Interpreting Longitudinal Growth in Alternate Assessments Through Complementary Evidence

Brock Rowley

Gerald Tindal

Behavioral Research and Teaching

University of Oregon

Prepared for the Oregon Department of Education

July 31, 2026

Table of Contents

Abstract

Longitudinal growth is central to educational assessment, yet no single indicator fully characterizes change. This study evaluated whether growth in alternate assessments can be interpreted more defensibly by integrating complementary evidence. Longitudinal Oregon Extended Assessment records from 2014-15 through 2025-26 were ana-lyzed for vertically scaled Grades 3-8 English Language Arts and Mathematics. Analyses considered continuous scale-score movement, achievement-level transitions, progress toward proficiency, reliable-growth classifications, and multi-year longitudinal sequences. Three conditions organized interpretation: reliable measurement addressed whether observed change exceeded expected measurement uncertainty; developmental context located change across subjects, grades, starting achievement levels, and student records; and longitudinal continuity described whether longitudinal patterns showed temporal consistency or changes in direction over time. Meaningful growth was ob-servable across the study period but varied systematically across students and contexts. Complementary indicators often converged while contributing distinct information, and their divergence identified questions that a single mea-sure would have obscured. Reliable-growth classifications separated changes exceeding expected uncertainty from those consistent with measurement variability, while multi-year analyses revealed frequent changes in direction and showed that uninterrupted growth patterns were comparatively uncommon. The strongest inferences emerged when distinct indicators converged and the conditions of change were explicit. The study contributes an interpretive frame-work for alternate-assessment growth reporting centered on the convergence of complementary evidence under ex-plicit conditions of reliable measurement, developmental context, and longitudinal continuity.

1. Introduction

Growth has become a central objective of contemporary educational systems and an increasingly important component of educational assessment, accountability, and instructional decision-making (Betebenner & Linn, 2009; Castellano & Ho, 2013). Unlike achievement status, which summarizes student performance at a single point in time, growth characterizes change in student performance across time. Contemporary accountability systems increasingly distinguish among growth models, growth measures, and growth indicators because each contributes distinct information to educational decision-making (Domaleski & Hall, 2016; Winter et al., 2022).

Despite the importance of growth, its interpretation remains challenging. Multiple indicators of student progress are commonly available, including scale-score movement, achievement-level movement, movement relative to performance standards, and measures of score precision. These indicators often provide related but not identical evidence about student performance (Domaleski & Hall, 2016; Winter et al., 2022). Consequently, growth interpretation extends beyond selecting a single statistical model or indicator; it requires synthesizing multiple sources of evidence into coherent, technically defensible interpretations of student progress across diverse educational contexts.

Research on alternate assessments has likewise advanced substantially over the past two decades. Previous investigations have documented alternate-assessment development and validation (Tindal et al., 2003), participation patterns (Saven et al., 2016), longitudinal achievement (Tindal et al., 2016), growth modeling (Farley et al., 2017; Nese et al., 2017), and transition-matrix approaches (Tindal et al., 2017). Collectively, this body of work demonstrates that growth can be measured, modeled, and reported for students with the most significant cognitive disabilities. However, previous research has primarily described growth through individual analytic approaches or specific indicators.

Comparatively less attention has been devoted to evaluating whether conclusions about growth remain consistent across multiple complementary sources of evidence and the educational contexts in which those indicators are interpreted. This investigation builds directly upon that established body of research. Rather than proposing another growth metric or advocating a particular analytic model, the study evaluates whether multiple complementary sources of evidence can be synthesized into a coherent and technically defensible framework for interpreting longitudinal student growth. The framework integrates evidence from scale-score movement, achievement-level change, progress toward proficiency, reliable-growth classifications, and repeated observations across time.

Within this framework, three conditions organize the interpretation of growth: **reliable measurement**, **developmental context**, and **longitudinal continuity**. Reliable measurement concerns whether observed score change can be distinguished from expected measurement uncertainty and is evaluated through reliable-growth classifications. Developmental context situates growth within subject, grade, starting achievement level, and the student's sequence of observations. Longitudinal continuity describes how the direction and pattern of growth unfold across repeated observations and is examined through multi-year sequences and the persistence of direction across successive periods. Temporal consistency is evident when direction is sustained within those sequences, whereas changes in direction reveal interruption, recovery, or other variation across time.

These conditions do not replace the analytic indicators used in the study. Rather, they organize the interpretation of scale-score movement, achievement-level change, progress toward proficiency, reliable-growth classifications, longitudinal patterns, and persistence. Applied consistently throughout the manuscript, they provide a structure for evaluating where evidence converges, where it diverges, and how measurement, developmental, and temporal conditions influence conclusions about longitudinal student development.

The investigation was conducted within the broader context of Oregon's ongoing evaluation of statewide growth reporting. As the Oregon Department of Education evaluated alternative frameworks for interpreting longitudinal growth within the general assessment system, parallel questions emerged regarding the interpretation of growth within the Oregon Extended Assessment. This investigation addresses that complementary need by evaluating an evidence-informed interpretive framework that synthesizes multiple complementary sources of evidence for longitudinal growth within Oregon's alternate assessment system (Oregon Department of Education, 2024).

To evaluate this framework, longitudinal Oregon Extended Assessment (ORExt) records spanning school years 2014-15 through 2025-26 were analyzed using multiple complementary indicators of student growth. The analy-ses examine where evidence converges, where it diverges, and how growth varies across adjacent-year transitions, longer longitudinal sequences, achievement levels, grades, subjects, and changing assessment contexts. Together, these analyses evaluate the central claim that meaningful growth is observable for students with the most significant cognitive disabilities, while demonstrating that its interpretation depends upon the convergence of complementary evidence considered within explicit conditions of reliable measurement, developmental context, and longitudinal continuity.

2. Assessment Context and Measurement Considerations

2.1 Oregon Growth Context and ORExt Positioning

Because longitudinal growth is interpreted within the context of the assessment system that produces it, understanding the characteristics of the Oregon Extended Assessment (ORExt) provides an essential interpretive warrant for the analyses that follow. Assessment design, participation policies, achievement standards, and measurement characteristics collectively define the evidentiary context from which growth inferences are drawn (American Educational Research Association et al., 2014; Winter et al., 2022). Interpreting growth within alternate assessments presents additional challenges. Students participating in alternate assessments represent a heterogeneous population with diverse instructional experiences, developmental profiles, communication characteristics, and achievement characteristics (Buzick & Laitusis, 2010; Winter et al., 2022). Consequently, responsible interpretation of longitudinal growth extends beyond observed score change to include the educational and measurement contexts in which that change occurs. These contextual factors provide essential interpretive warrants for evaluating whether evidence of growth converges across complementary indicators.

The Oregon Extended Assessment (ORExt) provides a distinctive setting for evaluating this interpretive framework. Since the 2014-15 redesign, ORExt has reported student performance on a vertically scaled RIT metric intended to support interpretation across grades and years. The availability of longitudinal student records spanning more than a decade creates a unique opportunity to evaluate multiple complementary sources of evidence within a common developmental measurement framework. This continuity establishes the evidentiary foundation for examining where indicators of growth converge, where they diverge, and how context influences their interpretation.

Although ORExt maintained the core measurement properties necessary to support longitudinal interpretation, important policy, operational, and assessment-design changes occurred during the study period. These include establishment of the current cut-score structure during the 2015 redesign and standard-setting process, subsequent changes in form length and operational item counts, evolving eligibility guidance affecting alternate-assessment participation, minimum-participation rules governing test completion, and disruptions associated with the COVID-19 pandemic. These developments do not preclude longitudinal analysis, but they define the interpretive conditions under which growth should be evaluated. They provide essential context for understanding changes in score distributions, participation patterns, measurement precision, achievement-level movement, and the durability of growth across successive years.

2.2 Scale-Score and Cut-Score Interpretation

Interpretation of longitudinal growth depends upon the measurement properties of the underlying assessment. Within ORExt, scale scores provide the primary continuous measure of student performance and establish the principal evidentiary basis for evaluating developmental change (Rowley et al., 2019). Because the vertically scaled RIT metric places student achievement on a common developmental continuum, student performance can be interpreted across grades and years within a consistent measurement framework rather than through independent grade-specific score scales. Observed scale-score movement therefore provides one important source of evidence regarding longitudinal growth. However, the magnitude of score change alone does not necessarily establish its educational significance. Similar score changes may support different interpretations depending on a student's initial achievement level, proximity to performance thresholds, and the measurement characteristics of the assessment (Betebenner & Linn, 2009; Domaleski & Hall, 2016). Consequently, interpretation of longitudinal growth requires consideration of both observed score change and the measurement context in which that change occurred.

Performance-level classifications provide a complementary source of evidence by locating student performance relative to established achievement standards. Unlike continuous scale scores, performance levels are policy-based reporting categories established through formal standard-setting procedures and associated cut scores.

For ORExt, the current achievement-level descriptors and cut scores were established through Oregon's statewide standard-setting process conducted in June 2015 (Oregon Department of Education, 2015). Movement across performance-level boundaries may therefore represent meaningful educational change, whereas stability within a performance level should not be interpreted as evidence that meaningful developmental growth has not occurred. The 2015 standard-setting process also established grade-specific interpretive conditions that warrant consideration. In Mathematics, several grades contain relatively narrow Level 2 score bands, creating situations in which comparatively small scale-score changes may coincide with movement across performance-level boundaries. These characteristics illustrate why performance-level movement should be interpreted as complementary to, rather than a substitute for, continuous scale-score evidence. Considered together, the two indicators provide a more complete and educationally meaningful interpretation of longitudinal growth than either indicator can provide independently.

2.3 Form-Length, Participation, and Measurement Context

Assessment design and participation policies establish additional contextual conditions for interpreting longitudinal growth. Changes in assessment design influence measurement precision, blueprint representation, construct coverage, and the interpretation of score change, even when the underlying developmental scale remains stable (American Educational Research Association et al., 2014). Consequently, these characteristics should be considered when drawing longitudinal inferences.

During the study period, ORExt evolved from 48-item forms consisting of 36 operational and 12 field-test items to 36-item forms consisting of 30 operational and 6 field-test items. This transition preserved the assessment blueprint and vertically scaled measurement framework while reducing the number of operational and field-test items administered to students. Accordingly, the reduction in form length represents a change in assessment design rather than simply a shorter test and provides important context for interpreting longitudinal evidence (Rowley et al., 2025).

Participation policies likewise shape the evidentiary basis for longitudinal interpretation. ORExt participation is determined through IEP-team decisions guided by state eligibility criteria and federal alternate-assessment participation requirements. During the study period, Oregon implemented additional guidance intended to strengthen alternate-assessment participation decisions and improve alignment with federal participation expectations (Rowley et al., 2025). Consequently, changes in the composition of the tested population provide important context for evaluating longitudinal score distributions, linkage patterns, and growth outcomes.

Minimum-participation rules represent an additional component of this interpretive framework by defining when testing may be discontinued while still satisfying participation requirements. Although these policies influence the evidentiary structure of the longitudinal record, they do not diminish the value of the resulting evidence. Rather, they establish additional conditions that should be considered when interpreting observed growth.

Finally, the COVID-19 pandemic disrupted assessment participation, instructional continuity, and the temporal structure of longitudinal records. These disruptions altered the continuity and comparability of adjacent-year observations and therefore represent important contextual conditions for interpreting longitudinal evidence.

Taken together, vertical scaling, achievement-level structure, assessment design, participation policies, and changing measurement conditions define the context within which longitudinal growth should be interpreted. Together, these conditions support an interpretive framework in which evidence for student growth is evaluated through the convergence of complementary indicators across differing educational, developmental, and measurement contexts rather than through any single indicator in isolation.

3. Methods

3.1 Assessment Context

The Oregon Extended Assessment (ORExt) provides the assessment context for this investigation. ORExt is Oregon's alternate assessment aligned with alternate academic achievement standards (AA-AAAS) for students with the most significant cognitive disabilities. Participation is determined through Individualized Education Program (IEP) team decisions using Oregon's alternate-assessment participation criteria within the broader federal framework governing alternate-assessment participation (American Educational Research Association et al., 2014; Rowley et al., 2025).

Longitudinal analyses were restricted to vertically scaled Grades 3–8 English Language Arts and Mathematics because these content areas share a common vertically scaled RIT metric. This common scale provides the measurement foundation necessary to interpret student performance across grades and years within a continuous developmental framework rather than through independent grade-specific score scales. Science and high school assessments were excluded because they do not contribute evidence within this vertically scaled growth framework.

