establish a baseline or set SIMR growth targets over time based on expected impact of implementing evidence-based practices.

For states using large-scale assessment growth as their SIMR, an alternate assessment growth measure provides an opportunity to include alternate assessment results in the SSIP. It also holds states accountable for improving outcomes for entire cohorts of students with disabilities. A growth-to-target approach could help states set appropriate targets and monitor student progress over time without having to focus narrowly on proficiency.

DISTRICT, SCHOOL AND CLASSROOM USE

As described in Chapter 2, growth models differ in the degree to which they produce easily-interpreted growth measures. The more interpretable growth measures could be useful at the local level to guide educational programming and instruction. For example, a district that did not meet its target on the academic growth indicator for students with disabilities might review its data to identify specific areas to target for improvement. Such a district also would benefit from having equivalent types of data on students with disabilities who take the alternate or general academic assessment, because service-delivery models for these subgroups often look different. Taking a data-based decision-making approach enables states to use disaggregated growth measures based on groups of students who take the alternate or general assessment, and it is especially useful for those states that have robust reporting systems with role- or user-specific access.

Districts that use interim assessments to monitor academic progress may use an assessment for the general population that produces some measure of within-year growth. If a state has an alternate interim assessment system that is appropriate for students with SCD, the assessment might produce a within-year growth measure to inform instruction — which may be more feasible for outside the high-stakes type of accountability associated with ESSA. In this case, any of the same prerequisites and processes described earlier in this chapter would apply: the district would start by clearly specifying the purpose and use of the growth measure, then conceptually define expectations and growth for the population in a way that can be supported by the assessment as it is designed. A state would need to check the evidence of validity and technical quality and the defensibility of the assessment scores w before selecting, evaluating and operationally using a growth model.

Note that, while many of the general steps would apply, the selected approaches may differ substantially, depending on the prioritized use cases. For example, a state or district that privileges providing instructional information to educators might develop a system based on more frequent, focused assessments that sacrifice standardization and broad generalization to be more customized to individual student learning goals. Districts developing their own alternate interim assessments might have limited resources available to complete these steps or to produce their own growth measures. Districts using commercially-available interim assessments would need to evaluate evidence provided by the test publisher. A recent review of commercially-available interim assessments indicated that no alternate interim assessments could be found (Boyer & Landl, 2021). More work is needed in this area.

An alternate assessment growth measure could also be used at the individual student level to inform instruction and communicate with parents. For example, individualized education program (IEP) teams could include information about academic growth when describing a student's present levels of performance or setting targets for IEP goals and objectives. In a system where the alternate assessment scores can produce information about within-year growth, this information could be used during parent-teacher conferences to discuss student progress and next steps.

MONITORING, EVALUATION AND RESEARCH

Assuming states have selected, implemented and evaluated alternate assessment growth measures for operational use and found them to produce the intended results, the growth measure may also be useful for other research, evaluation or monitoring purposes. For example:

1. The growth measure could be used as an outcome measure in intervention research evaluating the impact of new instructional strategies on student learning. It also could be used to guide curriculum changes at the school or district level.
2. Researchers and states could use growth measures as part of research designed to evaluate the effectiveness of various education models (e.g., inclusive, self-contained or, public separate schools).
3. States that use a growth model which produces actual growth relative to an expected amount of growth could use the growth measure to monitor policy implementation or evaluate instructional practice. For example:
   a. Growth and status can be used together to help states and districts identify students who are potentially misidentified for participation in the alternate assessment and are more appropriately assessed using the general academic assessment. A state could include language about growth in its definition of SCD and the participation guidelines its IEP teams use to make assessment participation decisions.
   b. Outliers at both ends of the growth-measure distribution may be worth examining further. Outliers at the upper end of the distribution may have had exceptionally strong instruction or a change in conditions (e.g., access to augmentative and alternative communication) that supported greater than expected growth. Outliers at the lower end of the distribution, especially if there are multiple students with the same teacher, may indicate that the teacher needs additional support or professional development.

CONCLUSION

This handbook describes processes and considerations for selecting growth models, creating alternate assessment growth measures and including alternate assessment growth measures in accountability growth indicators. Key questions and considerations are woven through each chapter, and states will follow different processes and arrive at different decisions. In closing, it is valuable to reiterate several key principles to guide this work:

1. Policy uses and intended interpretations should be the primary driver of states' design decisions.
2. The technical adequacy of the AA-AAAS and its scores, the growth measure and the growth indicator must support the state's policy uses and interpretations.
   - ESSA growth indicators are intended to describe or predict growth over time for a group of students. The indicator is part of a larger accountability system and is not used in isolation for high-stakes program evaluation. States should seek a level of technical adequacy for its AA-AAAS growth measure that will allow it to be meaningfully incorporated into the state's growth indicator. This may not require using a complex growth model or the same model used for its general assessment.
3. AA-AAAS growth data may be reported at the individual student level, where families and teachers are key stakeholders. It also may be reported at an aggregated level, to be used by a range of stakeholders. If the data support intended interpretations at the desired reporting level and also support the state's policy goals, states should consider including the AA-AAAS in the academic growth indicator to call attention to the importance of equitable treatment of all students.
4. Be cautious with regard to the potential for negative consequences and unintended uses of any AA-AAAS growth data. During the design and development phase, identify potential issues that might be associated with interpretations based on the alternate assessment and/or any alternate standards or extensions. Mitigate potential challenges associated with any technical constraints of the alternate assessment by matching an appropriate model to the data available. Collect data during operational use to evaluate whether the anticipated risks manifested — and if they did, whether the mitigation plans worked.
5. States should pursue this work using an equity lens. As the field continues to evolve, it is important to remain open to new ways of thinking about and evaluating equity and new ways to use AA-AAAS growth data to evaluate educational equity in programs and outcomes. Equity should be considered with regard to both how students can access the assessment fairly and how all students are represented in the state's accountability system.
6. States should continue to conduct research that helps them evaluate the efficacy of their growth indicators — including monitoring whether the indicators are having the desired effects — and be prepared to make necessary revisions.

This handbook is written at a time when few states include alternate assessment in their growth indicators, and there are limited examples from the research or best-practice literature to serve as models for states to consider. Significantly more work is needed in this area to lay a stronger foundation for the conceptual, technical, operational and policy aspects of alternate assessment growth.

While this work is complex, it is worth doing for the sake of equity, inclusion and authentic representation of students with significant cognitive disabilities in U.S. public education. States that engage in this important work are encouraged to share their lessons learned through professional networks and organizations such as the Council of Chief State School Officers (CCSSO). Sharing insights and information in the field of alternate assessments will prevent states from pursuing solutions that are unlikely to be viable and will serve to elevate and hasten the identification of best practices for students with SCD.

As best practices are identified and evolve over time, states also will need to consider their legislative and policy contexts when making decisions about whether and how to create AA-AAAS growth measures or to include alternate assessment growth in an accountability growth indicator. Grounding all decisions in the ultimate goal of promoting equity and improving student outcomes will maximize the benefits of alternate assessments and expand the positive applications of AA-AAAS results.

"While this work is complex, it is worth doing for the sake of equity, inclusion and authentic representation of students with significant cognitive disabilities in U.S. public education."

REFERENCES

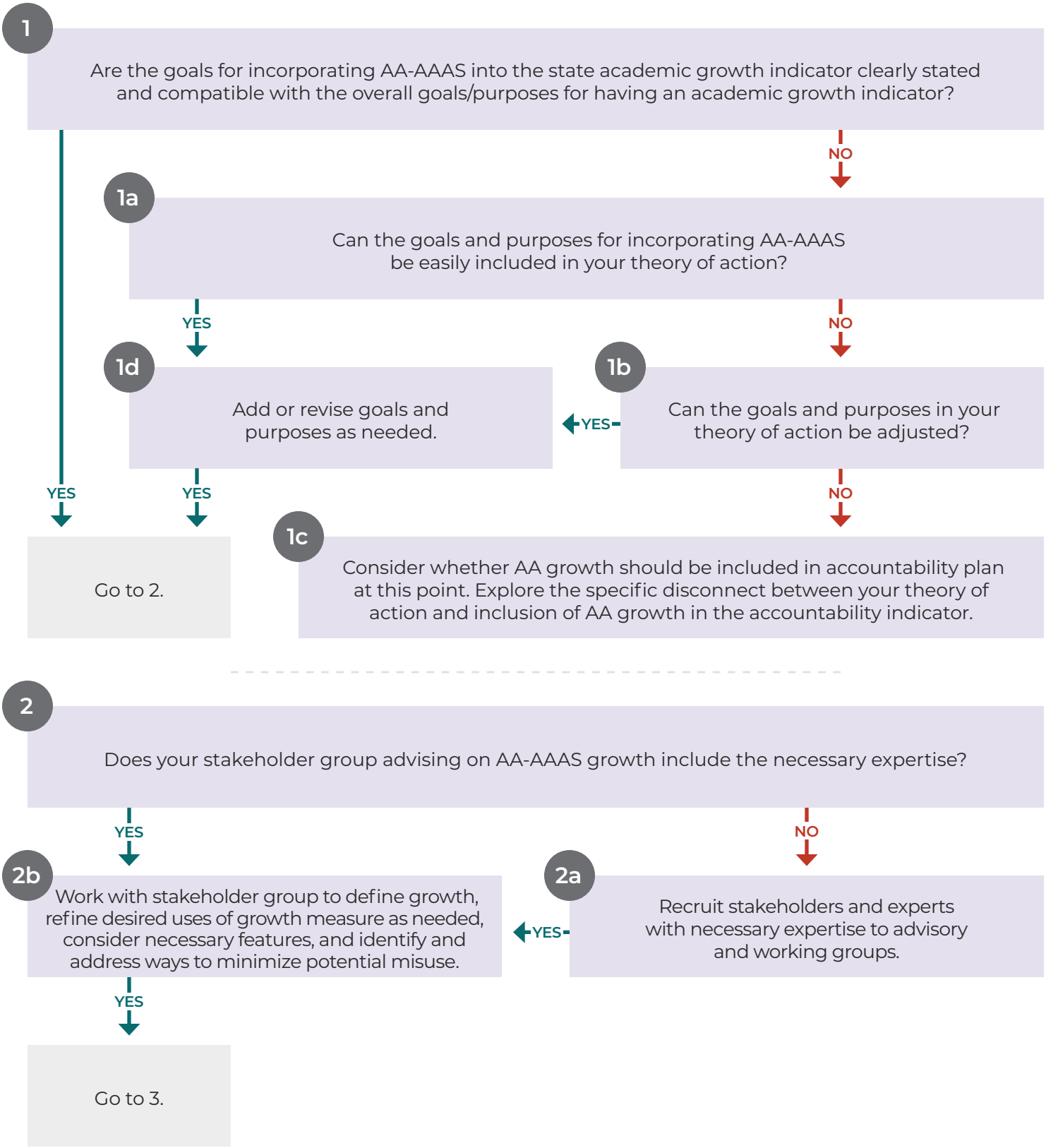
Boyer, M., & Landl, E. (2021, April). *Interim Assessment Practices for Students with Disabilities* (NCEO Brief #22). University of Minnesota, National Center on Educational Outcomes and National Center for the Improvement of Educational Assessment. Retrieved from https://nceo.umn.edu/docs/OnlinePubs/NCEOBrief22.pdf

