White Paper: Considerations for Measuring Academic Growth on Dynamic Learning Maps® (DLM®) Alternate Assessments

---

June 2019

Nehler, C., Clark, A., & Karvonen, M. (2019). *White Paper: Considerations for Measuring Academic Growth on Dynamic Learning Maps® (DLM®) Alternate Assessments*. Lawrence, KS: University of Kansas, Center for Accessible Teaching, Learning, and Assessment Systems (ATLAS).

Contents

White Paper: Considerations for Measuring Academic Growth on Dynamic Learning Maps Alternate Assessments

Monitoring students' academic growth has been a central policy issue since the passage of the No Child Left Behind (NCLB) Act of 2002 due to its emphasis on students making adequate yearly progress toward academic proficiency. While the Every Student Succeeds Act (ESSA) of 2015 reduced this emphasis and provides more state flexibility, most state accountability plans still include indicators for academic growth or progress. This policy practice has important implications for alternate assessments based on alternate achievement standards (AA-AAS) due to the heterogeneity in the population of students with the most significant cognitive disabilities (SCD) who take AA-AAS and how their knowledge, skills, and understandings are expressed and measured. ESSA does not require growth measures for AA-AAS to be included in accountability formulas, but some argue that omitting AA-AAS from growth calculations disadvantages the students who take them. However, growth measures that are neither valid nor reliable may introduce unintended consequences for districts, schools, teachers, and students.

The Dynamic Learning Maps (DLM) Consortium, its Governance Board, and its Technical Advisory Committee (TAC) have been discussing the topic of growth for several years. The consortium's current stance (as of February 2019) is that the priority for reporting growth should be to provide teachers with an instructionally useful metric for evaluating student progress over time. At this time, we are working toward measures of student progress, but due to the long history of the term "growth", we will use that term throughout the rest of the paper to avoid making the discussion cumbersome. Upon advice from the TAC, DLM staff are exploring the issues associated with calculating and reporting growth on DLM assessments and identifying a research agenda to support interpretable and valid measures for reporting academic growth. This paper describes growth considerations for Dynamic Learning Maps (DLM) alternate assessments, including an overview of methods and implications for using them to report growth both across and within academic years.

What is Growth?

Academic growth is the degree of change in academic performance of a single student or a cohort of students across two or more time points (Castellano & Ho, 2013). Therefore, measuring academic growth requires two elements at a minimum: the same students or groups of students (i.e., cohorts) and data from at least two time points. Student-level growth either describes student growth over time or predicts future growth (Council of Chief State School Officers, 2017). Reporting growth at the student level can provide teachers with information they can use to inform classroom instruction. Similarly, predictive growth measures can inform teachers whether students need remediation to meet grade-level targets (Buzick & Laitusis, 2010). In contrast, cohort-level growth can be reported to describe the average growth for a group of students over time or to predict their average growth. Cohort-level growth has been used to evaluate school or district effectiveness for state accountability purposes.

Metrics for measuring growth vary from tracking within-student change over time to normative measures of between-student growth relative to peers. In addition to the granularity or type of information that stakeholders may be interested in, metric selection is also guided by the characteristics of the scores and scoring system used for the assessment. Within-student measures include *categorical gains* and *gain scores*. Categorical gains describe growth via ordinal, performance-level comparisons across consecutive grades (e.g., fourth grade to fifth grade), and gain scores describe growth via quantitative score comparisons across grades. In contrast, normative growth measures quantify growth relative to one's peers. *Student growth percentiles* (SGPs) are a common normative method for quantifying growth; they compare a student's growth at a point in time to that of similarly performing peers at a previous point in time.

Growth is often reported as the change in academic performance from one grade level to another. In some cases, it may also be possible to track academic growth within a single school year. The same common metrics can be applied to a within-year growth model. This model may be more desirable to some stakeholders who wish to evaluate growth relative to instruction.

Growth Considerations for Alternate Assessments

While calculation of academic growth has been widely studied using assessment results from general education assessments (e.g., Betebenner, 2009; Guarino, Reckase, Stacy, & Woolridge, 2015; Lockwood & Castellano, 2015), reporting growth for AA-AAS has received less emphasis for a number of reasons. Students who take AA-AAS are a widely diverse group of students; students eligible for AA-AAS are those with SCD, representing roughly 1% of students in the United States. The most common disability diagnoses among these students are intellectual disabilities, autism spectrum disorders, and multiple disabilities (Kearns, Towles-Reeves, Kleinert, & Kleinert, 2015; Nash, Clark, & Karvonen, 2015). Students taking AA-AAS are more likely to communicate at a presymbolic level, less likely to be socially engaged, more likely to have limited motor skills, and more likely to have general health problems than students taking general assessments (Towles-Reeves, Kearns, Kleinert, & Kleinert, 2009). Students may also have degenerative conditions that may be expected to lead to negative or slowing growth over time (Wei, Blackorby, & Schiller, 2011). Each of these factors adds complexity to evaluating academic growth both within and across students who take AA-AAS.

Historically, students with SCD were excluded from the general education curriculum, and their instruction prioritized functional skills over academic content (Jackson, Ryndak, & Wehmeyer, 2008). In spite of legislation and policies intended to improve access to the general curriculum, their opportunity to learn grade-level academic content continues to be uneven, with instruction often limited to narrow strands of content (Elliott, 2014; Taub, McCord, & Ryndak, 2017). When students with SCD have not had access to the full breadth of grade-level academic content, the calculation of growth relative to peers or proficiency expectations may be uninterpretable at best, and harmful at worst.

An additional challenge to reporting growth is cross-year variability in the population of students taking AA-AAS. ESSA limits participation in AA-AAS to 1% of students statewide and allows states and schools significant latitude in determining eligibility year over year. IEP teams typically determine whether a student is eligible for the AA-AAS, most commonly based on the student's SCD and/or whether the student requires substantial adjustments to the curriculum (Albus & Thurlow, 2012). Variation in eligibility policies may lead to inconsistent participation across years or subjects (Saven, Anderson, Nese, Farley, & Tindal, 2016), making reliable calculation of growth at either the individual or group level more complex (Domaleski & Hall, 2016). This issue with cross-year participation may be partially addressed using a within-year growth model; we will explore this option later in the paper.

Finally, decisions regarding how AA-AAS growth results are incorporated into accountability systems have implications for interpretability and consequences. One possible approach to incorporating AA-AAS growth results is to convert an AA-AAS growth measure into the same metric used for general large-scale assessments and then pool the results into a single growth index for the school. This approach may lead to the loss of nuance in interpretation of that growth information (if it is different for AA-AAS). This approach may also mask differences in observed growth for students who take general versus alternate assessments—differences that in theory should inform decisions about resource allocation to improve future instruction. On the other hand, treating AA-AAS growth as a separate criterion in accountability formulas may bring unintended consequences if the formula allows for a disproportionate effect. For example, if a school's overall accountability classification is affected by an indicator derived from the AA-AAS growth of very few students in the school and those students do not make the expected growth, the school's outcome may foster resentment in district and school staff toward those students and teachers and, in turn, decrease their future educational opportunities. This consequence is wholly inconsistent with the original goals for inclusive large-scale assessment that were behind legislation such as IDEA 1997 and NCLB. Models for evaluating educator effectiveness also require some year-to-year data to estimate teacher effects. The literature on including students with disabilities in educator-effectiveness calculations still primarily focuses on students who take general assessments and has not addressed AA-AAS (Buzick & Jones, 2015; Holdheide, Goe, Croft, & Reschly, 2010; Jones, Buzick, & Turkan, 2013).

Brief History of Growth for Alternate Assessments

As alluded to previously, evaluating student growth results from AA-AAS and using them in accountability systems has been less common than for general assessments to date. In a 2016 survey of 19 state education agencies, 14 states did not include measures of student growth on AA-AAS in their accountability metrics (Domaleski & Hall, 2016). For other states, it was not clear whether growth for AA-AAS was included. Consequently, no extensive precedent is available to inform AA-AAS growth calculations.