3.2 Analytic Dataset

The analytic dataset comprised longitudinal Oregon Extended Assessment (ORExt) records spanning school years 2014-15 through 2025-26. Analyses were restricted to vertically scaled Grades 3-8 English Language Arts and Mathematics to ensure that every longitudinal comparison was supported by a common developmental measurement framework. This restriction ensured that each analysis contributed evidence relevant to the central interpretive questions addressed in this study.

The final analytic dataset contained 60,820 assessment records representing 14,302 unique students. Available variables included student identifiers, school year, tested grade, enrolled grade, subject, vertically scaled RIT scores, performance levels, conditional standard errors of measurement (CSEMs), theta estimates, demographic characteristics, disability classifications, and school and district identifiers. Assessment records were distributed nearly evenly across subjects, including 30,644 English Language Arts assessments (50.4%) and 30,176 Mathematics assessments (49.6%). These records supported multiple complementary units of analysis, including annual observations, student-subject longitudinal records, adjacent-year comparisons, and longer uninterrupted sequences. Collectively, these structures provided converging evidence regarding developmental continuity across different time periods.

3.3 Measurement Framework

ORExt reports both continuous scale scores and achievement-level classifications. Continuous RIT scores provide the primary evidence for evaluating developmental change because they preserve both the magnitude and direction of student movement across years. Achievement-level classifications contribute complementary evidence by locating student performance relative to established performance standards. Because achievement levels intentionally summarize information contained within the continuous scale, they should not be interpreted as direct measures of developmental growth (Betebenner & Linn, 2009; Domaleski & Hall, 2016).

Measurement precision and reliable growth serve related but distinct functions. Measurement precision is the technical property represented by each score's conditional standard error of measurement (CSEM), whereas reliable growth is the interpretive classification assigned when observed change exceeds the uncertainty associated with the two annual scores. Longitudinal inferences therefore require consideration of both observed score change and the precision with which that change was estimated. This distinction follows directly from established measurement theory and provides the technical foundation for the reliable-growth classifications used throughout this manuscript (American Educational Research Association et al., 2014).

Collectively, these measurement properties establish the interpretive framework adopted in this study. Stable achievement-level classifications should not automatically be interpreted as evidence that developmental growth has not occurred. Students may demonstrate positive scale-score movement without crossing an achievement-level boundary (Betebenner & Linn, 2009; Winter et al., 2022). Accordingly, longitudinal growth was evaluated through converging evidence from vertically scaled RIT scores, achievement-level classifications, measurement precision, and repeated observations across time rather than through any single indicator considered in isolation.

3.4 Longitudinal Architecture

Longitudinal growth cannot be inferred from isolated annual assessment results. Rather, repeated observations of the same student provide the evidence necessary to evaluate developmental change across time. Accordingly, the evidentiary structure of this study extends beyond individual assessment events to include adjacent-year comparisons and multi-year longitudinal sequences. Each successive structure contributes distinct evidence regarding student growth while strengthening confidence in developmental inferences across increasingly longer time spans.

Because every annual assessment is reported on a common vertically scaled developmental scale, observations can be linked across school years to characterize developmental continuity and change. Incorporating measurement precision further strengthens these longitudinal inferences by allowing repeated observations to be interpreted as evidence of reliable change, developmental continuity, or stability rather than simply as a sequence of independent annual scores.

The final analytic dataset included 14,302 unique students contributing 31,312 student-year observations. Because growth was evaluated separately for English Language Arts and Mathematics, the principal longitudinal unit of analysis was the student-subject record, resulting in 27,450 unique longitudinal records available for analysis.

Table 1

Longitudinal Analytic Evidence Supporting Reliable-Growth Analyses

Longitudinal evidence	Analytic unit	N	Primary purpose
Students	Student	14,302	Describe the study population
Student-year observations	Student × Year	31,312	Describe annual participation
Student-subject records	Student × Subject	27,450	Define longitudinal analytic units
Adjacent-year pairs	Student × Subject × 2 years	29,624	Estimate annual growth
Three-year sequences	Student × Subject × 3 years	13,840	Characterize short-term developmental patterns
Four-year sequences	Student × Subject × 4 years	5,543	Characterize intermediate developmental continuity
Five-year sequences	Student × Subject × 5 years	1,544	Characterize long-term developmental continuity

Table 1 illustrates how the evidentiary base becomes progressively more selective as additional longitudinal requirements are imposed. The broader adjacent-year reporting dataset contained 31,234 student-subject growth events. This total included 29,624 standard adjacent instructional-year pairs and 1,610 comparisons spanning the COVID interruption from 2018-19 to 2021-22. The broader 31,234-event dataset was used for historical descriptive summaries because it preserves all observed score-change events and provides context for the interruption period. Analyses intended to characterize standard annual growth, recurring adjacent-year patterns, and longer developmental sequences were based on the 29,624 standard adjacent instructional-year pairs.

The 29,624 linked pairs therefore provided the principal foundation for annual longitudinal analyses within the manuscript. Requiring three consecutive annual observations yielded 13,840 three-year sequences, whereas four consecutive observations yielded 5,543 four-year sequences. Only 1,544 complete five-year student-subject sequences were available across the uninterrupted pre-COVID (2014-15 through 2018-19) and post-COVID (2021-22 through 2025-26) assessment periods. This hierarchical longitudinal structure allows the complete reporting dataset to describe the full historical record while reserving the standard adjacent-year cohort for analyses requiring comparable annual instructional intervals and uninterrupted developmental sequences.

3.5 Growth Indicators

Educational measurement research has demonstrated that different growth indicators emphasize different aspects of student performance and therefore support different interpretive claims regarding longitudinal growth (Betebenner & Linn, 2009; Castellano & Ho, 2013; Domaleski & Hall, 2016). Rather than identifying a single preferred indicator, responsible interpretation requires evaluating how complementary indicators contribute evidence under differing educational and measurement contexts.

Accordingly, growth was evaluated using multiple complementary indicators. Continuous scale-score change quantified movement on the vertically scaled RIT score, whereas performance-level transitions summarized movement across policy-defined achievement categories. Additional indicators evaluated student position relative to proficiency standards through distance-to-standard and cut-score proximity measures. Reliable-growth classifications incorporated conditional standard errors of measurement (CSEMs) to distinguish observed score differences that exceeded expected measurement uncertainty from those that did not (American Educational Research Association et al., 2014; Rowley et al., 2025).

Collectively, these indicators provide complementary evidence regarding longitudinal student development. Their value lies not in competing to identify a single preferred measure of growth, but in demonstrating where evidence converges, where it diverges, and how differing educational, developmental, and measurement contexts influence interpretation. Throughout this manuscript, conclusions regarding student growth are therefore based on the convergence of complementary evidence rather than on any single indicator considered in isolation.

3.6 Interpretive Framework for Longitudinal Growth

Educational measurement standards and the broader growth literature recognize that longitudinal interpretations are strengthened when multiple defensible sources of evidence are considered together rather than in isolation (American Educational Research Association et al., 2014; Betebenner & Linn, 2009; Castellano & Ho, 2013). Consistent with this guidance, the analytic strategy was designed to evaluate the central interpretive claim of this manuscript: that meaningful longitudinal growth is interpreted more defensibly through the convergence of complementary sources of evidence than through reliance on any single growth indicator. Rather than extending the literature through another growth model or statistical approach, this study evaluates how multiple forms of evidence converge under differing measurement, developmental, and temporal conditions to support longitudinal interpretations of student growth (Betebenner & Linn, 2009; Castellano & Ho, 2013; Domaleski & Hall, 2016).

Accordingly, continuous scale-score movement, achievement-level transitions, progress toward proficiency, reliable-growth classifications, longitudinal patterns, and persistence are treated as complementary sources of evidence regarding student development. None is interpreted as sufficient evidence in isolation.

Reliable measurement concerns whether observed score change can be distinguished from expected measurement uncertainty and is evaluated through reliable-growth classifications. For each adjacent-year score difference, measurement uncertainty was calculated by combining the prior- and current-year CSEMs:

$$SE_{\text{difference}} = \sqrt{CSEM^2_{\text{prior}} + CSEM^2_{\text{current}}}.$$

A 95% threshold for reliable change was then calculated as:

$$1.96 \times SE_{\text{difference}}.$$

Each adjacent-year transition was assigned to one of three classifications. A reliable increase occurred when the observed scale-score change exceeded the positive threshold. A reliable decrease occurred when the change exceeded the threshold in the negative direction. Changes that remained within the positive and negative thresholds were classified as not distinguishable from measurement uncertainty.

Developmental context situates the resulting evidence within subject, grade, starting achievement level, and the student's sequence of observations. Longitudinal continuity describes how growth unfolds across repeated observations and is evaluated through longitudinal patterns and the persistence of direction across successive periods. Temporal consistency is evident when direction is sustained, whereas interrupted or redirected patterns indicate recovery, fragility, or other variation across time.

The methodological contribution of this study is therefore interpretive rather than computational. Rather than proposing another growth statistic, this investigation evaluates a framework grounded in the convergence of complementary evidence under three explicit conditions: reliable measurement, developmental context, and longitudinal continuity. These conditions do not replace the analytic indicators used in the study. Instead, they organize the interpretation of where those indicators converge, where they diverge, and what their combined evidence supports regarding longitudinal student development. Together, they shift the emphasis from identifying a preferred measure of growth to evaluating how distinct sources of evidence strengthen confidence in longitudinal inference.

3.7 Analytic Strategy

Analyses proceeded in three stages. First, the assessment context, measurement framework, and longitudinal architecture established the conditions under which growth could be interpreted. Second, adjacent-year and multi-year analyses examined the observability, distribution, and reliability of growth across developmental contexts. Third, the indicators were synthesized to determine where their conclusions converged, where they diverged, and what each pattern supported. This sequence allowed each indicator to answer a distinct question while evaluating the coherence of the evidence across contexts and time periods.

Student and program context analyses were conducted as descriptive representativeness checks across categories of gender, economic disadvantage, English learner status, race/ethnicity, and primary disability. The continuation analysis was based on prior-year assessment records in Grades 3-7 for which a subsequent tested grade was expected within the vertically scaled English Language Arts and Mathematics sequence. A record was classified as continuing when the same student-subject record was linked to an assessment in the next adjacent instructional year. Grade 8 records were excluded because Grade 8 was the terminal grade in the vertically scaled sequence.

For each characteristic and subject, the continuation rate represented the percentage of eligible prior-year records linked to the next adjacent instructional year. Continuation-rate variation was summarized using the spread between category-specific rates within each characteristic and subject. Missing subgroup values were retained and reported as "Missing / not provided." Category-specific cell sizes were examined, and results for sparsely represented categories, particularly within primary disability, were interpreted cautiously.

These context analyses described representation within the adjacent-year growth frame. They were not used to estimate causal effects of student characteristics on longitudinal growth.

4. Results

4.1 Growth Is Observable

Meaningful growth was observable across the ORExt longitudinal record, but it was not uniform. Its magnitude, direction, and interpretation varied by year pair, grade, subject, starting achievement level, and developmental context. Across the historical series, three conclusions recurred: growth was observable; its distribution was systematic; and no single indicator was sufficient. Confidence increased when scale-score movement, achievement-level movement, progress toward proficiency, and reliable-growth classifications converged. Across the adjacent-year reporting dataset, 28.8% of transitions were classified as reliable increases, 17.9% were classified as reliable decreases, and 53.3% were not distinguishable from measurement uncertainty. These findings demonstrate that meaningful positive and negative movement was observable across the longitudinal record, while also showing that more than one-half of observed transitions did not exceed the uncertainty associated with the scores.

Figure 1 summarizes the historical distribution of reliable adjacent-year growth observed across the ORExt longitudinal record. By placing the 2024–25 to 2025–26 administration within the full longitudinal record, the figure establishes the historical context against which subsequent findings are interpreted.