D'Brot (in press). *Operational Best Practices for Accountability*. Council of Chief State School Officers.

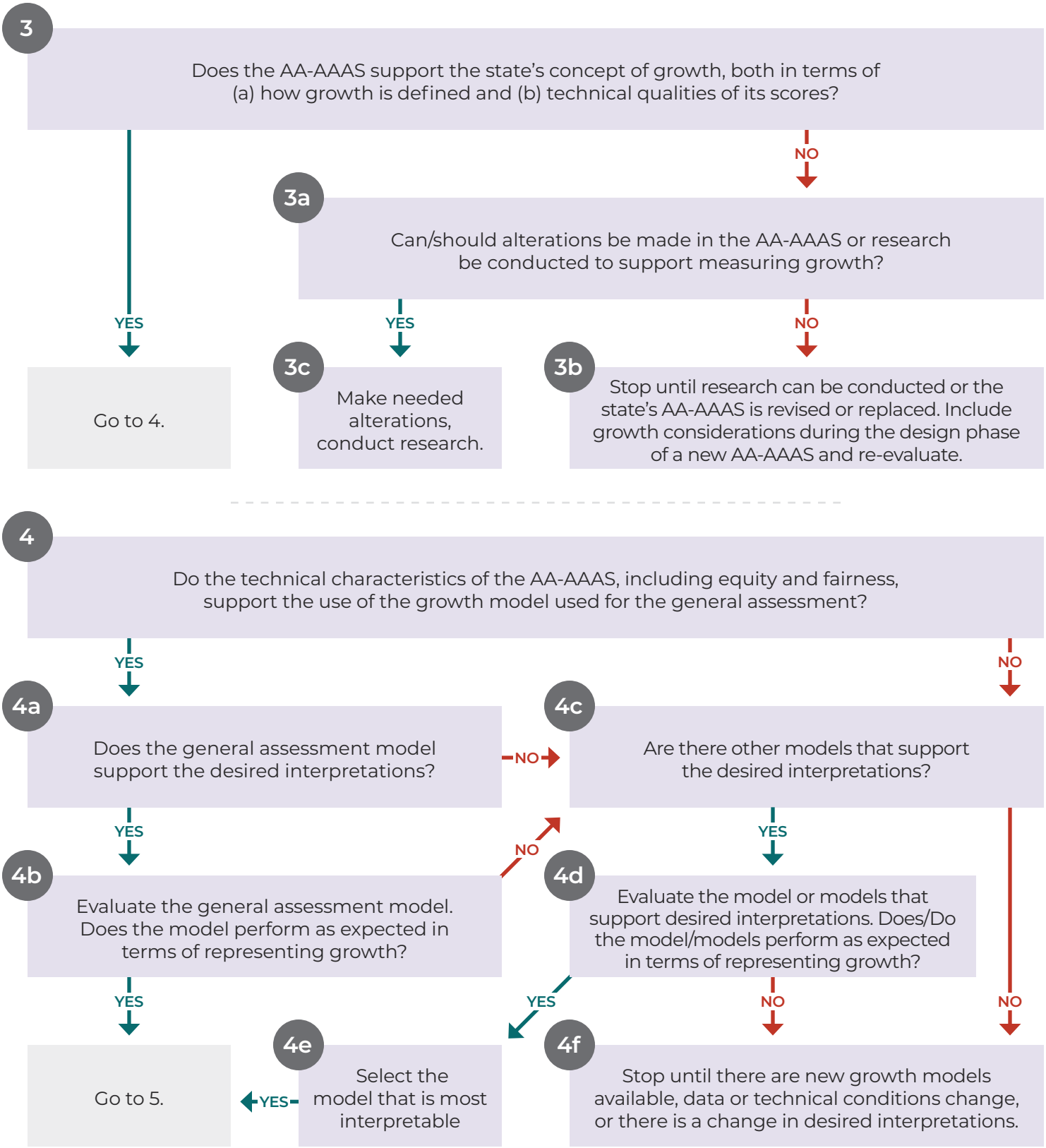
Lazarus, S. S., Hayes, S. A., Nagle, K., Liu, K. K., Thurlow, M. L., Dosedel, M., Quanbeck, M., & Olson, R. (2021). *The Role of Assessment Data in State Systemic Improvement Plans (SSIPs): An Analysis of FFY 2018 SSIPs* (NCEO Report 425). National Center on Educational Outcomes. Retrieved from https://nceo.umn.edu/docs/OnlinePubs/NCEOReport425.pdf