Among the few states in the 2016 survey that did indicate they included growth on AA-AAS in their accountability calculations, two states reported using categorical methods to calculate growth. Results from the Nebraska alternate assessments were used in a decision matrix for accountability purposes, comparing whether students' performance in the current year did not meet, met, or exceeded expectations based on their performance levels on the previous year's assessment (Domaleski & Hall, 2016). Florida used a similar performance-level-based approach for evaluating growth on its alternate assessment, whereby students demonstrated growth by maintaining or increasing their performance level (from among nine available levels) across years (Domaleski & Hall, 2016). Only one state reported using quantitative values to indicate growth. Michigan used SGPs on its AA-AAS to quantify growth among students taking the highest of three tiers of the assessment (Domaleski & Hall, 2016). This means that only a subset of Michigan's AA-AAS results were included in accountability calculations. Vertically scaled assessments can also be used to report growth by calculating gain scores, or the change in performance over time. However, Domaleski and Hall (2016) found no evidence that gain scores from AA-AAS have been used for accountability purposes.

DLM Alternate Assessments

The DLM Consortium administers AA-AAS to approximately 90,000 students in grades 3 through 8 and high school across 18 states, the District of Columbia, and a Bureau of Indian Education school (referred to collectively hereafter as *states*). DLM assessments measure students' knowledge and skills relative to grade-level alternate content standards, called *Essential Elements* (EEs), in English language arts (ELA), mathematics, and/or science (see Appendix A for state requirements by grade and subject).[1]

[1] Two states administer only the science assessment. Three states administer only ELA and mathematics assessments.

The basis of DLM assessments is a learning map model, which consists of interconnected *nodes* (or skills) and pathways between them denoting the order of skill acquisition. The map itself is not tied to grade-specific expectations, but rather specifies the order of skill acquisition, from foundational through college- and career-ready skills. For each grade level, EEs are measured by neighborhoods of nodes in the map. Assessment *blueprints* stipulate the grade-specific EEs measured in each subject. Nodes are organized into *linkage levels* to provide students multiple access points to grade-level academic content. In ELA and mathematics there are five linkage levels, including three *Precursor* levels leading up to the grade-level expectation and a *Successor* linkage level extending beyond the grade-level expectation. In science, there are three linkage levels, with two Precursor levels leading up to the grade-level expectation. Each linkage level measures one or more nodes in the underlying learning map model.[2] A single node may be assessed at one or more grades, depending on the skills and EEs being assessed. For example, a node measured at the grade-level expectation in a lower grade may be measured at an *Initial Precursor* level in a higher grade.

States administering ELA and mathematics assessments select from either an integrated or year-end assessment model. The *integrated model* measures student knowledge and skills via through-course, instructionally embedded assessments, which are included in assessment scoring. The blueprint specifies the criteria teachers should meet when selecting EEs for instruction and assessment (see Appendix B for an example). Teachers can choose to assign an EE more than once at the same or a different level as part of system use. Through spring 2019, the administration model also included an end-of-year spring assessment for a subset of tested EEs or to assign new EEs to meet remaining blueprint-coverage criteria to update performance

[2] Science maps are currently under development.

after a full year of instruction. Beginning in 2019–2020, states using the integrated model will transition to two instructionally embedded assessment windows; during each window, teachers will select content to cover the full blueprint.

The *year-end assessment model* provides optional access to the instructionally embedded assessments for teachers who want to use them, but only results from a year-end spring assessment are used to report student achievement. Through spring 2019, administration was limited to 30 items, with the blueprint generally covering a wider breadth of EEs than the integrated model, resulting in many EEs being measured by only one or two items. Beginning in spring 2020, changes in the assessment model will ensure all EEs are measured by at least three items. Science assessments also use a year-end model for administration and scoring; all science EEs are measured by at least five items.

Assessment Features Affecting Growth Calculations

The DLM Consortium is guided by the belief that students with SCD can and do learn and progress academically in supportive environments. However, there is a dearth of research on how that process may unfold and what the most valid and reliable method to evaluate academic growth may be for these students. This paper expands on the reasons DLM results are inappropriate for application of common growth metrics generally and for accountability purposes especially. General assessments that use certain metrics to calculate growth among the broader population of students do not have the same population and measurement challenges that AA-AAS have. A subsequent paper will describe the many features of the DLM assessment and professional-development systems that enable monitoring student progress within the boundaries of the intended and currently supported uses of DLM assessment results. The advantages provided by the learning map structure and the instructionally embedded assessments available throughout the school year may provide instructionally useful and actionable information for teachers and parents concerning their students' academic progress.

A number of unique aspects of DLM assessments affect the calculation of academic growth, including the scoring model, the variability in underlying map structure and assessment design within and across grades, variability in student performance, and matching student records. The DLM TAC has advised against using growth results from DLM assessments for accountability purposes, as there are currently no known valid or reliable metrics for use with DLM assessment results.

Scoring model. Dichotomous student mastery of linkage levels, rather than a raw- or scaled-score value, serves as the basis for reporting student results. The system uses diagnostic classification modeling, in the form of discrete latent class analyses, to determine the probability of linkage-level mastery for each assessed EE. To avoid overly penalizing students, two additional scoring rules are applied: (a) a percentage-correct scoring rule, whereby students who correctly respond to at least 80% of items are considered masters; and (b) a two-down scoring rule, whereby students who test at a higher linkage level but do not demonstrate mastery are classified as masters of the linkage level two levels below the lowest level tested. Therefore, even masters of the same linkage level may have demonstrated their knowledge in different ways (and on different linkage levels).

For example, if a student in grade 5 tested at the *Target* linkage level for RI.5.1 (i.e., "Identify words in the text to answer a question about explicit information"), she is expected to demonstrate that she "can identify words or details to answer a question about explicit information presented in a text." If she does not demonstrate mastery at the Target level, she is classified as a master of the linkage level two levels below this one, or the *Distal Precursor*, where it is assumed that she “can understand a familiar text read aloud or through oral or other media by answering questions posed by others.” Another student tested at the Distal Precursor level and the scoring model classified him as a master based on his performance on the relevant items. These two students are both masters of the same skill, but that determination was made under two different circumstances. In practice, across all grades, subjects, and models, the most common way of demonstrating mastery is via model-based probability (66%–94%), followed by percentage correct (1%–30%) and the two-down rule (5%–30%; DLM Consortium, 2018a and 2018b).

Variability across grades. While there is some coherence in EEs across grades (similar to the trajectories evident in the Common Core State Standards), the number and content of EEs vary by grade. EEs are grade specific, as are the total number of available skills for which students can demonstrate mastery (where total skills equals the number of EEs multiplied by five linkage levels in ELA and mathematics and three linkage levels in science).

Because of the underlying map structure, there is also added complexity to interpreting total linkage levels mastered across grades. The skill acquisition necessary to demonstrate mastery of the Target level on one EE may be more or less than is needed to develop mastery for the same linkage level for another EE in a subsequent grade (e.g., EE.7.G.2 “Recognize shapes with specified attributes” and EE 8.G.2 “Recognize congruent figures”). Further, demonstrating mastery across linkage levels is also affected by the number of intervening nodes in the underlying learning map neighborhood. For instance, the five linkage levels indicated in an EE map neighborhood in high school span a substantially broader portion of the underlying map structure than a third-grade map neighborhood. This is done to encompass adequately the expanding range of skills reflected from the Initial Precursor to Successor level in higher grades.

These complexities mean that the distance, or learning, between linkage levels in a total linkage-level scale is not the same across linkage levels, EEs, grades, or subjects. Therefore, an interval scale of measurement—a scale in which a given distance between measures has the same meaning across the entire scale—cannot be assumed across grades, which is required for most calculations of growth.

Variability within grades. Similarly, an interval scale of measurement cannot be assumed within grades either. The total number of linkage levels mastered does not convey the skill acquisition necessary to demonstrate mastery across EEs. For example, summing student mastery for “use bar graphs to read data” (Target level M.EE.3.MD.3) and “tell time to the hour” (Target level M.EE.3.MD.1) does not account for variation across the EEs. Further, variation may exist within EEs, such as the amount of learning needed for a student to develop mastery at an Initial Precursor level on M.EE.3.MD.3 (e.g. “Recognize attribute values”) compared to the grade-level Target (e.g. “Use bar graphs to read the data”). Treating each linkage level equally to calculate change in total linkage levels mastered is conceptually misleading.