Figure 1
Reliable Growth Across Adjacent-Year Cohorts

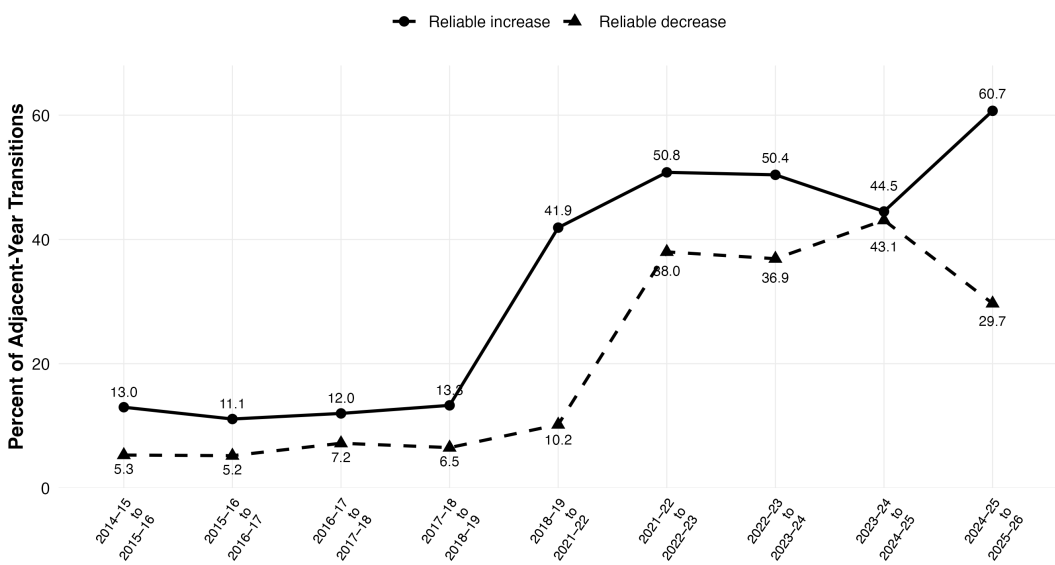


Figure 1 demonstrates that the 2024-25 to 2025-26 administration produced the strongest growth signal among the standard adjacent instructional-year comparisons in the available longitudinal record. This pattern was supported across complementary indicators, including scale-score movement, achievement-level movement, progress toward proficiency, and reliable-growth classifications. The larger descriptive scale-score gains observed from 2018-19 to 2021-22 reflect the extended interval created by the COVID interruption and are not directly comparable with standard annual transitions.

The elevated 2024-25 to 2025-26 growth signal was neither uniform nor attributable to a single indicator. Multiple analyses demonstrated that growth was concentrated within specific grade-subject combinations and among students entering the assessment at lower achievement levels. Accordingly, the strongest standard annual comparison in the longitudinal record is best interpreted as a comparatively strong but uneven period of annual growth rather than a system-wide shift of equal magnitude across all educational and developmental contexts.

Within the post-COVID period, the 2024-25 to 2025-26 cohort produced the strongest reliable-growth signal. The reliable-increase rate was 60.7%, compared with an average of 48.6% across the three preceding post-COVID annual comparisons. The elevated growth signal was observed across both English Language Arts and Mathematics while remaining concentrated within a relatively small number of grade-subject combinations. Across the broader longitudinal record, related analyses show that growth was broadly distributed across the assessment system but was comparatively strongest within several cross-indicator grade-subject contexts rather than uniformly expressed across all grades and subjects.

The post-COVID longitudinal record further strengthens this conclusion. Although the 2023-24 to 2024-25 year pair represented the weakest descriptive period within the recent longitudinal record, particularly in Mathematics, the subsequent 2024-25 to 2025-26 administration demonstrated substantially stronger scale-score gains and more pronounced upward movement across performance levels in both subjects. The pattern therefore supports interpreting recent post-COVID performance as strengthening within its developmental context rather than as a continuation of the immediately preceding year's results.

The observable growth signal varied across educational, developmental, and measurement contexts, requiring closer examination of where growth occurred and whether complementary indicators converged.

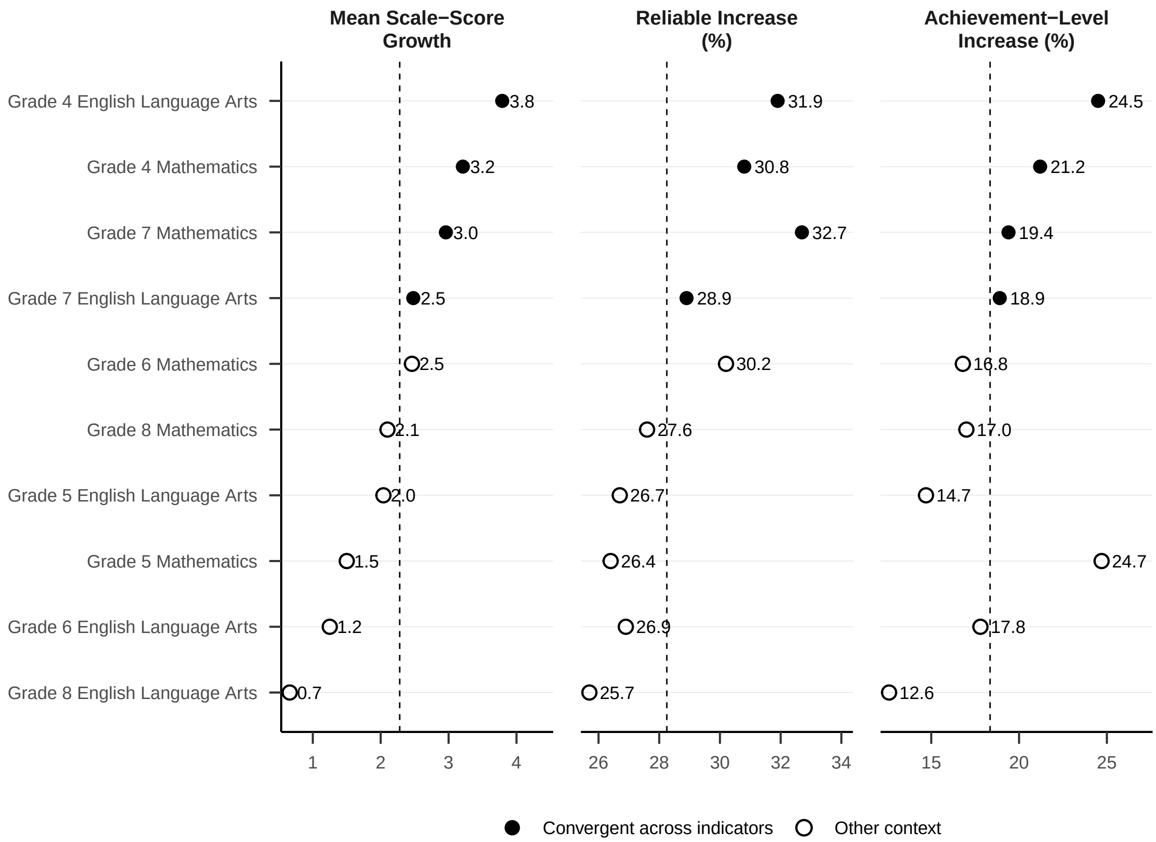
4.2 Cross-Indicator Convergence and Developmental Variation

Growth varied systematically rather than uniformly across grade-subject contexts. Figure 2 displays three complementary indicators for each context: mean scale-score growth, the percentage of reliable increases, and the percentage of achievement-level increases. The figure summarizes historical adjacent-year results across the longitudinal record and does not combine the indicators into a composite score.

A grade-subject context was classified as cross-indicator convergent when it ranked in the upper half of the ten contexts on all three indicators. This transparent descriptive criterion identifies contexts that were comparatively strong across multiple indicators rather than exceptional on only one measure.

Figure 2

Historical Convergence of Growth Indicators Across Grade-Subject Contexts



Note. Filled points identify grade-subject contexts that ranked in the upper half on all three indicators. Open points identify the remaining contexts. Dashed vertical lines mark the indicator-specific division between the upper and lower halves. Growth events were assigned to the student's current tested grade.

Across the three indicators, Grade 4 English Language Arts, Grade 4 Mathematics, Grade 7 English Language Arts, and Grade 7 Mathematics ranked in the upper half of the ten grade-subject contexts. The criterion required comparatively strong results across all three indicators; leading on a single indicator was insufficient. For example, Grade 6 Mathematics had a comparatively high reliable-increase rate but weaker achievement-level movement, whereas Grade 5 Mathematics had comparatively high achievement-level movement but weaker scale-score growth and reliable-increase results. These contrasts illustrate why the indicators are complementary rather than interchangeable.

Growth events were assigned to the student's current tested grade. Accordingly, Grade 4 contexts represent Grade 3-to-4 transitions, and Grade 7 contexts represent Grade 6-to-7 transitions. The cross-indicator pattern should therefore be interpreted as comparatively strong change across those grade transitions, not as a standalone property of Grade 4 or Grade 7.

Complementary analyses also demonstrated that growth varied by subject, starting achievement level, and adjacent-year transition. Patterns across successive observations were more frequently variable than consistently increasing or decreasing, indicating that longitudinal growth unfolded dynamically rather than uniformly across students and educational contexts.

Analyses of persistence extended these findings across longer longitudinal sequences. The observed patterns included recovery, defined as decline followed by improvement; fragility, defined as improvement followed by decline; stability with little directional change; and repeated shifts between increases and decreases. Confidence in interpretation increased when complementary evidence converged across developmental contexts and time periods.

4.3 Student and Program Context Findings

Student and program characteristics provided descriptive context for the longitudinal sample but were not interpreted as explanations for observed growth. Differences in continuation rates across gender categories were minimal, with spreads of 0.74 percentage points in English Language Arts and 0.04 percentage points in Mathematics. The corresponding spreads across economic disadvantage categories were 6.05 and 5.24 percentage points, respectively.

English learner status produced continuation-rate spreads of 1.36 percentage points in English Language Arts and 5.28 percentage points in Mathematics. Race/ethnicity showed greater descriptive variation, with spreads of 13.59 percentage points in English Language Arts and 14.20 percentage points in Mathematics. Primary disability showed the largest variation among the available contextual characteristics, with spreads of 50.27 percentage points in English Language Arts and 45.81 percentage points in Mathematics.

These findings indicate that continuation rates differed across some student characteristics, particularly primary disability category. They do not establish that those characteristics caused the longitudinal growth patterns reported elsewhere in the study. Interpretation of growth therefore remained grounded in students' starting achievement levels, grade-subject contexts, adjacent-year transitions, and reliable-growth classifications.

Disability-related findings warrant particular caution because the categories represent heterogeneous groups of students with different instructional histories, educational experiences, and sample sizes. The comparatively large spreads may therefore reflect differences in category composition, cell size, or other unmeasured characteristics. They should be interpreted as descriptive features of the assessed population rather than as evidence that disability category explains why growth occurred.

4.4 Composition and Contribution

Cohort composition shaped where growth originated, but it did not explain the magnitude of the statewide signal. The distribution of students across achievement levels changed substantially over time, and those shifts affected each level's contribution to observed growth. Contribution and explanation, however, are not interchangeable.

Across the longitudinal record, the proportion of students beginning at Performance Level 1 increased from approximately 29% in the earliest year pairs to more than 50% during the 2024-25 to 2025-26 administration, whereas the proportion beginning at Performance Level 4 declined from approximately 18% to less than 3%. These shifts demonstrate that the achievement-level composition of linked cohorts changed substantially over time and therefore represent an important contextual condition for interpreting longitudinal growth.

Figure 3 distinguishes changes in cohort composition from each achievement level's contribution to the observed longitudinal growth signal. By separating who comprised the linked cohort from where growth originated, the figure provides the context needed to evaluate whether changes in overall growth reflect shifts in student composition, differential contributions across achievement levels, or both.

Figure 3 demonstrates that students beginning at Performance Level 1 generated a disproportionately large share of the observed growth signal relative to their representation within the linked cohort. Accordingly, changes in cohort composition help explain where observed growth originated but do not fully explain why some periods produced substantially stronger overall growth than others.