U.S. Department of Education, Office of Special Education Programs (2019). *Part B FFY 2017 SPP/APR Indicator Analysis Booklet*. Retrieved from https://osep.grads360.org/#communities/pdc/documents/18278

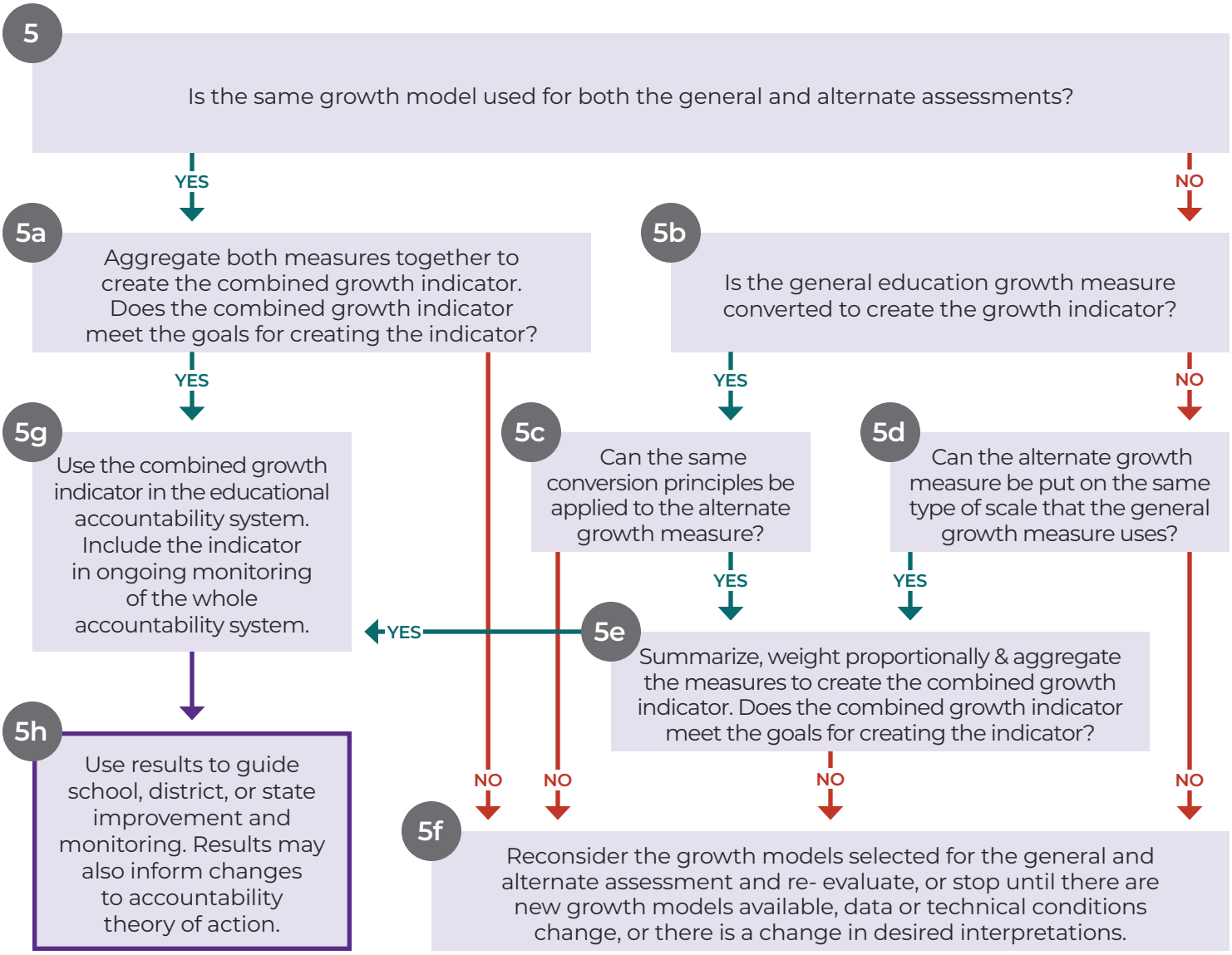
APPENDIX A: FLOWCHARTS TO GUIDE STATE DECISIONS ABOUT AA-AAAS GROWTH



APPENDIX A: FLOWCHARTS TO GUIDE STATE DECISIONS ABOUT AA-AAAS GROWTH



APPENDIX A: FLOWCHARTS TO GUIDE STATE DECISIONS ABOUT AA-AAAS GROWTH



APPENDIX B: VIGNETTES OF STATE DECISIONS RELATED TO ALTERNATE ASSESSMENT GROWTH MODELS, MEASURES AND INDICATORS

This appendix summarizes two states' hypothetical approaches to considering AA-AAAS growth. They illustrate how different states might follow the steps in the flowchart described in Chapter 6 and answer questions related to policy and technical considerations introduced in Chapter 4. Table 4-1 is duplicated here, with each consideration labeled, to illustrate their correspondence to the steps in the vignettes. Parenthetical numbers in the vignettes (e.g., P3) correspond with the considerations listed in the table below.