Additionally, the flexible integrated-model blueprints give teachers latitude in determining which (and how many) EEs and linkage levels students are tested on for each conceptual area (CA). It is possible that within the same classroom and year, some students will have tested on different (or more) EEs or linkage levels than their peers. In this case, disparities in total linkage-level mastery across students in a classroom are attributable to the teacher’s discretion and are not necessarily inherent differences in skills and knowledge among the students.

Variability in performance. To avoid interpretation challenges, the DLM Consortium does not report total linkage levels mastered in state return files (known as general research files or GRFs) or on Individual Student Score Reports (ISSR) shared with parents and teachers. This was intended to prevent stakeholders from mistakenly interpreting total linkage levels mastered as a total score or scaled-score value that they may be accustomed to receiving from general assessments. However, the total linkage levels mastered were used to define cut points between the four consortium-defined *performance levels*, which are then used to describe student achievement in the subject (Clark, Nash, Karvonen, & Kingston, 2017).

To illustrate the variability in performance, Figure 1 and Figure 2 summarize the total linkage levels students mastered by grade in ELA and mathematics, for integrated and year-end models respectively, and Figure 3 summarizes total linkage levels mastered for science. These figures include students who completed at least one *testlet* (short for *instructionally relevant testlet,* an engagement activity followed by three to nine items measuring the student's knowledge, skills, and understandings) during the 2017–2018 academic year. Completion of one testlet is also the definition of participation in AA-AAS for many consortium states and therefore these students represent the population most likely to be included in accountability calculations. Across both models, total linkage levels mastered in ELA were typically uniform across the full distribution. In mathematics, total linkage levels mastered tended to be positively skewed. In the year-end model, across grades and subjects, the most common number of total linkage levels mastered was generally zero. Because states have varying policies for which grade(s) and subjects are required for assessment, sample sizes and distributions of total linkage levels mastered vary. For example, in science, only one state administered the biology end-of-course exam, which is not depicted in these figures.

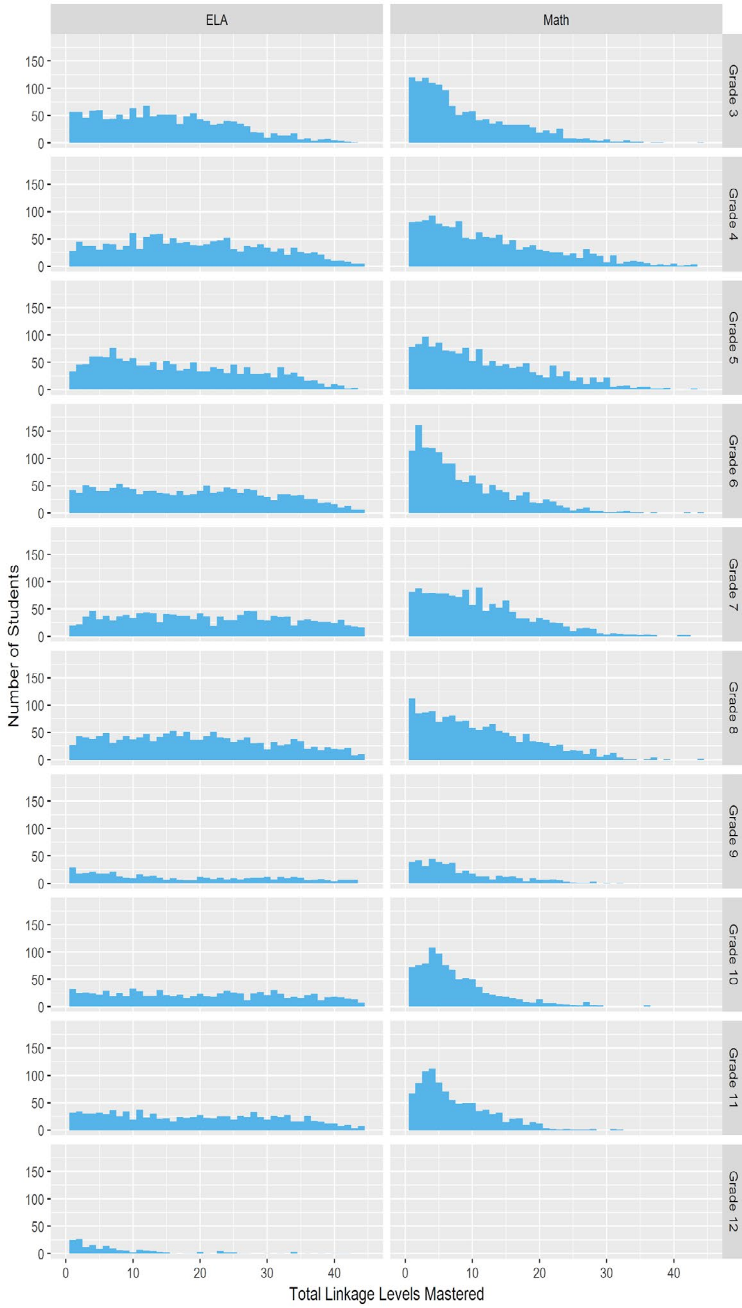

Figure 1. Integrated model total linkage levels mastered.

Figure 4 and Figure 5 summarize performance-level results for ELA and mathematics in the integrated and year-end models, respectively. Figure 6 summarizes performance-level results for science. In mathematics and science, students most frequently achieve at the *Emerging* performance level across all grades. This is also true for ELA in year-end model states, whereas the distribution for ELA in integrated-model states varies more across grades.

In addition to reporting performance level in the subject, ISSRs also summarize the percentage of mastered skills by CA, and states using the integrated model also receive linkage-level-mastery classifications. Together, the assessment results communicate to parents and teachers an individual student's academic achievement relative to grade-level alternate content standards.

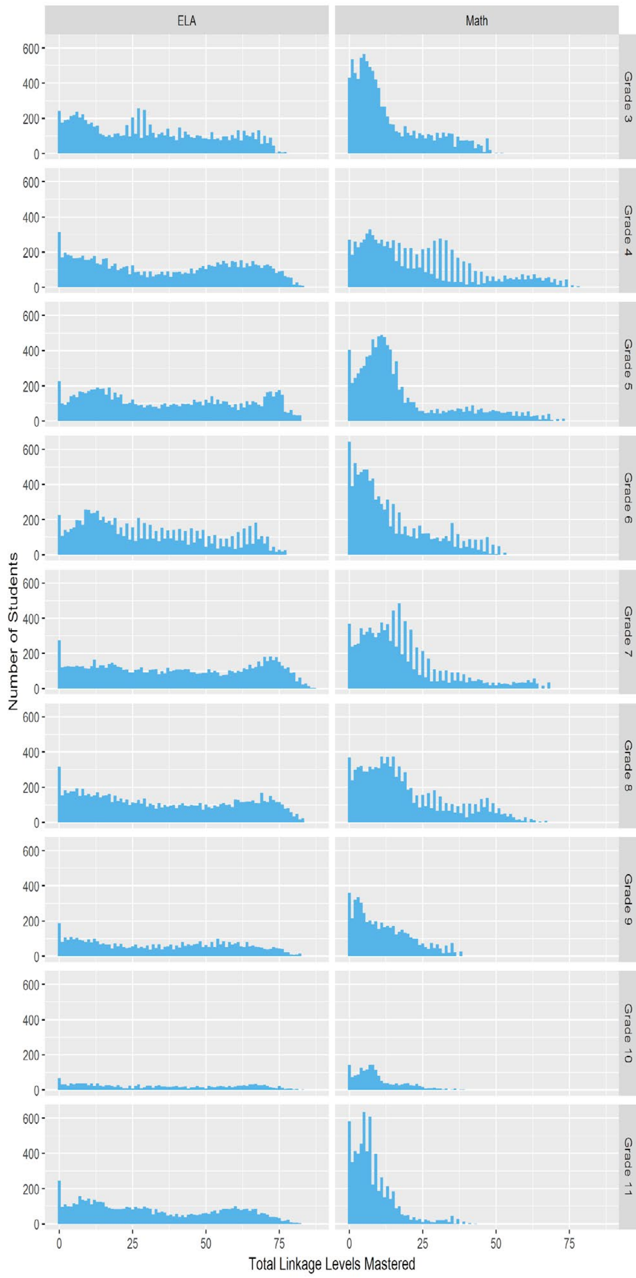

Figure 2. Year-end model total linkage levels mastered 2017-2018.