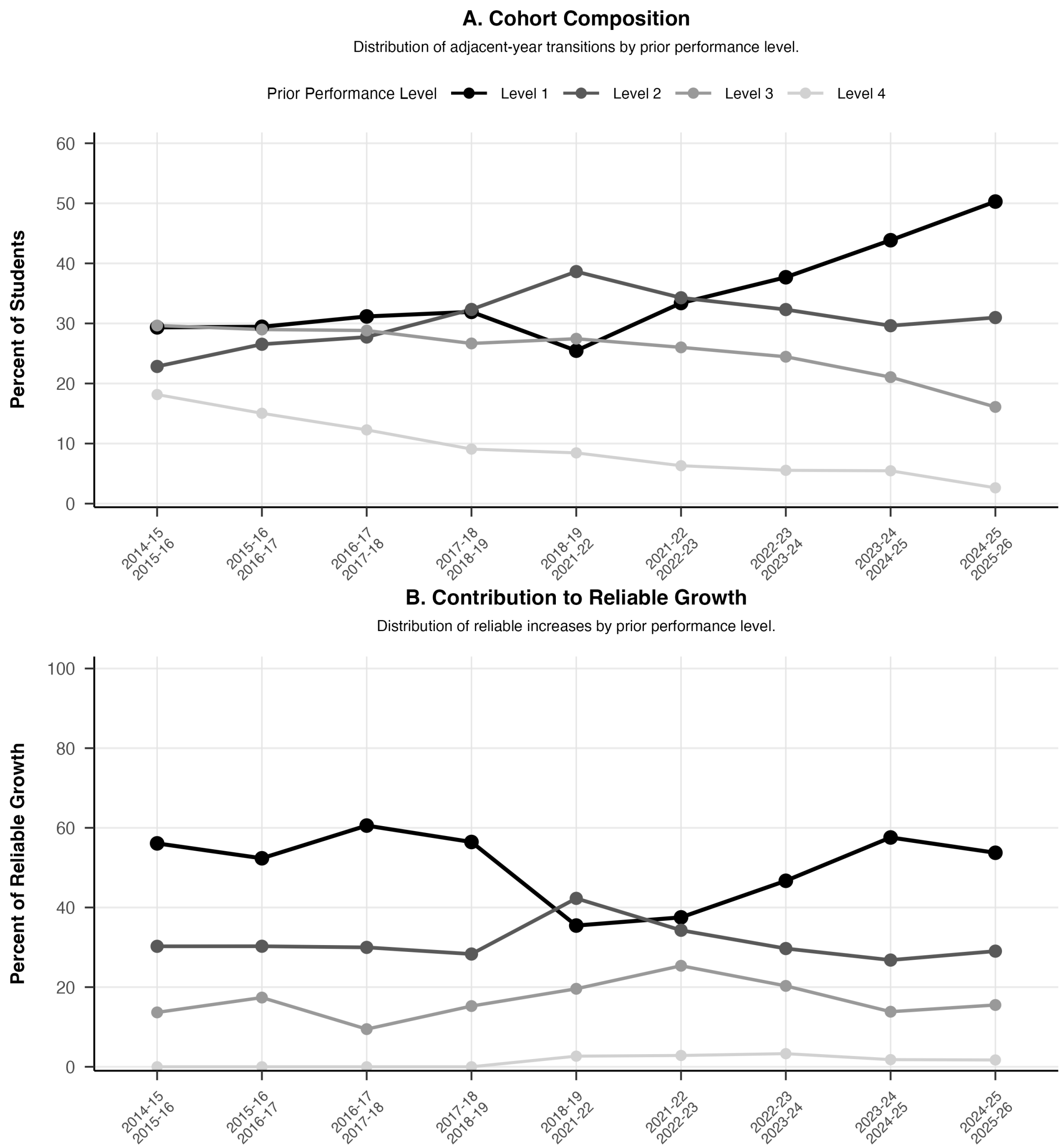
Additional analyses reinforce this interpretation. Performance Level 1 composition was strongly associated with the proportion of total growth generated by students beginning at Performance Level 1 ($r = .972$; $R^2 = .944$), yet exhibited only a comparatively weak relationship with overall growth magnitude ($r = .298$). These findings distinguish contribution from explanation. Cohort composition identifies where the observed growth signal originates, whereas broader educational, developmental, and measurement contexts are required to explain differences in the magnitude of longitudinal growth across assessment years.

Subject-specific analyses further reinforce this distinction. Mathematics exhibited a substantially larger increase in Performance Level 1 representation than English Language Arts, resulting in correspondingly larger contributions to the observed growth signal. During the 2024-25 to 2025-26 administration, students beginning at Performance Level 1 represented approximately one-half of the linked cohort and contributed a disproportionate share of the total observed growth signal.

Collectively, these findings indicate that students' initial achievement position strongly influences both the distribution of observed growth and the relative contribution of different student groups. At the same time, they demonstrate

Figure 3

Cohort Composition and Contribution to Reliable Growth



that cohort composition alone is insufficient to explain the broader longitudinal patterns observed across the longitudinal record.

4.5 Growth and Achievement-Level Movement

Scale-score movement and achievement-level movement were related but not interchangeable. Scale scores represented the magnitude and direction of change; achievement levels located that change relative to established performance standards. Their agreement strengthened interpretation, while their divergence identified progress that either indicator alone could obscure.

Figure 4 summarizes movement toward the Level 3 performance standard among students beginning below proficiency. By relating continuous growth to changes in achievement-level status, the figure illustrates how students may make meaningful developmental progress even when they remain within the same performance level.

Figure 4
Movement Toward Proficiency

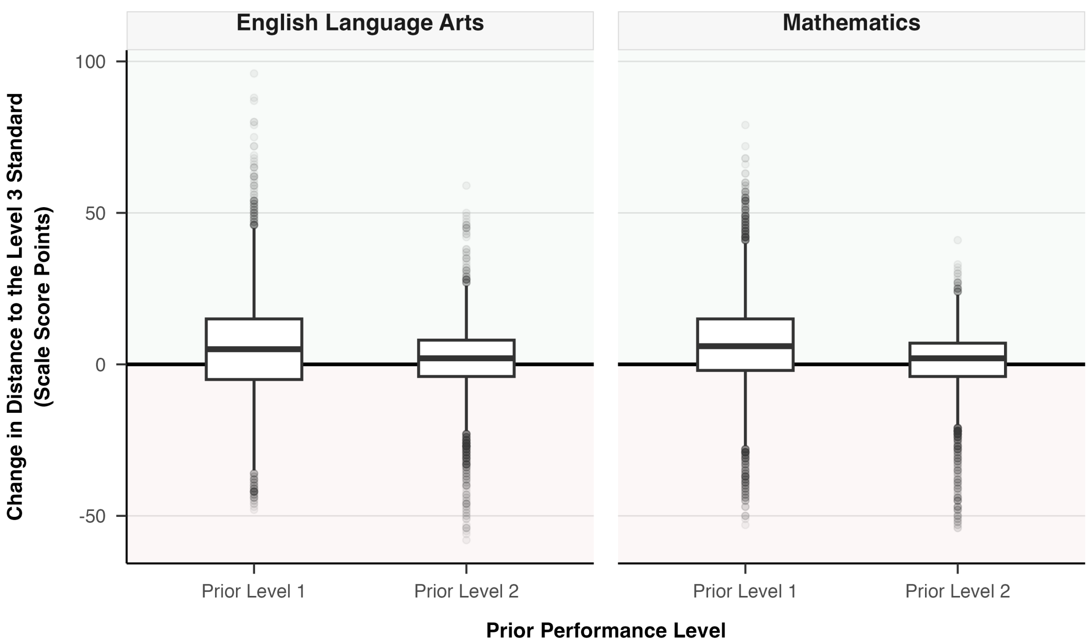


Figure 4 demonstrates that substantial positive movement frequently occurred without students crossing the Level 3 performance threshold. Students beginning farther below proficiency often exhibited meaningful developmental gains while remaining within the same achievement level. These findings demonstrate that continuous scale-score movement captures important evidence of student progress that is not always reflected by achievement-level transitions.

Distance-to-standard analyses reinforce this interpretation. Students beginning closest to the Level 3 cut score were substantially more likely to attain Level 3 during the subsequent administration than students beginning farther below the standard. However, many students beginning well below the cut score demonstrated measurable movement toward proficiency without crossing the reporting threshold. Together, these findings demonstrate that movement toward proficiency and attainment of proficiency represent complementary, rather than interchangeable, longitudinal indicators of student progress.

Historical comparisons further strengthen this distinction. During the pre-COVID period, most students remaining within the same achievement level were classified as not distinguishable from measurement uncertainty. In contrast, post-COVID analyses demonstrated that stable achievement-level classifications frequently contained both reliable increases and reliable decreases.

Collectively, these findings indicate that stability in achievement-level classification should not be interpreted as evidence that meaningful developmental growth did not occur. One notable exception emerged within Grade 6 Mathematics. Across multiple analyses, Grade 6 Mathematics demonstrated positive scale-score growth and a comparatively high reliable-increase rate while simultaneously exhibiting negative achievement-level movement. Multiple explanatory hypotheses, including cohort composition, cut-score spacing, measurement precision, and floor or ceiling effects, were evaluated but did not fully account for the observed divergence. Rather than weakening the overall findings, this unresolved pattern illustrates why complementary growth indicators should not be interpreted interchangeably. Divergence among indicators identifies important interpretive questions that would remain undetected if longitudinal growth were evaluated using any single measure in isolation.

Taken together, these findings support the conclusion that achievement-level movement and continuous scale-score movement provide complementary perspectives on longitudinal growth rather than interchangeable measures of student progress.

4.6 Reliable Growth

Measurement precision is the technical basis for reliable-growth classification. Conditional standard errors of measurement quantify the uncertainty associated with each annual score, whereas the reliable-growth classification indicates whether observed change exceeds that combined uncertainty. The two terms are therefore related but not interchangeable: measurement precision describes score uncertainty, and reliable growth describes the resulting interpretation of change.

Figure 5 summarizes reliable-growth classifications across prior achievement levels. By showing which observed score changes exceeded the uncertainty quantified through measurement precision, the figure illustrates the resulting reliable-growth interpretation beyond observed score change alone.

Figure 5

Reliable Growth by Prior Performance Level

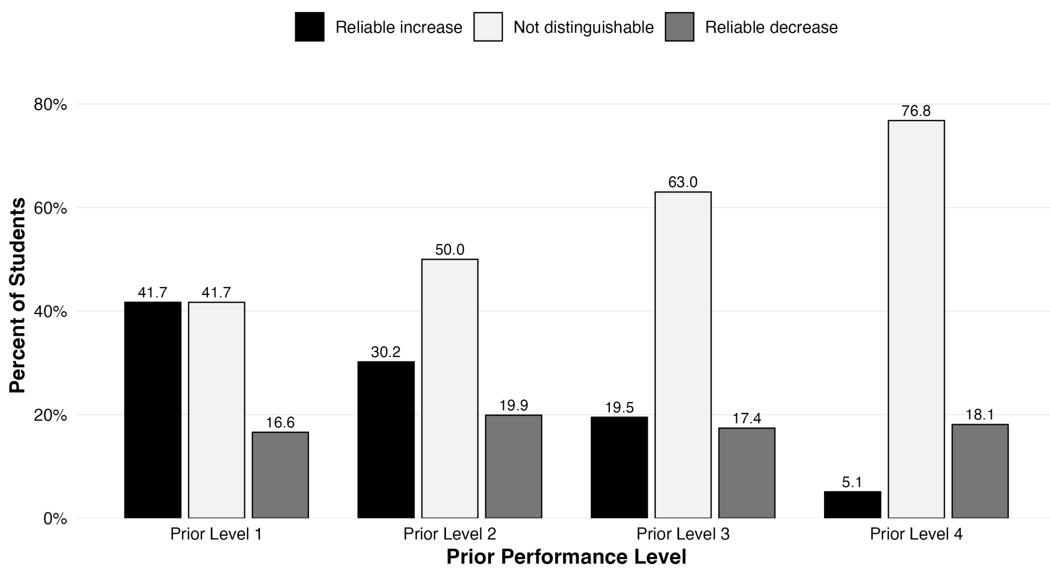


Figure 5 demonstrates a clear relationship between students' prior performance level and their subsequent reliable-growth classification. Students beginning at Performance Level 1 exhibited the highest proportion of reliable increases, while the likelihood of reliable increase declined steadily across higher prior performance levels. Conversely, the proportion classified as exhibiting reliable decreases increased systematically with prior achievement level, whereas the percentage whose score changes were not distinguishable from measurement uncertainty remained comparatively stable across levels.

Together, these findings indicate that starting achievement level strongly influenced reliable-growth outcomes and underscore the value of incorporating measurement precision when interpreting longitudinal score changes.