FIGURE 4-1: CONSIDERATIONS WHEN SELECTING A GROWTH MODEL (adapted from D'Brot 2017)

POLICY CONSIDERATIONS	TECHNICAL CONSIDERATIONS
P1 - What policy goals or signals does the state want to send in relation to student growth on the alternate assessment?	T1 - What are the technical characteristics of the AA-AAAS? What scores are produced? How many students take the AA-AAAS?
P2 - In what ways should growth on the AA-AAAS inform conversations about individual student progress?	T2 - Are growth-model calculation requirements supported by the technical characteristics of the AA-AAAS? (This may include issues associated with the stability of the population, not just the sample size of students.)
P3 - How can the application of a growth model on the alternate assessment improve outcomes by promoting progress (rather than making a summative determination or punitive judgments) for both lower- and higher- performing students?	T3 - What are the technical characteristics of the growth measure? How might they inform inclusion in the growth indicator (discussed in Chapter 5)?
P4 - How can the growth model be applied equitably when characterizing student growth across critical subgroups? How does the state develop the measure, and how can the state communicate about its meaning in way(s) that minimize unintended consequences?	

Caveats

1. These vignettes are not intended to be compared or contrasted to one another, but are provided as alternative approaches based on different contexts.
2. The vignettes are intended to show examples of how the process of selecting and implementing a growth model based on the AA-AAAS could work. While the flowchart steps and technical and policy questions are expected to be useful to many states, each state will use its own processes and reach its own decisions.

APPENDIX B: VIGNETTES OF STATE DECISIONS RELATED TO ALTERNATE ASSESSMENT GROWTH MODELS, MEASURES AND INDICATORS

State A framed its goals from a systems perspective, whereas **State B** framed its goals from a student perspective. Both frameworks are appropriate, as long as the decisions made throughout the design, development, implementation and evaluation phases of developing a growth model based on the AA-AAAS are coherent within the state. While these vignettes do not exhaustively represent approaches to growth, they are intended to provide concrete examples of some of the issues and decisions states should consider.

STEP	STATE A	STATE B
	State A uses a categorical growth model, based on its four achievement levels — Advanced, Proficient, Approaching Proficient and Basic — for general academic assessments. (T3) *Their AA-AAAS produces scale scores (scaled within each grade) that are converted to the same four achievement levels.* *Approximately 500 students per grade (3,500 total) statewide take the AA-AAAS annually. (T1)*	*State B uses SGP as a growth model for their general assessment and calculates the proportion of students on track to meet the high school-Proficient standards for each school and subgroup. (T3)* *Their AA-AAAS produces scale scores on a vertical scale that are converted to four achievement levels.* *Approximately 1,200 students per grade (8,400 total) take the AA-AAAS. (T1)*
1. Are the goals for incorporating AA-AAAS into the state academic growth indictor clearly stated & consonant with the overall goals/ purposes for having an academic growth indicator?	The state wants to signal that growth on the AA-AAAS is equally as important as growth on the general assessment. (P1, P3) The state's overall goal for its academic growth indicator is to characterize academic growth clearly and easily to the public and educators. Its purposes include ensuring that all students are on track for post-secondary success in higher education or the job market. The state's accountability theory of action describes its goals and purposes broadly enough to be inclusive of desired outcomes for all students, including those who take AA-AAAS. The state's general growth model values students moving toward and into the Proficient range, or staying in the Proficient range, more than moving beyond Proficient. (P1)	The state wants to indicate that students with SCD are expected to progress in their academic knowledge, skills and understanding over time. (P1, P3) The state's overall goal for its academic growth indicator is to measure the degree to which a school's students are on track to become Proficient or above-Proficient and to provide teachers and leaders with information that can help them direct resources to improve instruction. The state's accountability theory of action focuses on the goal of preparing students for college and careers after graduation. The theory of action was developed before the state had a graduation option for students who take AA-AAAS. Once the state has developed its alternative diploma option, the state will revise its accountability theory of action to be inclusive of both pathways to graduation and articulation of postsecondary readiness for both groups. (P1)