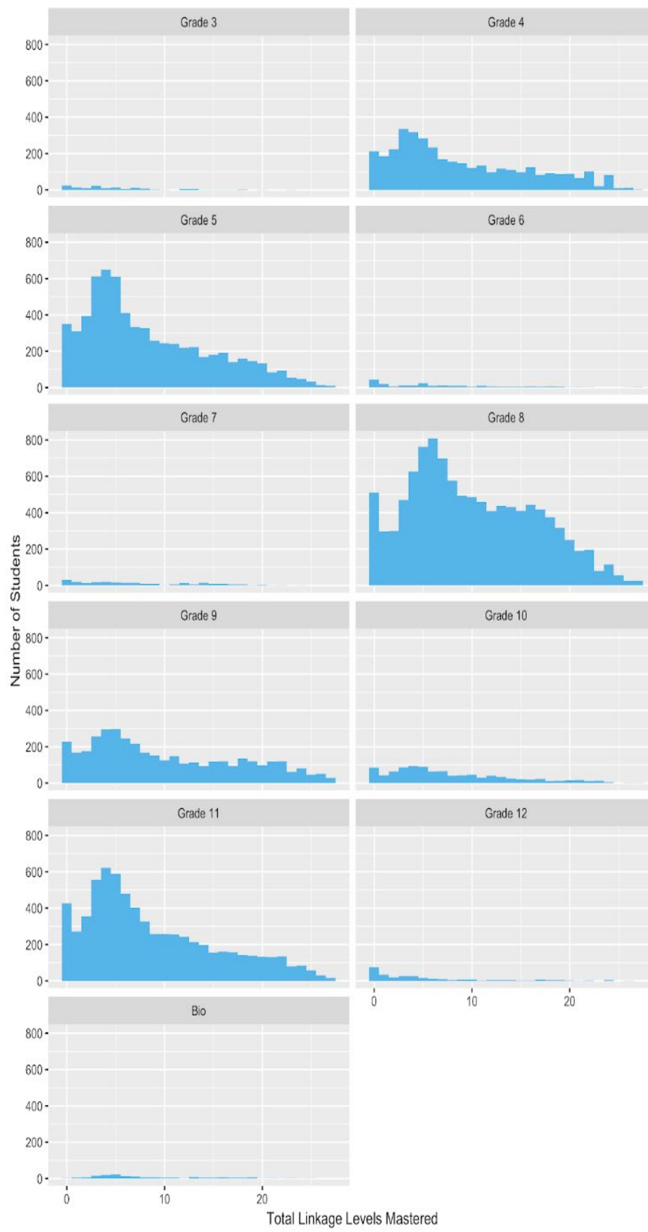

Figure 3. Science total linkage levels mastered 2017-2018.

Figure 4. Integrated model performance level distribution 2017–2018.

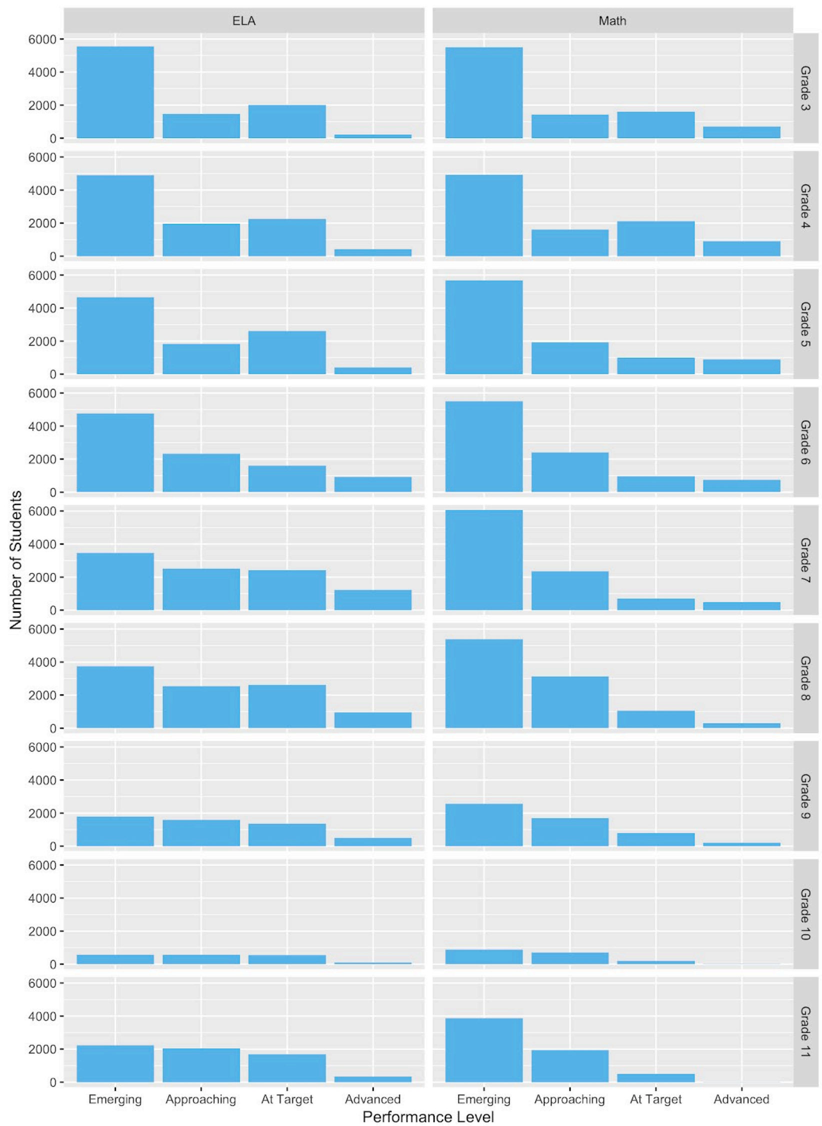

Figure 5. Year-end model performance level distributions 2017–2018.

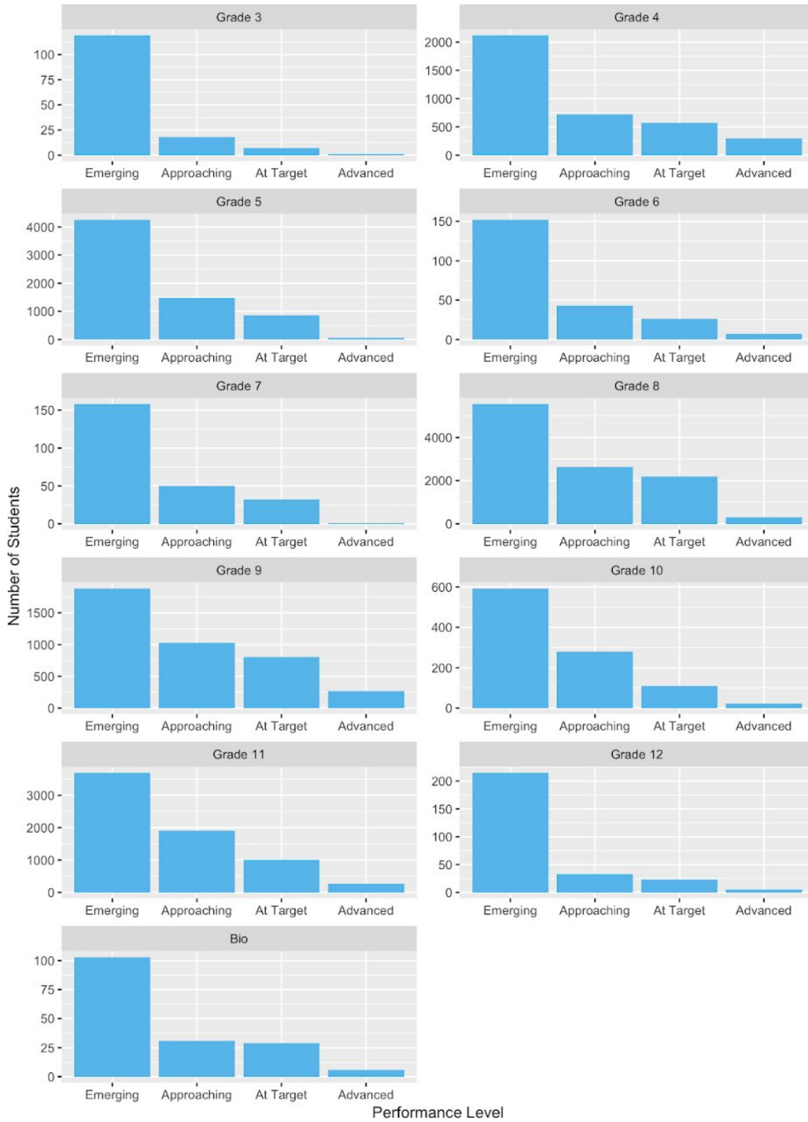

Figure 6. Science performance level distribution 2017–2018.