Across adjacent-year transitions, 28.8% were classified as reliable increases, 17.9% as reliable decreases, and 53.3% as not distinguishable from expected measurement uncertainty. Reliable increases occurred more frequently than reliable decreases, while most observed transitions did not exceed the 95% threshold for reliable change. These findings demonstrate that reliable-growth classifications distinguish observed changes that exceed expected measurement uncertainty from those that do not.

Reliable-growth classifications also provided evidence beyond that available from performance-level movement alone. Most reliable increases occurred within existing achievement levels rather than through movement across reporting thresholds. Similarly, reliable decreases were distributed across both same-level and downward transitions. Crosswalk analyses therefore demonstrated that reliable growth and achievement-level movement were clearly related while contributing distinct information regarding longitudinal growth. The consistency of these findings across English Language Arts and Mathematics, historical periods, and analytic approaches further strengthens the evidentiary basis for interpretation.

Starting achievement position emerged as one of the most consistent organizing characteristics within the longitudinal record and was associated with the likelihood of reliable increase, reliable decrease, and subsequent longitudinal patterns. The repeated convergence of these findings supports the interpretation that starting achievement position represents a structural characteristic of the longitudinal record rather than an artifact of a particular analytic method. Taken together, reliable growth provides an additional source of evidence by identifying observed change that exceeds expected measurement uncertainty while complementing scale-score movement and achievement-level classifications.

4.7 Persistence

Persistence analyses examined whether the direction of growth was sustained or changed across longer longitudinal sequences. Figure 6 summarizes persistence across extended sequences. By examining whether reliable-growth classifications continued in the same direction across successive adjacent-year transitions, the figure evaluates the stability and continuity of observed growth beyond individual year-to-year comparisons.

Figure 6
Persistence of Student Growth

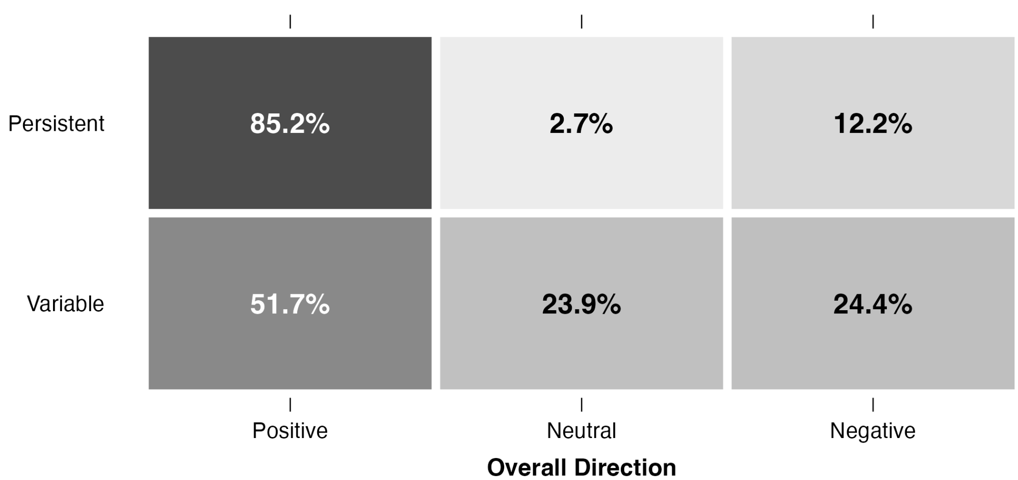


Figure 6 demonstrates that uninterrupted patterns were comparatively uncommon. Variable sequences were substantially more frequent and included changes in direction across successive observations. Persistence therefore complements adjacent-year analyses by evaluating whether the direction of growth continues across longer periods. Persistence analyses were conducted using student-subject sequences containing three or more consecutive observations. A total of 2,356 student-subject sequences satisfied this requirement and provided the evidence used to evaluate persistence across extended developmental windows.

Persistence was operationally defined as uninterrupted movement in the same direction across consecutive longitudinal observations, whereas one or more directional changes classified a sequence as variable. Variable patterns occurred substantially more frequently than uninterrupted progression, indicating that longitudinal growth frequently involved periods of both advancement and setback rather than continuous movement in a single direction. These findings extend longitudinal interpretation beyond adjacent-year comparisons. Whereas earlier analyses demonstrated that meaningful growth was observable and structurally organized, persistence analyses show that the direction of growth commonly changes over time. Evidence from extended longitudinal sequences therefore strengthens the interpretation that student growth is a dynamic developmental process rather than a sequence of independent annual outcomes.

Appendix A presents the longitudinal growth chapter developed for inclusion in the draft 2025-26 Oregon Extended Assessment Technical Report as an operational application of the interpretive framework described in this study. The chapter also illustrates one possible approach to statewide reporting of longitudinal growth results.

Appendix B extends this application by showing how the complementary indicators examined in the study may be integrated within a prototype longitudinal student report for educators and families.

5. Discussion

The findings are consistent with the broader growth literature, which emphasizes that growth models, measures, and indicators answer different questions and should not be treated as interchangeable (Betebenner & Linn, 2009; Castellano & Ho, 2013; Domaleski & Hall, 2016). The ORExt results provide an empirical illustration of that principle. Scale-score movement, achievement-level movement, progress toward proficiency, reliable-growth classifications, longitudinal patterns, and persistence frequently converged, but each preserved a distinct technical and interpretive function.

The study also extends prior alternate-assessment research. Earlier work demonstrated that achievement and growth can be documented and modeled for students with the most significant cognitive disabilities (Farley et al., 2017; Nese et al., 2017; Tindal et al., 2016), while related research described participation patterns and the heterogeneous conditions under which alternate-assessment results are produced (Buzick & Laitusis, 2010; Saven et al., 2016; Winter et al., 2022). The present findings move the question beyond whether growth can be estimated. They show how confidence in a longitudinal interpretation changes when multiple indicators are examined together under explicit measurement, developmental, and temporal conditions.

The distinction between achievement status and developmental change was especially important. Transition matrices remain useful for describing movement among policy-defined performance categories (Behavioral Research and Teaching, 2013; Tindal et al., 2017). Consistent with prior guidance, however, the ORExt evidence shows that performance-level movement provides only one component of a broader growth interpretation (Domaleski & Hall, 2016; Winter et al., 2022). Students frequently demonstrated meaningful scale-score movement without crossing an achievement-level boundary, and stable achievement-level classifications contained both reliable increases and reliable decreases. Status stability therefore could not be interpreted as an absence of developmental change.

The reliable-growth findings reinforce the importance of measurement precision. The *Standards for Educational and Psychological Testing* and related growth guidance emphasize that observed score differences should be interpreted in relation to the uncertainty associated with the scores (American Educational Research Association et al., 2014; Domaleski & Hall, 2016). In this study, reliable-growth classifications did not establish that every change represented a permanent developmental effect. Rather, they distinguished changes that exceeded the combined uncertainty of two annual scores from those that did not. The finding that more than one-half of historical adjacent-year transitions were not distinguishable from measurement uncertainty therefore supports caution against interpreting every observed score difference as meaningful growth.

The multi-year findings also qualify interpretations based on annual averages. Positive average growth and cross-indicator grade-subject patterns did not imply uninterrupted improvement for individual students. Variable sequences were much more common than persistent sequences, and the direction of change frequently shifted across successive observations. This result is consistent with the heterogeneity emphasized in the alternate-assessment literature (Buzick & Laitusis, 2010; Winter et al., 2022), while adding evidence that longitudinal variability remains visible even when annual indicators show an overall positive signal.

Divergence among indicators was informative rather than merely inconvenient. Grade 6 Mathematics demonstrated positive scale-score growth and a comparatively high reliable-increase rate while showing negative achievement-level movement. The available analyses did not fully resolve that pattern. Nevertheless, the divergence exposed an interpretive question that would have been hidden by a single summary measure. In this sense, contradiction among indicators serves a diagnostic function by identifying where additional measurement or contextual investigation is needed.

The principal contribution of this study is therefore interpretive rather than computational. The study does not propose another growth statistic or collapse distinct indicators into a composite index. Instead, it provides an empirically evaluated framework in which complementary indicators retain their technical meanings and are interpreted through the convergence, divergence, and conditions of the available evidence.

6. Limitations

The findings should be interpreted within several evidentiary boundaries.

First, the analyses are descriptive rather than causal. The evidence supports claims regarding the interpretation of longitudinal growth within ORExt but does not support causal inferences regarding instructional effectiveness, program effectiveness, school effects, district effects, or policy interventions.

Second, adjacent-year growth pairs provided the principal evidentiary unit for many analyses. Although supplemental analyses incorporated longer longitudinal sequences, those sequences represented a progressively more selective subset of students with uninterrupted records. Additional analyses of extended records are needed to strengthen claims regarding long-term developmental processes.

Third, the evidence was derived from the Oregon Extended Assessment. Many of the interpretive challenges identified here are likely relevant to other alternate assessment systems, but evidence from independent AA-AAAS programs is required before the framework can be generalized more broadly.

Fourth, reliable-growth analyses were constrained by the availability and comparability of precision-related information across assessment years. Consequently, several longitudinal inferences relied on the years for which comparable measurement evidence was available.

Fifth, the cross-indicator designation used in Figure 2 is comparative and descriptive. Ranking in the upper half on all three indicators identifies consistency across the ten observed contexts, but it does not establish statistical significance, a causal explanation, or a permanent grade effect.

Finally, the principal contribution is an interpretive framework rather than an operational reporting model. The evidence supports the claim that complementary indicators strengthen longitudinal interpretation, but it does not establish an optimal weighting system, composite growth score, or reporting algorithm. Those claims require additional empirical evaluation.

These limitations define the scope of the manuscript's claims and the conditions under which the framework should be applied.

7. Conclusions

The limitations above constrain causal, generalizable, and operational claims. Nevertheless, three conclusions remain well supported by the ORExt longitudinal evidence.

First, meaningful longitudinal growth is observable within ORExt when interpreted through multiple complementary sources of evidence rather than through any single indicator. Second, growth varies systematically across students' starting achievement levels, subjects, grade transitions, and longitudinal records rather than being distributed uniformly across the assessment system. Third, the strongest longitudinal inferences arise when distinct indicators converge under explicit conditions of reliable measurement, developmental context, and longitudinal continuity.

The framework therefore treats growth as an evidentiary interpretation rather than a score difference, achievement-level transition, or statistical classification considered alone. Reliable measurement addresses whether observed change can be distinguished from expected measurement uncertainty. Developmental context locates that change within the conditions under which it occurs. Longitudinal continuity describes how change unfolds across repeated observations, including whether direction is sustained or changes across time.

Although developed within ORExt, the framework has broader relevance for alternate assessments and statewide growth reporting. Its contribution is not a new growth metric, but a disciplined approach to determining what the combined evidence supports. The limitations prevent causal attribution and broad, unqualified generalization, but they do not overturn the central finding that longitudinal interpretation is strengthened when complementary indicators retain their distinct meanings and conclusions are supported through converging evidence.

8. Implications

The conclusions support several practical and research implications while remaining within the evidentiary boundaries described above.

8.1 Implications for ORExt

The findings support several actions for future ORExt growth reporting and program monitoring:

- **Preserve** achievement status, scale-score change, movement toward proficiency, reliable growth, and persistence as distinct reporting indicators.
- **Monitor** grade-subject contexts in which indicators consistently converge or diverge.
- **Interpret** annual results relative to students' starting achievement levels and the broader longitudinal record.

- **Separate** standard adjacent instructional-year comparisons from extended intervals such as the COVID interruption.
- **Develop** student reports that communicate complementary growth evidence without reducing progress to a single label.
- **Evaluate** whether cross-indicator grade-subject patterns remain stable across future administrations.

Appendix A illustrates how this framework may be incorporated into statewide technical reporting, while Appendix B provides one possible application to longitudinal student reporting.

8.2 Implications for Alternate Assessments

The findings support three broader applications:

- **Combine** complementary indicators rather than treating one measure as a complete representation of growth.
- **Account** for heterogeneous developmental contexts and comparatively small longitudinal cohorts.
- **Replicate** the framework across independent AA-AAAS systems before drawing broader generalizations.

8.3 Implications for State Accountability Systems

The findings support three interpretive safeguards:

- **Treat** individual growth metrics as components of a broader evidentiary framework.
- **Distinguish** operational accountability summaries from comprehensive interpretations of student development.
- **Retain** contextual information needed to prevent overinterpretation of a single reported indicator.