APPENDIX B: VIGNETTES OF STATE DECISIONS RELATED TO ALTERNATE ASSESSMENT GROWTH MODELS, MEASURES AND INDICATORS

STEP	STATE A	STATE B
2. Does your stakeholder group advising on AA-AAAS growth include the necessary expertise?	The state engaged a member of its special education department when designing its original ESSA accountability plan. They will bring that person back to the table for this growth conversation. Lacking internal capacity, they will modify their contract with their AA-AAAS vendor to add time needed for conducting additional analyses to inform this work. Their TAC, which advises on assessment and accountability, includes a member with expertise in alternate assessments. The state anticipates forming an ad-hoc group of educators with expertise in special education and the relevant content areas, along with parents and technical advisors, to review and provide input on draft plans. (P2).	The state developed its accountability theory of action and ESSA plan with minimal input from stakeholders responsible for identified subgroups of students (e.g., students with disabilities, English learners, students living with poverty). Because there were enough students taking the AA-AAAS in each grade level, the state initially explored using SGP for its AA-AAAS, so that the same model would be used for both general and alternate assessments. Because an SGP model is normative, the Exceptional Children's Office (ECO) staff had concerns about whether it was appropriate for use with the state's AA-AAAS. ECO staff immediately asked questions of the accountability staff about how growth is defined and the expected rate of growth for students who take AA-AAAS. The state decided to convene a broader stakeholder group to address some of these questions before taking any additional steps. Their TAC has a member with general expertise on students with disabilities but not alternate assessments. On the TAC's recommendation, the state assessment and accountability staff identified an alternate assessment expert to serve as an ad-hoc member of the TAC to advise on this line of accountability work. (P2)
3. Does the AA-AAAS support the state's concept of growth, both in terms of (a) how growth is defined and (b) technical qualities of its scores?	The state defines growth as moving toward or maintaining proficiency from grade to grade. The state expects students who take AA-AAAS to be able to make progress toward that goal, recognizing that growth is likely to be slower or more variable for students who take AA-AAAS. (P2) The AA-AAAS has four achievement levels, as does the general assessment. The state's peer-review committee determined that it had sufficient reliability evidence for achievement level classifications. However, the scores of most students who are ranked as Advanced on the AA-AAAS tend to cluster near the Proficient/Advanced cut score. Therefore, the range of earned score points in the Advanced level is smaller than in the other three levels. (T1) The state decides to adjust the measure for alternate growth to reflect this gap in scores at the upper level.	Based on the stakeholder input, State B decided that a growth model which supports normative interpretations, as SGP does, was not appropriate. The state picked from among the options for calculating absolute growth and decided to evaluate two models: simple gain score and categorical gain. (P2) The AA-AAAS has sufficient reliability evidence for achievement level classifications. In some grades and subjects, there are few items to anchor the upper and lower ends of the scale. The conditional standard error of measurement is larger than would be expected, especially at the lower end of the scale. (T1)

APPENDIX B: VIGNETTES OF STATE DECISIONS RELATED TO ALTERNATE ASSESSMENT GROWTH MODELS, MEASURES AND INDICATORS

STEP	STATE A	STATE B
4. Do the technical characteristics of the AA-AAAS, including equity and fairness, support the use of the growth model used for the general assessment?	The AA-AAAS has substantially met peer-review requirements, including sufficient evidence for the critical elements addressing validity, technical quality and achievement standards. None of its remaining areas for additional evidence are in those three sections. (T1)	The AA-AAAS has substantially met peer-review requirements, including sufficient evidence for most of the critical elements addressing validity, technical quality and achievement standards. The state has been asked to provide additional evidence of its response to external-alignment study findings that highlighted a need for improved alignment in high school ELA. In addition, the state needs to review and take steps to address differential item functioning found in some mathematics items that performed differently for non-white and white students. (T1, P4) The state has implemented a policy decision in which the test administrator stops the assessment for any student who does not respond to five consecutive items. These students do not receive a scale score and are automatically assigned to achievement level I. Approximately 20% of students exit the assessment early due to the stopping rule. (T1)

APPENDIX B: VIGNETTES OF STATE DECISIONS RELATED TO ALTERNATE ASSESSMENT GROWTH MODELS, MEASURES AND INDICATORS

STEP	STATE A	STATE B
5. Is the same growth model used for both the general and alternate assessments?	The state can use a categorical gain model for AA-AAAS, as it does for general assessments. In the general assessment growth model, students receive points for different degrees of growth across categories that are defined by subsets of their academic achievement levels. Each of the three levels below Advanced (i.e., Basic, Approaching Proficency and Proficiency) is divided into two categories, Low and High. Weights range from 1.0 to 3.0, depending on how many subcategories the student gains from Year 1 to Year 2 (see *Value Table for General Assessment, State A*, at the end of this document). (T3) Recognizing that growth for students who take AA-AAAS may be more incremental, the state would like to implement a similar model for AA-AAAS growth data but consider adding a third subcategory in its lowest two achievement levels (see *Value Table for Alternate Assessment, State A*, at the end of this document). (T2) The state has been implementing its AA-AAAS for three years, yielding two waves of gain data. They are able to identify thresholds that support a decision on the number of subcategories that are sufficiently consistent and meaningfully differentiate subgroups in terms of content. (T2) The state decides to implement its plan to include AA-AAAS growth in its growth indicator as a pilot in the next school year. They will evaluate the impact on school and district determinations, especially for special schools and for schools in high-poverty areas. In addition, the state will communicate the intended purposes and uses of the growth measure; monitor how the measure is being used by educators, parents and the press; and identify misinterpretations/misuse and provide information about the appropriate uses of AA-AAAS findings. (P4) The state will attempt to add more alternate assessment items in the higher range of Advanced so that the subcategories can be the same for the general and alternate. They will monitor the effects of having more difficult items on the validity of inferences on the assessment as well as on the utility and fairness of the growth measure. (T1, P4)	Based on stakeholder feedback, the state previously decided against using the SGP, which is used for the general assessment. Because of CSEM concerns and the lack of scale scores for students who left the assessment early due to the stopping rule, the state has ruled out the simple-gain score. The state now wants to explore a categorical-gain model, realizing it needs to carefully consider how to weight movement from the policy-driven Level I designation (due to the stopping rule) into Level I based on a scale score after the assessment is administered in full. The state has accrued three years of AA-AAAS implementation data on its current AA-AAAS. So far, there is no clear pattern of data supporting a decision about the number of subcategories that are sufficiently consistent and yet meaningfully different in terms of content. The state discusses a research plan with its TAC. The plan includes *a priori* definitions and criteria for "meaningfully different" and "sufficiently consistent." The state will collect data for two more years and look across four waves of score change data before re-evaluating.