Matching student records across years. States vary in how (or whether) they retain student identifiers over years and which grade(s) they require to be assessed in high school and in science. Fluctuations in student participation across years and eligibility variation, both within and across states, also affects the ability to match student assessment performance records across time. DLM staff matched records of students who participated in DLM assessment for 3 consecutive years between 2016 and 2018. This number included 40,419 ELA students (48.3%), 40,375 mathematics students (46.6%), and 1,877 science students (5.5%).[3] Approximately 83% of these 3-year students take the DLM assessment in year-end-model states. Matching data across at least two time points is a requirement for each growth model, so the discussion of growth models assumes that stakeholders are able to match records across years and addresses other important considerations for each model with respect to DLM assessments.

In summary, the unique scoring method used to report student mastery of skills, the underlying map and assessment structure, and the feasibility of matching records over time raises a number of concerns for reporting growth on DLM assessments. Figure 7 summarizes growth considerations for AA-AAS and specifically DLM assessments.

[3] Percentages are based on combined 2018 DLM assessment participation in year-end and integrated-model states.

1. Heterogeneity among students taking DLM assessments
2. Eligibility variation within and across states
3. Diagnostic scoring model
4. No vertical scaling or interval-level scales
    a. EE content and number vary across grades
    b. EE level of challenge varies within grades
    c. Integrated model has flexible blueprint; students assess one or more times a year per EE at the same or different linkage level
    d. Year-end model limited by the number of items measuring the EE and linkage level; students assessed only in spring; most common number of mastered linkage levels is 0
5. Four performance levels describe achievement relative to standards
6. Cut points made for each grade separately
7. Distribution assumptions
8. Sample size reductions after splitting by model and grade
9. High school assessments typically not required annually
10. Matched records available for subset of population

Figure 7. Growth considerations for DLM assessments.

Evaluating Options for Reporting Growth for DLM Assessments

This section discusses methods and considerations for evaluating across- and within-year approaches to reporting academic growth for DLM assessments. Although there are numerous growth models, we limit our discussion here to those that are most commonly applied or that DLM Consortium states have indicated they currently use or may intend to use. We present the most salient considerations for each model generally and then discuss each in relation to the DLM assessment's particular growth considerations described in Figure 7.

In spring 2018, DLM staff distributed a survey on state accountability plans under ESSA to 17 partner states. Of the 13 states that responded to the survey, 10 reported that they do include growth on general assessments in their accountability metrics. All but two of these states indicated that they use SGPs; the others use categorical methods (e.g., transition matrices). Seven said that they report within-student growth, and three report cohort-based growth. Of responding states, only three indicated they calculate growth for AA-AAS, using either SGPs or residual gain methods. Two of these states pool general and alternate-assessment growth information, resulting in one metric for all students. One state computes and reports general and alternate-assessment growth results separately.

Year-to-Year Measures

Year-to-year growth metrics summarize student growth across academic years (i.e., grades). All growth models discussed here require data for students or cohorts from at least two time points. Considerations are presented for integrated and year-end assessment models and for science.

Gain scores. Gain scores summarize numeric change in performance across two time points. Simple gain scores describe the absolute magnitude and trajectory of a student's academic growth by subtracting the difference between the two values, therefore requiring both values be on the same scale (i.e., interval vertical scale). Residual gain scores use linear regression to empirically derive expectations for test performance based on the student's past scores and demographic variables of interest. The expectations can predict future scores or be contemporaneous with currently available scores. They describe whether a student's growth will meet, has met, will fall short of, or has fallen short of expectations, based on their past performance, by reporting the residual (i.e., the difference between their predicted and observed scores; Blackorby, Taylor, & Wei, 2016). Gain scores can be reported at the cohort level as an average gain score (Welch, Dunbar, & Rickels, 2016), for example, over subgroups or aggregation levels (e.g., class, school, district). Requirements for reporting gain scores are summarized in Table 1.

Table 1
Requirements for Simple and Residual Gain Scores Compared to DLM Assessment Properties

Simple gain	Residual gain	DLM assessments
Vertical scaling	Interval scale or better	No vertical scaling
Interval scale or better	Past performance information	Not interval scale or better
Two or more administrations		

Because they are numeric indicators of academic growth, both simple and residual gain scores have data requirements that DLM assessment results do not meet, such as vertical and interval scaling. Because of differing blueprints, subtracting raw, total linkage levels mastered across years is not appropriate; converting gain scores to percentages or *z* scores would retain the theoretical challenges created by treating all linkage levels as equivalent. For integrated-model assessments without additional constraints, students may demonstrate what might appear to be “growth” by exceeding blueprint requirements and testing only at lower linkage levels, which may reflect greater breadth of instruction, but not necessarily depth of content knowledge or progress toward the grade-level expectation in the EE. Finally, reporting growth based on a metric (i.e., total linkage levels mastered) that differs from actual score reporting would be incongruous and potentially confusing.

Student growth percentiles. SGPs are common metrics to demonstrate year-to-year academic growth. Most DLM Consortium states report using SGPs for their general assessments and may desire to report growth consistently across general and alternate assessments. SGPs describe a student’s current performance relative to academic peers. Quantile regressions compute SGPs by fitting 99 quantile regression lines, one for each percentile, to predict future achievement. They indicate both the student’s current status and growth relative to other students with comparable prior year scores, rather than to all students.

There are a number of requirements for using SGPs to report academic growth, especially for students served by DLM assessments, as summarized in Table 2. SGPs are a common growth metric because they do not require vertical scaling, allow stakeholders to compare students with similar performance histories, and they allow for variation in summarizing results (Castellano & Ho, 2013). These comparisons can include the median or simple average for peer groups (e.g., grade) or aggregate levels (e.g., school, district; Georgia Department of Education, 2017; New York State Education Department, 2018). However, SGPs typically require sample sizes of at least 5,000 students to achieve stable estimates of rank (Castellano & Ho, 2013; Grady, Lewis, & Gao, 2010).

Table 2
Requirements for Student Growth Percentiles and DLM Assessment Properties

Student growth percentiles	DLM assessment
Interval scale or better	Not interval scale or better
Sample sizes greater than 5,000 students	Integrated model; some grades have small samples
Some equivalence among peers (e.g., access to instruction, blueprint coverage)	Heterogeneous population
	Intended flexibility in integrated model

If used for DLM assessments, calculation of SGPs would compare the prior grade's performance (e.g., third-grade total linkage levels mastered) to performance in the subsequent grade (e.g., fourth-grade total linkage levels mastered). SGP calculations must be performed separately for each grade and subject by assessment model. Consortium membership is subject to change over time; whereas 5,000 students per grade may be feasible for current year-end model states, integrated-model membership does not support sample sizes of 5,000 per grade and subject. Four integrated-model states administered the 2018 assessment to 12,638 students, with fewer than 1,700 students at each grade level (DLM Consortium, 2018a).[4]

Norm-referenced growth measures like SGPs introduce additional challenges when comparing individual student performance to that of their peers. The variation in student population characteristics and individual state guidance on student eligibility for AA-AAS make peer comparisons difficult and present conceptual issues with interpreting a growth percentile for a student within this population. For instance, it is not conceptually clear what it would mean for a student to have demonstrated more progress than 90% of his or her peers on DLM assessments when many students have dramatically different disabilities, academic goals, and opportunities to learn the full breadth of academic content.