8.4 Implications for Growth Reporting

The findings support four reporting practices:

- **Report** achievement-level advancement, reliable growth, movement toward proficiency, and persistence separately.
- **Show** progress that occurs within an achievement level.
- **Summarize** multi-year patterns without obscuring annual variation.
- **Translate** technical evidence into concise, family-accessible narratives.

The practical importance of these distinctions is evident among students beginning at Achievement Level 1 who demonstrated reliable increases across contiguous sequences spanning three or more grades. Although 68.7% remained within Achievement Level 1, 66.5% moved closer to the proficiency standard. Meaningful developmental progress therefore frequently occurred without corresponding achievement-level advancement.

8.5 Future Research

The most important next step is replication across additional alternate assessment systems to determine whether the convergence observed under conditions of reliable measurement, developmental context, and longitudinal continuity represents a broader characteristic of alternate academic achievement assessments or reflects conditions unique to ORExt.

Future work should also refine the explanatory basis for longitudinal growth interpretation. The Grade 6 Mathematics divergence, differences associated with starting achievement position, and the stability of longitudinal patterns across extended developmental periods remain important questions. Research should also evaluate how complementary indicators can be translated into reporting systems that remain technically defensible while providing meaningful information for educators, families, districts, and state agencies.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Behavioral Research and Teaching. (2013). *Analysis of growth on state tests for students with significant cognitive disabilities* (Technical Report No. 1311). Behavioral Research & Teaching, University of Oregon.

Betebenner, D. W., & Linn, R. L. (2009). Growth in student achievement: Issues of measurement, longitudinal data analysis, and accountability. *ETS Research Report Series*, *2009*(2), i–31. https://doi.org/10.1111/j.2333-8504.2009.tb02187.x

Buzick, H. M., & Laitusis, C. C. (2010). Using growth for accountability: Measurement challenges for students with disabilities and recommendations for research. *Educational Researcher*, *39*(7), 537–544. https://doi.org/10.3102/0013189X10382294

Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models*. Council of Chief State School Officers.

Domaleski, C., & Hall, E. (2016). *Guidance for estimating and evaluating academic growth*. National Center & State Collaborative.

Farley, D., Anderson, D., Irvin, P. S., & Tindal, G. (2017). Modeling reading growth in grades 3 to 5 with an alternate assessment. *Remedial and Special Education*, *38*(4), 195–206. https://doi.org/10.1177/0741932516650441

Nese, J. F. T., Kamata, A., & Tindal, G. (2017). A two-step sampling weight approach to growth mixture modeling for emergent and developing skills with distributional changes over time. *Journal of School Psychology*, *61*, 55–74. https://doi.org/10.1016/j.jsp.2016.12.001

Oregon Department of Education. (2015). *Oregon Extended Assessment technical documentation*.

Oregon Department of Education. (2024). *Setting standards for student growth in Oregon* [Draft report].

Rowley, B., Tindal, S., Anderson, D., & Tindal, G. (2025). *Oregon Extended technical report*. Behavioral Research & Teaching, University of Oregon.

Rowley, B., Tindal, S., Irvin, P. S., Anderson, D., & Tindal, G. (2019). *Oregon Extended technical report*. Behavioral Research & Teaching, University of Oregon.

Saven, J. L., Anderson, D., Nese, J. F. T., Farley, D., & Tindal, G. (2016). Patterns of state assessment participation for students with significant cognitive disabilities. *The Journal of Special Education*, *49*(4), 209–220. https://doi.org/10.1177/0022466914567376

Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., Crawford, L., & Hollenbeck, K. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Ex-*

ceptional Children, *69*(4), 481–494. https://doi.org/10.1177/001440290306900406

Tindal, G., Nese, J. F. T., Farley, D., Saven, J. L., & Elliott, S. N. (2016). Documenting reading achievement and growth for students taking alternate assessments. *Exceptional Children*, *82*(3), 321–336. https://doi.org/10.1177/0014402915585492

Tindal, G., Nese, J. F. T., & Stevens, J. J. (2017). Estimating school effects with a state testing program using transition matrices. *Educational Assessment*, *22*(3), 189–204. https://doi.org/10.1080/10627197.2017.1344093

Winter, P., Karvonen, M., D'Brot, J., Hoover, J. C., & Hopfengardner Warren, S. (2022). *State accountability systems: Measuring and reporting academic growth for students with significant cognitive disabilities*. Council of Chief State School Officers.

Appendix A

Appendix Note.

The following chapter is reproduced from the draft 2025-26 Oregon Extended Assessment Technical Report to illustrate one operational application of the longitudinal growth framework developed in this study. The manuscript presents the primary research analyses conducted across the longitudinal record using the linked student-subject analytic framework. The accompanying technical report chapter uses the broader operational reporting dataset and related summary procedures to place the 2025-26 administration within its historical context. Consequently, some sample sizes and descriptive summaries differ between the manuscript and this appendix because they reflect distinct analytic datasets and reporting purposes. The chapter is included to illustrate the translation of research into statewide assessment practice rather than as a separate research contribution.

Technical Report Section 6.5: Longitudinal Growth

The primary purpose of this technical report is to document the operational characteristics and performance of the 2025-26 Oregon Extended Assessments. The longitudinal analyses presented in this chapter provide historical context for interpreting the 2025-26 administration rather than serving as a separate longitudinal research study. By examining adjacent-year growth, reliable growth, and multi-year longitudinal patterns across the complete longitudinal record, the 2025-26 results can be interpreted relative to established historical patterns. Throughout this chapter, particular attention is given to where the 2025-26 administration aligns with, or differs from, the broader longitudinal evidence.

Adjacent-Year Scale Score Growth

Adjacent-year scale-score movement was examined to quantify changes in student performance between consecutive administrations. Scale scores provide the primary continuous measure of longitudinal change and establish the foundation for all subsequent growth analyses presented in this chapter. Table A1 summarizes adjacent-year scale-score movement across nine adjacent-year comparisons representing 31,234 growth events from 7,848 unique students.

Table A1

Summary of Adjacent-Year Scale Score Growth

Subject	Average growth	Strongest year	Peak growth	Most recent growth
English Language Arts	2.18	2018-19 to 2021-22	5.27	3.59
Mathematics	2.72	2018-19 to 2021-22	7.48	5.33

Average annual gains were positive across nearly every adjacent-year comparison in both English Language Arts and Mathematics. English Language Arts demonstrated average adjacent-year gains of 2.18 scale-score points, while Mathematics averaged 2.72 points. Although average gains varied among administrations, positive movement was consistently observed across the longitudinal record. The largest average gains occurred during the 2018-19 to 2021-22 comparison period, reflecting the extended interval associated with the COVID interruption. Average gains were substantially smaller during the 2022-23 through 2024-25 administrations before increasing again during the 2024-25 to 2025-26 administration. Standard deviations remained relatively stable across all adjacent-year comparisons (approximately 12–16 scale-score points), indicating that differences among administrations were primarily associated with shifts in average growth rather than substantial changes in score variability.

Figure A1
Mean Adjacent-Year Scale Score Change

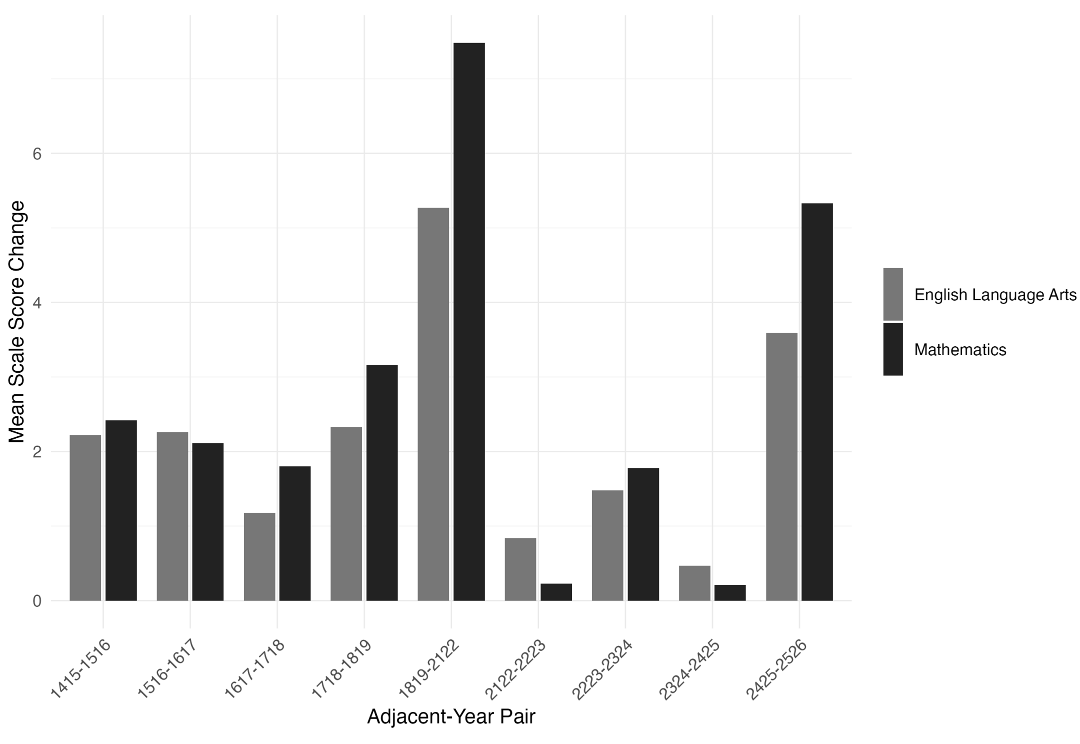


Figure A1 illustrates these longitudinal patterns graphically. Although average annual growth fluctuated across administrations, positive adjacent-year movement remained the predominant pattern throughout the observable longitudinal record. These historical patterns provide the benchmark against which the 2025-26 administration is interpreted throughout the remainder of this appendix.

Growth Patterns by Grade and Subject

Historical adjacent-year growth across the complete longitudinal record was examined by grade-subject combination to identify where growth was strongest within the ORExt assessment system. In addition to average scale-score growth, reliable-increase rates and achievement-level improvement were examined to provide complementary perspectives on student progress.

Table A2

Historical Growth Patterns by Grade and Subject

Rank	Grade	Subject	Mean scale growth	Reliable increase (%)	Achievement-level increase (%)
1	4	English Language Arts	3.79	31.9	24.5
2	4	Mathematics	3.21	30.8	21.2
3	7	Mathematics	2.96	32.7	19.4
4	7	English Language Arts	2.48	28.9	18.9
5	6	Mathematics	2.46	30.2	16.8
6	8	Mathematics	2.10	27.6	17.0
7	5	English Language Arts	2.04	26.7	14.7
8	5	Mathematics	1.50	26.4	24.7
9	6	English Language Arts	1.25	26.9	17.8
10	8	English Language Arts	0.66	25.7	12.6

Table A2 summarizes adjacent-year growth across grade-subject combinations. Grade 4 demonstrated the largest average scale-score gains in both English Language Arts (3.79 points) and Mathematics (3.21 points). Mathematics Grade 7 (2.96 points) and English Language Arts Grade 7 (2.48 points) also demonstrated comparatively strong average growth. Reliable-increase rates exceeded 25% across every grade-subject combination, with the highest rate observed in Grade 7 Mathematics (32.7%). Achievement-level improvement occurred across all grade-subject combinations but consistently represented a smaller proportion of students than reliable increases. Together, these findings indicate that longitudinal growth varied systematically across grades and subjects rather than occurring uniformly throughout the assessment system.