APPENDIX B: VIGNETTES OF STATE DECISIONS RELATED TO ALTERNATE ASSESSMENT GROWTH MODELS, MEASURES AND INDICATORS

VALUE TABLE FOR **GENERAL** ASSESSMENT, STATE A

Business Rules

Students who are below Proficient (Basic Low to Approaching High) in Year 1:

- do not receive any points if the stay in the same subcategory in Year 2
- receive .25 points for every increase in subcategory below Proficient
- receive an additional 1.25 points for movement between below-Proficient into Proficient, then .25 increase for every additional subcategory in Proficient High to Advanced High

Students who are at or above Proficient (Proficient Low to Advanced High):

- receive 1 point for staying in the same subcategory in Year 2, plus .25 points extra for every increase in subcategory in Year 2.

		YEAR 2 LEVEL							
		Advanced High	Advanced Low	Proficient High	Proficient Low	Approaching High	Approaching Low	Basic High	Basic Low
YEAR 1 LEVEL	Advanced High	1	0	0	0	0	0	0	0
	Advanced Low	1.25	1	0	0	0	0	0	0
	Proficient High	1.5	1.25	1	0	0	0	0	0
	Proficient Low	1.75	1.5	1.25	1	0	0	0	0
	Approaching High	2	1.75	1.5	1.25	0	0	0	0
	Approaching Low	2.25	2	1.75	1.5	0.25	0	0	0
	Basic High	2.5	2.25	2	1.75	0.5	0.25	0	0
	Basic Low	2.75	2.5	2.25	2	0.75	0.5	0.25	0

APPENDIX B: VIGNETTES OF STATE DECISIONS RELATED TO ALTERNATE ASSESSMENT GROWTH MODELS, MEASURES AND INDICATORS

VALUE TABLE FOR **ALTERNATE** ASSESSMENT, STATE A

Business Rules

Students who are below Proficient (Basic Low to Approaching High) in Year 1:

- do not receive any points if they stay in the same subcategory in Year 2.
- receive .25 points for every increase in subcategory below Proficient, **except** that for any upward movement within Basic, they receive .25 points (.25 points from Basic Low to Basic Medium, from Basic Low to Basic High, or from Basic Medium to Basic High).
- receive an additional 1.25 points for movement between below-Proficient into Proficient, then a .25-point increase for every additional subcategory in Proficient High to Advanced High.

Students who are at or above Proficient (Proficient Low to Advanced High):

- receive 1 point for staying in the same subcategory in Year 2, plus .25 points extra for every increase in subcategory in Year 2.

		YEAR 2 LEVEL								
		Advanced	Proficient High	Proficient Low	Approaching High	Approaching Medium	Approaching Low	Basic High	Basic Medium	Basic Low
YEAR 1 LEVEL	Advanced	1	0	0	0	0	0	0	0	0
	Proficient High	1.25	1	0	0	0	0	0	0	0
	Proficient Low	1.5	1.25	1	0	0	0	0	0	0
	Approaching High	1.75	1.5	1.25	0	0	0	0	0	0
	Approaching Medium	2	1.75	1.5	0.25	0	0	0	0	0
	Approaching Low	2.25	2	1..75	0.5	0.25	0	0	0	0
	Basic High	2.5	2.25	2	0..75	.5	0.25	0	0	0
	Basic Medium	2.75	2.5	2.25	1	.75	0.5	0.25	0	0
	Basic Low	2.75	2.5	2.25	1	.75	0.5	0.25	0.25	0

APPENDIX C: STATE USE OF ALTERNATE ASSESSMENT DATA IN GROWTH INDICATORS

(Based on a spring 2021 survey of ASES and ASR SCASS member states; 25 responses.)