Because assessments are available at different linkage levels, the year-end-model spring administration includes adaptation between testlets (linkage level adjustment based on performance on the prior testlet), and the integrated model includes a flexible blueprint, students can master linkage levels any number of different ways. For this reason, total linkage levels mastered is not a meaningful value in this context. Comparisons between students with similar total linkage levels are murky and difficult to interpret. For example, two fourth-grade students may have mastered 10 linkage levels each in third grade. If calculating SGPs based on total linkage levels mastered on DLM assessments, these two students were academic peers and their growth would be compared against one another. The model assumes that mastery of 10 linkage levels has the same meaning for both students. However, because of individualized instructional priorities and DLM-assessment-system scoring rules, the students may have achieved 10 total mastered linkage levels on different EEs and in different ways (e.g., one student demonstrated

[4] Arkansas began implementing the integrated model in 2018–2019.

mastery at the Initial Precursor level through the two-down scoring rule from the Proximal Precursor level; another student demonstrated mastery on teacher-administered Initial Precursor testlets). It does not make conceptual sense to say that these two students are equivalently skilled in third grade and then compare their progress the following year when they will likely again demonstrate total linkage levels mastered in different ways.

Categorical gains. The categorical gains model describes growth based on performance-level results year to year. Implementation of the categorical model involves building a transition model, which is an N by N matrix in which N is the number of performance-level categories. The cells on the diagonal indicate students who maintained the same performance level across grades. The cells below the diagonal indicate students whose performance level decreased, and the cells above the diagonal indicate students whose performance level increased. See Table 3 for an example of the matrix.

Transitions between performance levels can be unweighted or weighted. In an unweighted model, all positive transitions are valued at +1, and all negative transitions are valued at -1, regardless of the number of performance levels that the student moved (i.e. even if the student improved by two performance levels, the unweighted model will still assign that student a growth value of +1). For weighted transitions, a value table specifies the assigned values for different transitions. For example, an increase from Level 1 to Level 3 can be assigned a value of +2, and a decrease from Level 2 to Level 1 can be assigned a value of -1. Table 3 depicts a weighted value table using DLM performance levels.

Table 3
Example DLM Weighted Value Table

Level	Emerging	Approaching	Target	Advanced
Emerging	0	+1	+2	+3
Approaching	-1	0	+1	+2
Target	-2	-1	0	+1
Advanced	-3	-2	-1	0

The categorical gains model could be applied in a year-to-year approach to individual student results from DLM assessments at the performance level (four levels). Cohort-level results could also be summarized as the percentage of students with a positive transition or the average transition value, and reported for specific subgroups or at varying aggregate levels (e.g. class, school, district, state). Table 4 lists the assumptions specific to categorical growth models.

Table 4
Assumptions for Categorical Growth Models and DLM Assessment Properties

Categorical gains	DLM assessment
Consistent performance levels across years Sensitivity to cut points Coarseness of information	Consistent performance levels; no vertical scaling and cut points determined for each grade separately Cut points derived using total linkage levels mastered Students with SCD show gains at finer-grained level

Note. SCD = significant cognitive disabilities.

Categorical models are naturally restricted in the number of categories available to describe student performance. DLM assessments report student achievement relative to four performance levels. There may be a great deal of variability in the skills and knowledge of students who achieve at each performance level, and improvements in students' skills may not be adequately captured by the overall performance level. For instance, a student with SCD may achieve significant progress toward IEP goals across years but score at the same performance level. A categorical gains model would not be sensitive to this type of growth. Additionally, the theoretical framework for interpreting growth scores for students with SCD is not well established. For example, a student who sustains the same performance level across two grades (even if below grade-level target) may, in fact, demonstrate positive growth rather than stagnation. If we view the student as having confronted higher (and thereby more complex) grade-level content with equivalent competence, in some ways this may indicate that the student has grown academically; whether this is the case cannot be ascertained from categorical models.

The categorical model requires that the performance-level categories be the same across grades and that each performance level represent the same relative degree of mastery across grades. For the DLM assessments, cut points between the four performance levels were set separately for each mode, subject, and grade. While statistical smoothing was used to arrive at final cuts for DLM assessments,[5] and every grade has the same four performance levels, the bounds of what constitutes achievement at each level varies by grade. For example, the Emerging performance level may look different in grade 5 than in grade 6 because of natural variability in panelists and the cuts they established during standard setting, the breadth of EEs covered by either blueprint, the complexity of linkage levels measuring the EEs, the underlying map neighborhoods, and grade-to-grade variability in EE content. These variations may result in a report of positive or negative growth that may not accurately represent the student's true academic progress over time.

[5] Science also included a vertical articulation process.

Within-Year Growth

Within-year growth summarizes student growth within a single academic year (i.e., grade). Reporting growth within an academic year for DLM assessments requires more than one administration of the same EE and/or linkage level testlet. For this reason, within-year growth could be explored only for states participating in the integrated model for ELA and mathematics. Considerations are summarized here for the revised integrated model that begins in 2019–2020, in which instructionally embedded assessments will have been administered in both fall and spring windows. Teachers will have had the option to assess the same or different EEs and linkage levels to fulfill blueprint requirements in each window and to control the timing of administration of parts of the assessment within windows that last approximately 15 weeks.

Gain scores. As with year-to-year growth, within-year growth could be measured on DLM assessments using simple or residual gain scores. If used, within-year gain scores require the same considerations as year-to-year gain scores (see Table 1) and the same interpretation of change in student performance over time. If applied to DLM assessments, within-year simple gain scores could be conceptualized within the academic year as the change in total skills mastered for the subject or CA, change in highest level mastered for each EE, or change in linkage-level-mastery status for each linkage level of every tested EE. Values would quantify the change in total linkage levels mastered from the first instructionally embedded window to the second, as shown below.

Subject or CA Gain Score = Total Spring Linkage Levels (within subject or CA) – Total Fall Linkage Levels (within subject or CA)

However, basing gain scores on total linkage levels mastered assumes results are provided on an interval-level scale, which, as we have demonstrated, they are not. As an alternative to using total linkage levels mastered, gain scores could also be provided at the EE or linkage level to summarize the number of levels mastered or change in mastery status, respectively.

EE Gain Score = Spring Highest Level Mastered – Fall Highest Level Mastered

Linkage Level Gain Score = Spring Mastery Status – Fall Mastery Status

Although they alleviate the problem with interval-level scale assumptions, gain scores based on EEs or mastery status pose additional challenges to reporting academic growth. Reporting EE gain scores requires that students be tested on the same EEs across the fall and spring embedded windows. Although teachers are expected to cover the full blueprint in each window, the intended flexibility of the integrated model allows teachers to choose EEs and linkage levels within coverage constraints (e.g., “Choose 3 EEs within Conceptual Area 1.1”). Variation across windows is expected and would introduce several issues if gain scores were used for growth on DLM assessments:

- Total linkage levels mastered in the subject or CA may be based on different EEs in each window, including more- or less-challenging ones.
- EE mastery is available only for EEs measured in both windows, which will likely vary across students, thereby creating disparate reporting issues.
- Exceeding blueprint-coverage requirements in either window may affect results positively or negatively.

The DLM scoring model may also affect interpretation. Linkage-level mastery is a categorical status indicator, and students can demonstrate mastery in three different ways. While the mastery probability threshold of 0.8 is consistent across all linkage levels, one cannot assume students possess the same degree of skill across all mastered linkage levels or that all linkage levels are equally challenging to master (e.g., see Chapter 5 of DLM Consortium, 2018b, for a summary of linkage-level mastery and non-mastery parameters).

Similarly, the two-down scoring rule may introduce examples of negative growth if a student tests at a higher linkage level in spring but does not demonstrate mastery. For example, in the fall of seventh grade, a student tested on RI.7.6 (i.e., “Determine an author’s purpose or point of view”) at the Proximal Precursor linkage level (i.e., “Can identify words or phrases for determining the point of view of an informational text’s author”) and was classified as a master of that linkage level. In the spring of seventh grade, the student’s teacher may decide to see whether the student had progressed and elected to test the student at the Target linkage level (i.e., “Can identify author’s point of view or purpose for writing an informational text on the topic at hand”). The student was classified as a non-master of this linkage level and, by the two-down scoring rule, is instead classified as a master of the Distal Precursor level, which is a level below the student’s performance in the fall.

Student growth percentiles. As with gain scores, SGPs could be reported in a within-year model. The results provide the same interpretation and carry the same considerations as year-to-year applications of SGPs (see Table 2).