Figure A2
Achievement-Level Movement

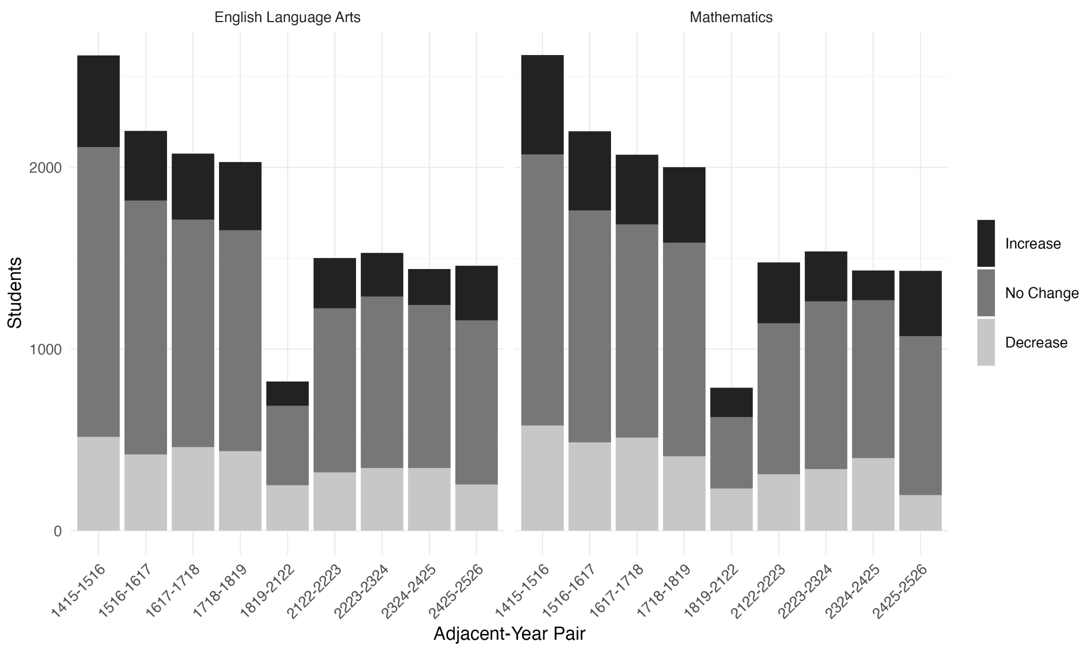


Figure A2 illustrates the distribution of achievement-level increases, decreases, and stable classifications across adjacent-year comparisons in English Language Arts and Mathematics. These grade- and subject-specific growth patterns provide important context for interpreting the distribution of student growth observed during the 2025-26 administration.

Movement Toward Proficiency

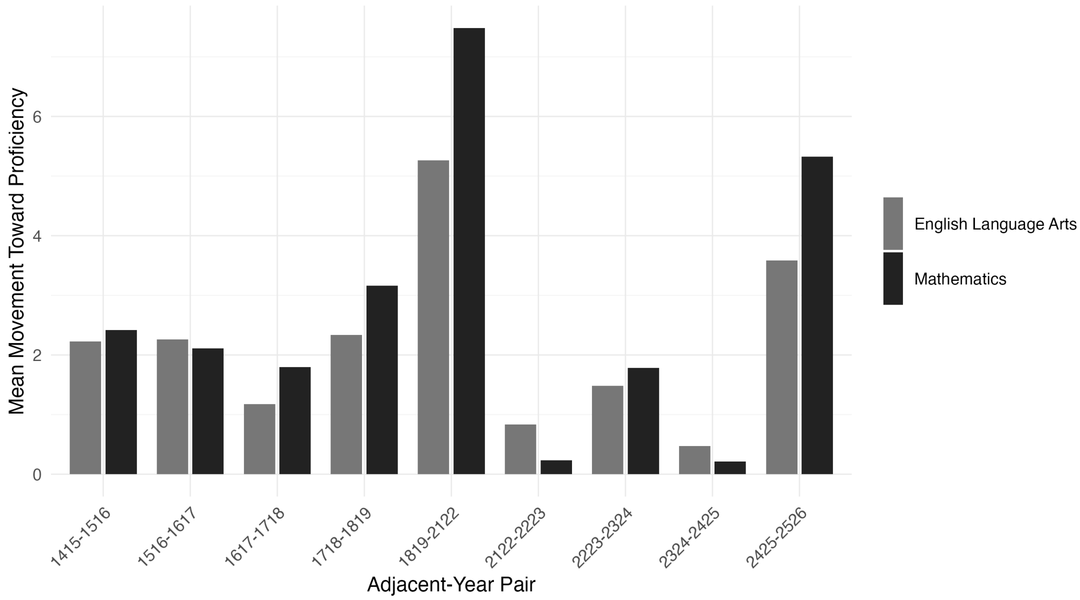
Movement toward proficiency provides an intuitive indicator of individual progress relative to the next achievement level. Figure A3 illustrates this metric across adjacent-year comparisons in English Language Arts and Mathematics. Because statewide movement toward proficiency is mathematically related to aggregate scale-score growth, it is presented as an applied reporting example rather than interpreted as an independent statewide growth indicator.

Figure A3 shows the distribution of changes in students' distance from the Level 3 performance standard. Positive values indicate movement toward the standard, whereas negative values indicate movement away from it.

Reliable Growth

Observed changes in scale scores reflect both changes in student achievement and expected variation arising from measurement error. Measurement precision, represented by each score's CSEM, quantifies that uncertainty. Reliable-

Figure A3
Movement Toward the Level 3 Performance Standard



growth classifications then distinguish observed changes that exceed the combined uncertainty from those that do not.

Table A4

Reliable Growth Classification

Reliable Growth Classification	N	Percent
Not distinguishable	16,651	53.31
Reliable decrease	5,598	17.92
Reliable increase	8,985	28.77

Table A4 summarizes reliable-growth classifications across 31,234 adjacent-year growth events. Overall, 16,651 transitions (53.3%) were not distinguishable from measurement uncertainty, 8,985 transitions (28.8%) demonstrated reliable increases, and 5,598 transitions (17.9%) demonstrated reliable decreases. Unlike traditional gain-score summaries, reliable growth evaluates whether observed change exceeds the amount of change expected from measurement error. Figure A4 shows how these reliable-growth classifications vary according to students' prior performance level.

Figure A4
Reliable Growth by Prior Performance Level

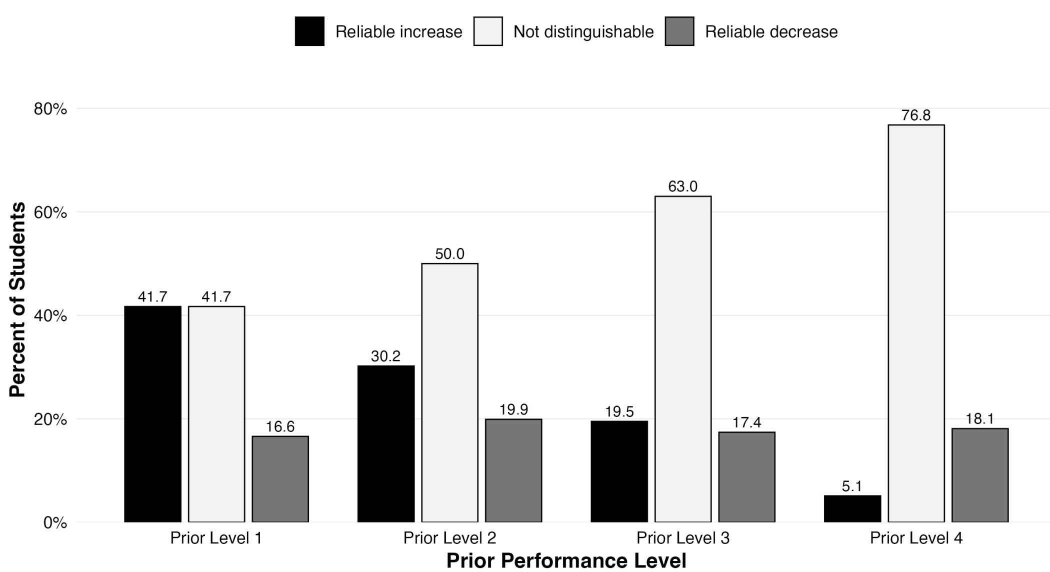


Figure A4 illustrates that the proportion of reliable increases declined systematically as prior performance level increased. Students entering at lower prior performance levels were substantially more likely to demonstrate reliable increases, whereas reliable decreases became more common as prior performance level increased and the proportion not distinguishable from measurement uncertainty remained comparatively stable. Consequently, students with similar observed score changes may receive different reliable-growth classifications when their measurement precision differs. Against this historical distribution, the 2024-25 to 2025-26 administration showed a markedly stronger pattern of reliable growth: 60.7% of adjacent-year transitions were classified as reliable increases, compared with 29.7% classified as reliable decreases and 9.6% classified as not distinguishable from measurement uncertainty. These year-specific results provide additional evidence that the gains observed during the 2025-26 administration reflected meaningful longitudinal progress rather than measurement variability alone.

Longitudinal Growth Patterns

Adjacent-year analyses describe year-to-year change, whereas longitudinal pattern analyses characterize growth across multiple consecutive administrations. Students with sufficient longitudinal records were classified into recurring pattern groups that summarize their overall direction of growth across time.

Table A5

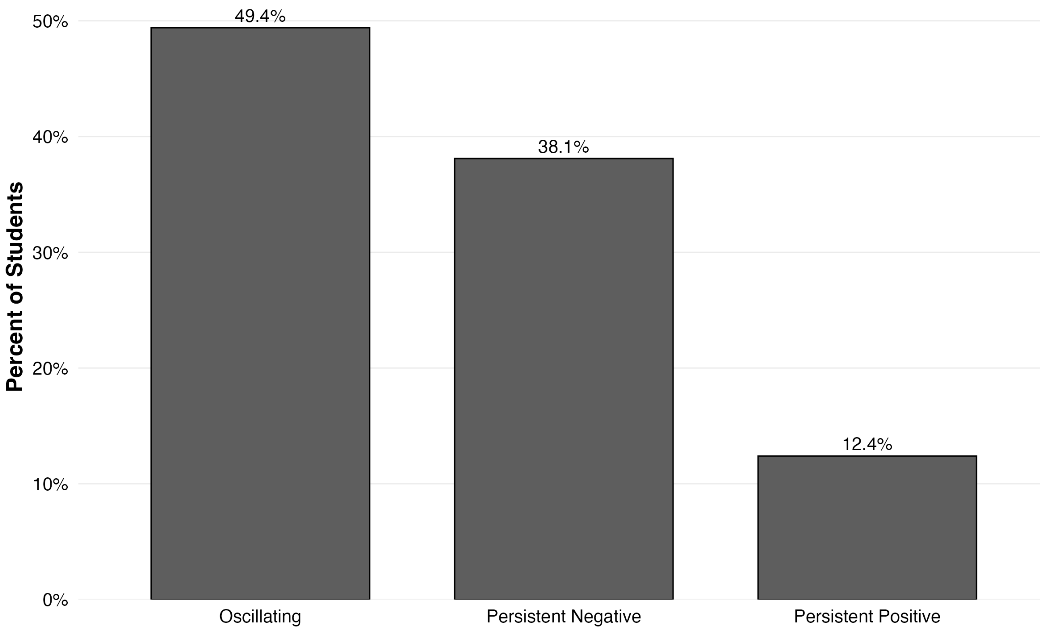
Longitudinal Growth Patterns

Family	N	Percent
Oscillating	1,165	49.4
Persistent Negative	898	38.1
Persistent Positive	293	12.4

Table A5 summarizes the three longitudinal patterns identified within the observable longitudinal cohort. Oscillating patterns were the most common longitudinal pattern (49.4%), indicating that many students experienced alternating periods of growth and decline across successive administrations rather than sustained movement in a single direction. Persistent negative patterns accounted for 38.1% of observable longitudinal records, while persistent positive patterns represented 12.4%. Together, these patterns demonstrate that longitudinal growth is heterogeneous and cannot be adequately described using a single adjacent-year transition. Figure A5 illustrates the relative frequency of the three longitudinal patterns.

Figure A5 shows that oscillating patterns were predominant, followed by persistent negative and persistent posi-tive patterns. These patterns establish the broader developmental context within which the 2025-26 administration should be interpreted.

Figure A5
Longitudinal Growth Patterns



Persistence

Persistence evaluates whether students maintained a consistent direction of growth across multiple consecutive administrations. Whereas longitudinal patterns describe the overall shape of a student's growth history, persistence examines whether growth continued in the same direction or reversed over time.

Table A6

Persistence

Persistence	Students	Percent
Variable	2,019	85.7
Persistent	337	14.3

Table A6 summarizes persistence across the observable longitudinal cohort. Among 2,356 student-subject sequences, 337 (14.3%) demonstrated persistent growth patterns, whereas 2,019 (85.7%) demonstrated one or more changes in growth direction and were classified as variable. Variable patterns indicate that students experienced both periods of improvement and periods of decline across successive administrations, whereas persistent patterns maintained a consistent direction of movement throughout the observable longitudinal record. Viewed together with the preceding longitudinal-pattern analyses, these findings indicate that longitudinal growth is typically dynamic rather than uniformly increasing or decreasing across time. Figure A6 summarizes the distribution of persistence across the observable longitudinal cohort.