STATE	How is the alternate academic assessment scored?	How are overall scores reported for the alternate academic assessment?	What growth model is used for the general academic assessment for federal accountability?	What growth model is used for the alternate academic assessment for federal accountability?	How is the general academic growth indicator calculated (if applicable)?	How is the alternate academic growth indicator calculated (if applicable)?	How are the alternate and general academic accountability measures/ indicators combined into a single academic indicator?
Arizona	Item response theory	Scale score, achievement level	Student growth percentiles	None	Median growth percentile	N/A	N/A
Arkansas	Diagnostic classification model	Achievement level, mastery profile	Value added	None	Aggregate across content; transform growth measure	N/A	N/A
Connecticut	Raw to scale	Scale score, achievement level	Growth to target/standard	None	Aggregate growth measures for groups; transform growth measures	N/A	N/A
District of Columbia	Item response theory (ELA and math); diagnostic classification model (science)	Scale score, achievement level	Student growth percentiles; growth to target/standard	None	Median growth percentile; growth to target/standard	N/A	N/A
Department of Defense	Raw to scale	Scale score, achievement level	None	In development	N/A	N/A	N/A
Georgia	Item response theory	Scale score, scale score range, achievement level	Student growth percentiles	None	Convert growth measure to categories	N/A	N/A
Hawai'i	Item response theory	Scale score, achievement level	Student growth percentiles	Categorical gain	Convert growth measure to categories; apply weights	Convert growth measure to categories; apply weights	Apply different rules to the two assessments, then aggregate within each assessment and weight the groups proportionally before combining
Indiana	Item response theory, vertical scale	Scale score	Student growth percentiles; growth to target/standard	None	Convert growth measure to categories	N/A	N/A

APPENDIX C: STATE USE OF ALTERNATE ASSESSMENT DATA IN GROWTH INDICATORS

(Based on a spring 2021 survey of ASES and ASR SCASS member states; 25 responses.)

STATE	How is the alternate academic assessment scored?	How are overall scores reported for the alternate academic assessment?	What growth model is used for the general academic assessment for federal accountability?	What growth model is used for the alternate academic assessment for federal accountability?	How is the general academic growth indicator calculated (if applicable)?	How is the alternate academic growth indicator calculated (if applicable)?	How are the alternate and general academic accountability measures/ indicators combined into a single academic indicator?
Iowa	Diagnostic classification model	Achievement level	Student growth percentiles	None	Aggregate measures for groups; apply weights; transform growth measures	N/A	N/A
Kansas	Diagnostic classification model	Achievement level, mastery profile	None	None	N/A	N/A	N/A
Kentucky	Raw to scale	Achievement level	None	None	N/A	N/A	N/A
Louisiana	Item response theory	Scale score, achievement level	Value added; growth to standard	None	Aggregate measures for groups; convert to categories	N/A	N/A
Michigan	Item response theory	Raw score, scale score, achievement level	Student growth percentiles	Percentile rank residuals	Percent meeting adequate growth	Percent meeting adequate growth	Apply the same rules to both measures
Mississippi	Raw to scale	Scale score, achievement level	Categorical gain	Categorical gain	Convert growth measure to categories and calculate average	Convert growth measure to categories and calculate average	Apply the same rules to both measures
Nebraska	Item response theory	Scale score, achievement level	Categorical gain; gain	Categorical gain; gain	Aggregate measures for groups; transform growth measures	Aggregate measures for groups; transform growth measures	Apply the same rules to both measures
New Jersey	Diagnostic classification model	Achievement level	Student growth percentiles	None	Aggregate measures for groups	N/A	N/A
North Carolina	Raw to scale	Scale score, achievement level	Value added	None	Aggregate measures for groups; transform measures; convert to categories	N/A	N/A

APPENDIX C: STATE USE OF ALTERNATE ASSESSMENT DATA IN GROWTH INDICATORS

(Based on a spring 2021 survey of ASES and ASR SCASS member states; 25 responses.)

STATE	How is the alternate academic assessment scored?	How are overall scores reported for the alternate academic assessment?	What growth model is used for the general academic assessment for federal accountability?	What growth model is used for the alternate academic assessment for federal accountability?	How is the general academic growth indicator calculated (if applicable)?	How is the alternate academic growth indicator calculated (if applicable)?	How are the alternate and general academic accountability measures/ indicators combined into a single academic indicator?
Ohio	Item response theory, adaptive	Scale score, achievement level	Value added	None	missing	N/A	N/A
Pennsylvania	Diagnostic classification model	Achievement level	Value added	None	Aggregate measures for groups; convert to categories	N/A	N/A
South Dakota	Item response theory	Scale score, achievement level, normative information	Student growth percentiles; growth to target/standard	Student growth percentiles; growth to target/standard	Convert to categories	Convert to categories	Apply the same rules to both measures
Texas	Item response theory	Scale score, achievement level, normative information	Gain	Categorical gain	Convert to categories	Convert to categories	Apply different rules to the two assessments, then aggregate
Utah	Diagnostic classification model	Achievement level, mastery profile	Student growth percentiles	None	Transform measures; convert to categories	N/A	N/A
West Virginia	Diagnostic classification model	Achievement level	Categorical gain	None	Convert to categories	N/A	N/A
Wisconsin	Diagnostic classification model	Achievement level	Student growth percentiles; value-added	None	Aggregate measures for groups; transform measures; apply weights	N/A	N/A
Wyoming	Item response theory	Scale score, raw score, achievement level	Student growth percentiles; growth to target/standard	Student growth percentiles; growth to target/standard	Aggregate measures for groups	Aggregate measures for groups	missing