If used with DLM assessments, within-year SGPs would theoretically evaluate spring performance based on total linkage levels mastered during fall administration. The SGP would provide a student’s percentile rank in the spring compared with students who performed similarly in the fall and allow comparisons among the students’ rates of academic growth in the year. These calculations and comparisons are not advised for DLM assessment results for the previously outlined reasons.

Categorical gains. Unlike a year-to-year model, for which a categorical gains model can be based only on performance-level transitions, a within-year categorical gains model used with DLM results could describe growth based on EE mastery or on linkage-level mastery. Implementation of the within-year categorical model involves building a transition model using the available categories for EE mastery (i.e., 6 or 4)[6] or linkage-level mastery (i.e., 2). Results can be summarized for groups of students, a hypothetical example is shown in Table 5. In this example of 1000 students, 700 of them began the year as non-masters of a particular linkage level, and 400 of those students went on to demonstrate mastery in the spring assessment. Put another way, of the 650 students who demonstrated mastery in the spring, 400 had been non-masters in the fall. This analysis of growth requires that students be assessed on the same EE and/or linkage level in the fall and spring. Cohort-level results can also be summarized as the percentage of students with a positive transition, or the average transition value, and reported at varying aggregate levels (e.g., class, school, district, state).

Table 5
Within-Year Categorical Gains by Fall and Spring Linkage Level Mastery Status

		Spring		
		Non-Master	Master	Total
Fall	Non-Master	300	400	700
	Master	50	250	300
	Total	350	650	1000

While overall performance in the subject can be used to calculate a student's performance level in fall and spring, reporting within-year, performance-level categorical gains does not make conceptual sense given that performance levels have been defined as a summative indicator of

[6] ELA and mathematics = 0, 1, 2, 3, 4, or 5 levels mastered; science = 0, 1, 2, or 3 levels mastered.

student achievement following a full year of instruction. Table 6 summarizes the assumptions for reporting within-year categorical gains at the EE or linkage level.

Table 6

Assumptions for Within-Year Essential Element or Linkage-Level Categorical Growth Models

Categorical gains	DLM assessment
Repeated data collection Time between administrations	EE/linkage-level assessment administration variable across students and windows leading to variable reporting Three ways to demonstrate linkage-level-mastery

Discussion

Synopsis of Models

Because of the highly variable population of students taking DLM assessments, the complexity of the two assessment models, and the use of diagnostic scoring, growth has not been reported for DLM assessments to date. Table 7 summarizes the key considerations for each model discussed in this white paper and the DLM assessment characteristics that do or do not align with them. Any future reporting of growth for DLM assessments must minimize the number of statistical and theoretical violations, including the absences of a vertical scale, an interval scale, and a suitable comparison group; unstable sample sizes; and the population's tendency to achieve finer-grained academic gains.

Table 7

Summary of Model Requirements and DLM Characteristics

Data requirements, considerations, or assumptions	**Simple gain scores**	**Residual gain scores**	**Categorical**	**SGPs**	**DLM property**
Vertical scale	Yes	No	Implicit	No	No vertical scaling
Interval scale	Yes	Yes	No	Yes	No interval scales
Comparison group	No	No	No	Yes	No comparison group
Large (i.e., >5,000) sample sizes	No	No	No	Yes	IM states and some grades have very small sample sizes
Sensitive to cut points	No	No	Yes	No	Cut points used to determine performance levels separately
Interpretation	How much has the student's or cohort's achievement improved? Is the student or cohort making progress toward a proficiency target?	Has the student or cohort performed at or beyond expectations based on previous achievement?	Has the student or cohort transitioned from one performance level to another?	What is the percentile rank of the student relative to students or cohorts with similar past scores? What is the median or average SGP?	—

Note. IM = integrated model; SGP = student growth percentile. Table adapted from Farley, Saven, Tindal, & Nese (2013)

Technical Advisory Committee Recommendations

The DLM TAC has provided advice regarding the calculation of growth for DLM assessments across several meetings. As the committee noted in August and October 2016 and in May 2018, norm-referenced growth measures pose many challenges, particularly for AA-AAS like DLM assessments. For these reasons, the TAC recommended against reporting SGP values for DLM assessments. The committee emphasized that, for DLM assessments, the priority for reporting growth should be to provide teachers with an instructionally useful metric for evaluating student progress over time. The committee recommended prioritizing map-based modeling research to support eventual reporting of growth at the node level.

In a subsequent meeting in February 2019, the TAC again indicated they did not support the use of DLM results for accountability-oriented growth evaluations. The TAC raised potential statistical and legal concerns about making high-stakes decisions based on use of DLM results beyond their supported purposes. The committee advised states to work within the boundaries of supported use for DLM results and the ESSA requirement that growth measures used for accountability purposes must be demonstrably valid and reliable. If states are obligated to use DLM results for accountability (e.g., due to a state legislative mandate), the TAC strongly advised that states take steps to minimize negative consequences so that students, teachers, schools, and districts are not penalized.

Future Directions

At this time, the DLM Consortium cannot make any recommendations for growth metrics that meet sufficient standards for reliability and validity and are consistent with supported uses of DLM results. Therefore, no growth metrics are calculated on behalf of the consortium at this time.

The DLM research agenda currently prioritizes efforts to pursue node-based data collection that may eventually support reporting growth at the node level. While this approach depends on modeling and test-development work, it allows for several options for evaluating student growth. We will also explore learning-progression-based methods, such as student learning objectives (Briggs et al., 2015) or using learning progression as the vertical scale for reporting growth (Yu, Kim, & Dunn, 2017).

The ultimate goal of reporting progress on DLM assessments is to provide information to stakeholders, especially teachers, to guide instructional decision-making. This requires a method for reporting change over time that is sensitive enough to effectively describe academic progress for students with SCD. Use of the instructionally embedded system as intended may support reporting of progress for within-student, within-classroom applications for growth through the school year. Stakeholders who are concerned with the academic progress of students who take AA-AAS may be interested in considering the opportunities described in the forthcoming follow-up paper.

References

Albus, D., & Thurlow, M. L. (2012). *Alternate assessments based on alternate achievement standards (AA-AAS) participation policies* (Synthesis Report 88). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved from https://nceo.umn.edu/docs/OnlinePubs/Synthesis88/SynthesisReport88.pdf

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4), 42–51. https://doi.org/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.* IDEA Data Center. Rockville, MD: Westat. Retrieved from ERIC database: https://files.eric.ed.gov/fulltext/ED584165.pdf

Briggs, D. C., Diaz-Bilello, E., Peck, F., Alzen, J., Chattergoon, R., & Johnson, R. (2015). *Using a learning progression framework to assess and evaluate student growth.* Boulder: University of Colorado, Center for Assessment Design Research and Evaluation. Retrieved from https://www.colorado.edu/education/sites/default/files/attached-files/CADRE.CFA-StudentGrowthReport-Final_0.pdf

Buzick, H. M., & Jones, N. D. (2015). Using test scores from students with disabilities in teacher evaluation. *Educational Measurement: Issues and Practice, 34*(3), 28–38. https://doi.org/10.1111/emip.12076

Buzick, H. M., & Laitusis, C. C. (2010). Using growth for accountability: Measurement challenges for students with disabilities and recommendations for research. *Educational Researcher, 39*, 537–544. https://doi.org/10.3102/0013189X10383560

Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models*. Washington, DC: Council of Chief State School Officers. Retrieved from https://scholar.harvard.edu/files/andrewho/files/a_practitioners_guide_to_growth_models.pdf

Clark, A., Nash, B., Karvonen, M., & Kingston, N. (2017). Condensed mastery profile method of setting standards for diagnostic assessment systems. *Educational Measurement: Issues and Practice, 36*(4). https://doi.org/10.1111/emip.12162

Council of Chief State School Officers. (2017). *Considerations for including growth in ESSA state accountability systems*. Washington, DC: Author. Retrieved from https://ccsso.org/resource-library/considerations-including-growth-essa-state-accountability-systems

Dynamic Learning Maps Consortium. (2018a). *2017–2018 Technical Manual Update – Integrated Model*. Lawrence: University of Kansas, Center for Educational Testing and Evaluation. Retrieved from https://dynamiclearningmaps.org/sites/default/files/documents/publication/2017-2018_IM_Technical_Manual_Update.pdf

Dynamic Learning Maps Consortium. (2018b). *2017–2018 Technical Manual Update – Year End Model*. Lawrence: University of Kansas, Center for Educational Testing and Evaluation. Retrieved from https://dynamiclearningmaps.org/sites/default/files/documents/publication/2017-2018_YE_Technical_Manual_Update.pdf

Domaleski, C., & Hall, E. (2016). *Guidance for estimating and evaluating academic growth*. Retrieved from National Center and State Collaborative website: http://www.ncscpartners.org/Media/Default/PDFs/Resources/EstimatingandEvaluatingStudentGrowth.pdf

Elliott, S. N. (2014). Measuring opportunity to learn and achievement growth: Key research issues with implications for the effective education of all students. *Remedial and Special Education, 34*(1), 58–64. doi: 10.1177/0741932514551282

Every Student Succeeds Act of 2015, P. L. 114-95 § 114 Stat. 1177 (2015–2016).