Figure A6
Persistence of Student Growth

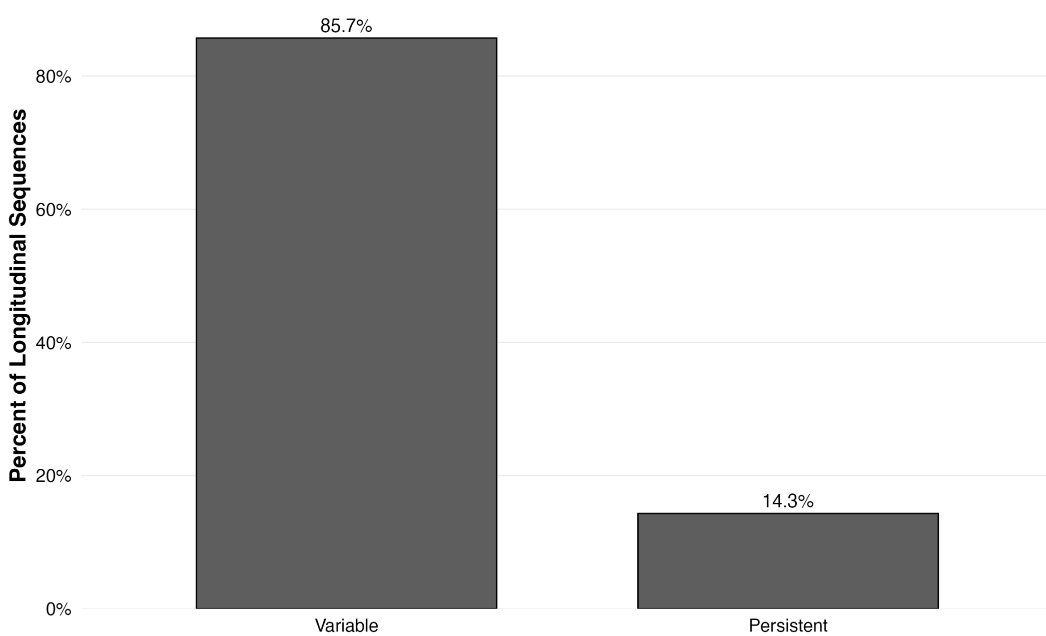


Figure A6 shows that variable patterns substantially outnumbered persistent patterns, indicating that changes in growth direction were considerably more common than uninterrupted longitudinal progress or decline. These findings reinforce the conclusion that adjacent-year growth should be interpreted within the broader context of students' complete longitudinal records rather than as isolated year-to-year transitions.

Together, these findings place the 2025-26 administration within the broader pattern of longitudinal growth observed across the Oregon Extended Assessment system.

2025–26 Growth Summary

The preceding longitudinal analyses establish the historical context within which the 2025-26 Oregon Extended Assessment results should be interpreted. Across adjacent-year growth, grade-specific growth patterns, reliable growth, multi-year longitudinal patterns, and persistence, the structure and distribution of growth observed during the 2025-26 administration generally aligned with established patterns across the assessment record. At the same time, the magnitude of the 2024-25 to 2025-26 growth signal was notably stronger than that observed during the two preced-ing administrations. Students demonstrated positive adjacent-year growth in both English Language Arts and Math-ematics. Reliable-growth analyses classified 60.7% of adjacent-year transitions as reliable increases, compared with 29.7% classified as reliable decreases and 9.6% classified as not distinguishable from measurement uncertainty. To-gether, these findings indicate that the 2025-26 results reflected meaningful changes in student achievement rather than measurement variability alone. Collectively, the findings show that the 2025-26 administration followed histori-cally recognizable patterns in where and how growth occurred while producing a comparatively strong growth signal in magnitude. This combination supports interpreting the 2025-26 results as both consistent with the broader devel-opmental structure of the ORExt longitudinal record and stronger than recent administrations. This longitudinal con-text provides the foundation for interpreting statewide growth during the 2025-26 administration and complements the student-level Growth Score Reports presented separately.

Appendix B

Research Translation: Prototype Student Growth Reporting

Appendix Note.

This appendix illustrates one possible application of the longitudinal growth framework to parent-facing student reporting. All student reports, historical graphics, identifiers, names, and assessment results are simulated and are included solely to demonstrate report structure and longitudinal reporting concepts. No personally identifiable information or actual student assessment records are presented.

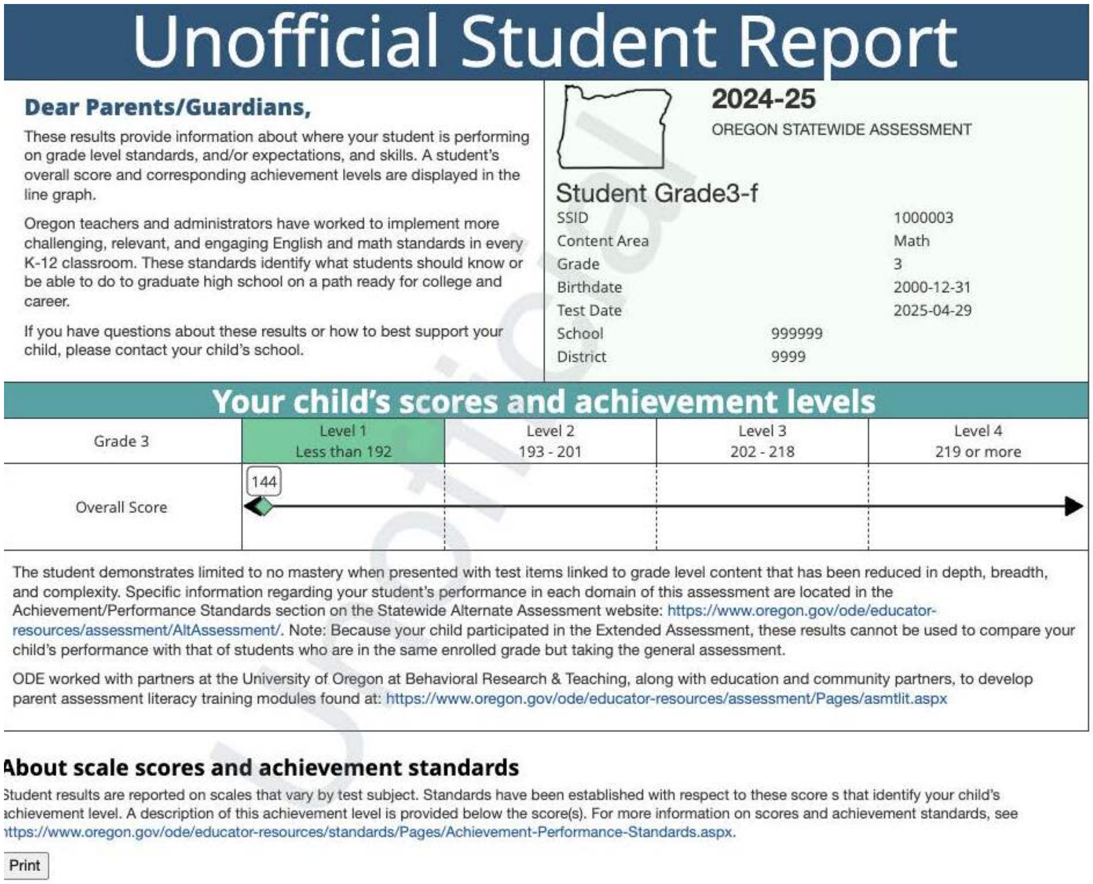
Purpose

The longitudinal analyses presented throughout this manuscript were conducted to improve the interpretation of student growth across successive administrations of the Oregon Extended Assessment. While these findings contribute to the technical understanding of longitudinal measurement, they also create opportunities to improve the way assessment results are communicated to students, families, and educators. This appendix illustrates one possible translation of the research into operational student reporting.

Current Oregon Student Growth Report

The current Oregon Extended Assessment student report provides families with a clear summary of annual assessment performance by reporting the student's current scale score, achievement level, and historical assessment results. These reporting elements are consistent with contemporary statewide assessment reporting practices and effectively communicate current student achievement.

Figure B1
Current Oregon Student Growth Report



Unofficial Student Report

Dear Parents/Guardians,

These results provide information about where your student is performing on grade level standards, and/or expectations, and skills. A student's overall score and corresponding achievement levels are displayed in the line graph.

Oregon teachers and administrators have worked to implement more challenging, relevant, and engaging English and math standards in every K-12 classroom. These standards identify what students should know or be able to do to graduate high school on a path ready for college and career.

If you have questions about these results or how to best support your child, please contact your child's school.

2024-25
OREGON STATEWIDE ASSESSMENT

Student Grade3-f

SSID	1000003
Content Area	Math
Grade	3
Birthdate	2000-12-31
Test Date	2025-04-29
School	999999
District	9999

Your child's scores and achievement levels

Grade 3	Level 1 Less than 192	Level 2 193 - 201	Level 3 202 - 218	Level 4 219 or more
Overall Score	144			

The student demonstrates limited to no mastery when presented with test items linked to grade level content that has been reduced in depth, breadth, and complexity. Specific information regarding your student's performance in each domain of this assessment are located in the Achievement/Performance Standards section on the Statewide Alternate Assessment website: https://www.oregon.gov/ode/educator-resources/assessment/AltAssessment/. Note: Because your child participated in the Extended Assessment, these results cannot be used to compare your child's performance with that of students who are in the same enrolled grade but taking the general assessment.

ODE worked with partners at the University of Oregon at Behavioral Research & Teaching, along with education and community partners, to develop parent assessment literacy training modules found at: https://www.oregon.gov/ode/educator-resources/assessment/Pages/asmtlit.aspx

About scale scores and achievement standards

Student results are reported on scales that vary by test subject. Standards have been established with respect to these score s that identify your child's achievement level. A description of this achievement level is provided below the score(s). For more information on scores and achievement standards, see https://www.oregon.gov/ode/educator-resources/standards/Pages/Achievement-Performance-Standards.aspx.

Print

Although the current report communicates current achievement effectively, longitudinal growth is primarily interpreted through changes in annual scale scores and achievement levels. The research presented in this manuscript demonstrates that meaningful longitudinal interpretation can be strengthened by incorporating multiple complementary indicators of student growth rather than relying upon a single annual score difference.

Prototype Longitudinal Student Growth Report

The prototype presented below illustrates one possible implementation of the longitudinal growth framework developed throughout this study. Rather than replacing the existing reporting system, the prototype preserves the familiar reporting elements currently provided to families while integrating additional evidence derived from the longitudinal analyses.

New Reporting Features

- **Longitudinal growth visualization.** Displays student performance across multiple administrations to illustrate developmental progress over time.
- **Reliable growth.** Distinguishes observed growth that exceeds expected measurement uncertainty from change that remains within that uncertainty.
- **Growth toward proficiency.** Indicates whether a student is moving closer to the next achievement level, even when achievement levels remain unchanged.
- **Multi-year growth interpretation.** Summarizes patterns across multiple years rather than emphasizing isolated annual score changes.
- **Parent-friendly narrative.** Integrates multiple indicators into a concise explanation intended to support interpretation by families and educators.

Table B1 summarizes the reporting features preserved from the current Oregon Extended Assessment student report and the additional longitudinal indicators introduced in the prototype.

Table B1

Comparison of Current and Prototype Student Reporting Features

Reporting feature	Current report	Prototype report
Current achievement	Yes	Yes
Scale score	Yes	Yes
Achievement level	Yes	Yes
Historical performance	Yes	Yes
Adjacent-year growth	No	Yes
Reliable growth	No	Yes
Movement toward proficiency	No	Yes
Multi-year growth interpretation	No	Yes
Parent-facing growth narrative	No	Yes

Together, these features illustrate how multiple complementary indicators can be integrated into a single parent-facing report while preserving the distinct information contributed by each indicator.

Relationship to the Research

Every new reporting element shown in the prototype corresponds directly to analyses presented within this manuscript. Consequently, the prototype does not introduce new reporting metrics solely for presentation purposes; instead, it illustrates how the empirical findings developed throughout this study may be translated into meaningful information for families while remaining grounded in psychometrically defensible longitudinal evidence.

Figure B2

Prototype Longitudinal Student Growth Report