Farley, D., Saven, J. L., Tindal, G., & Nese, J. F. T. (2013). *Analysis of growth on state tests for students with significant cognitive disabilities*. Eugene, OR: Behavioral Research and Teaching. Retrieved from ERIC database: https://files.eric.ed.gov/fulltext/ED545271.pdf

Georgia Department of Education. (2017). A guide to the Georgia Student Growth Model. Retrieved from https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/GSGM/FY18/SGPGuide%2006052017.pdf

Grady, M., Lewis, D., & Gao, F. (2010, April–May). *The effect of sample size on student growth percentiles*. Paper presented at the annual meeting of the National Council on Measurement in Education, Denver, CO.

Guarino, C., Reckase, M., Stacy, B., & Wooldridge, J. (2015). A comparison of student growth percentile and value-added models of teacher performance. *Statistics and Public Policy, 2*(1), 1–11. https://doi.org/10.1080/2330443X.2015.1034820

Holdheide, L. R., Goe, L., Croft, A., & Reschly, D. J. (2010). *Challenges in evaluating special education teachers and English language learner specialists*. National Comprehensive Center for Teacher Quality. Retrieved from ERIC database: https://files.eric.ed.gov/fulltext/ED520726.pdf

Jackson, L. B., Ryndak, D. L., & Wehmeyer, M. L. (2008). The dynamic relationship between context, curriculum, and student learning: A case for inclusive education as a research-based practice. *Research and Practice for Persons with Severe Disabilities, 33–34*, 175–195. https://doi.org/10.2511/rpsd.33.4.175

Jones, N. D., Buzick, H. M., & Turkan, S. (2013). Including students with disabilities and English learners in measures of educator effectiveness. *Educational Researcher, 42*, 234–241. doi:10.3102/0013189X12468211

Lockwood, J. R., & Castellano, K. E. (2015). Alternative statistical frameworks for student growth percentile estimation. *Statistics and Public Policy, 2*(1), 1–9. Retrieved from ERIC database: https://files.eric.ed.gov/fulltext/ED562562.pdf

Nash, B., Clark, A. K., & Karvonen, M. (2015). *First contact: A census report on the characteristics of students eligible to take alternate assessments* (Technical Report No. 15-02). Lawrence: University of Kansas, Center for Educational Testing and Evaluation. Retrieved from https://dynamiclearningmaps.org/sites/default/files/documents/publication/First_Contact_Census_2016.pdf

New York State Education Department. (2018). *Measuring student growth for institutional accountability in New York*. Retrieved from http://www.p12.nysed.gov/accountability/documents/NYSEDMonographMeasuringStudentGrowthforInstitutionalAccountabilityinNewYork2018.pdf

No Child Left Behind Act of 2001, P. L. 107-110, 20 U.S.C. § 6301 *et seq.* (2002).

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. *Journal of Special Education*, *49*, 209–220. https://doi.org/10.1177/0022466915582213

Taub, D. A., McCord, J. A., & Ryndak, D. L. (2017). Opportunities to learn for students with extensive support needs: A context of research-supported practices for all in general education classes. *Journal of Special Education, 51*, 127–137. https://doi.org/10.1177/0022466917696263

Towles-Reeves, E., Kearns, J., Kleinert, H., & Kleinert, J. (2009). An analysis of the learning characteristics of students taking alternate assessments based on alternate achievement standards. *Journal of Special Education, 42*, 241–254. https://doi.org/10.1177/0022466907313451

Yu, L., Kim, W., & Dunn, J. (2017, April). *Establishing learning progression-based vertical scales to measure growth.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Antonio, TX.

Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement of students with disabilities, ages 7 to 17. *Exceptional Children, 78*, 89–106. https://doi.org/10.1177/001440291107800106

Welch, C.J., Dunbar, S.B., & Rickels, H. (2016). Measuring student growth with the Iowa assessments. https://itp.education.uiowa.edu/ia/documents/Measuring-Student-Growth-in-Iowa-with-the-Iowa-Assessments.pdf

Appendix A
State Testing Requirements

State	Model	High school grades tested (Mathematics/ELA) 9	10	11	12*	Science grades tested
Alaska	YE	Required	No	No	No content available	4, 8, 10
Colorado	YE	Required	Required	Required		N/A
Delaware	YE	No	Required	Required		5, 8, 10 (BIO)
Illinois	YE	Required	Required	Required		5, 8, 11
Maryland	YE	No	No	No		5, 8, 11 (occasional 12)
Miccosukee	YE	No	Required	No		5, 8, 10
New Hampshire	YE	No	No	Required		5, 8, 11
New Jersey	YE	No	No	Required		5, 8, 11
New York	YE	Required	No	No		4, 8, 9
Oklahoma	YE	No	No	Required		5, 8, 11
Rhode Island	YE	No	No	Required		5, 8, 11
Utah	YE	Required	Required	Required		N/A
West Virginia	YE	No	No	Required		5, 8, 11
Wisconsin	YE	Required	Required	Required		4, 8–11
Arkansas	IM	Required	Required	Required	*	3–10
Iowa	IM	Optional	Required	Required	*	3–8, 10 & 11; 9 optional
Kansas	IM	No	Required	No	*	5, 8, 11
Missouri	IM	Optional	Optional	Required	*	5, 8, 11
North Dakota	IM	Optional	Required	Optional	*	N/A

Note. IM = integrated model; YE = year-end model.
* In grade 12, only ELA content is available.

Appendix B

Example Integrated-Model Blueprint Requirements for Grade 3 ELA

Grade 3: Available Essential Elements and minimum expectation for each student's assessment

Conceptual Area	EE	DESCRIPTION
ELA.C1.1	**Choose at least three EEs, including at least one RL and one RI.**	
	EE.RL.3.1	Answer who and what questions to demonstrate understanding of details in a text.
	EE.RL.3.2	Associate details with events in stories from diverse cultures.
	EE.RL.3.3	Identify the feelings of characters in a story.
	EE.RL.3.5	Determine the beginning, middle, and end of a familiar story with a logical order.
	EE.RI.3.1	Answer who and what questions to demonstrate understanding of details in a text.
	EE.RI.3.2	Identify details in a text.
	EE.RI.3.3	Order two events from a text as "first" and "next".
	EE.RI.3.5	With guidance and support, use text features including headings and key words to locate information in a text.
ELA.C1.2	**Choose two EEs in C1.2 (L, RL or RI) – EEs must be from different strands, i.e. RL and L, not RL and RL.**	
	EE.RL.3.4	Determine words and phrases that complete literal sentences in a text.
	EE.RI.3.4	Determine words and phrases that complete literal sentences in a text.
	EE.RI.3.8	Identify two related points the author makes in an informational text.
	EE.L.3.5.a	Determine the literal meaning of words and phrases in context.
	EE.L.3.5.c	Identify words that describe personal emotional states.
ELA.C1.3	**Choose at least one EE (RL or RI).**	
	EE.RL.3.9	Identify common elements in two stories in a series.
	EE.RI.3.9	Identify similarities between two texts on the same topic.
ELA.C2.1	**All students are assessed in both of these EEs through the writing assessment. In ITI, choose one Conventional EE or one Emergent EE. See Writing Testlet FAQ for more detail.**	
	EE.W.3.2.a	Select a topic and write about it including one fact or detail.
	EE.W.3.4	With guidance and support produce writing that expresses more than one idea.