Novel Interpretations of Academic Growth

By: Gary L. Williamson, Ph.D., Senior Research Scientist

May 2015

About MetaMetrics®

MetaMetrics, founded in 1984, is an educational measurement and technology organization whose mission is to connect assessment with instruction. The company's distinctive frameworks for English and mathematics bring meaning to measurement and are used by millions to differentiate instruction, individualize practice, and improve learning across all levels of education.

Abstract

Integrating a construct theory with Rasch measurement not only places persons and tasks on a common scale, it also resolves the indeterminacy of scale location and unit size when the scale is anchored in an operationalized task continuum based on the construct theory. Such an approach has several advantages for understanding academic growth as evidenced in a series of empirical examples which demonstrate how to: a) conjointly interpret student reading growth in the context of reading materials concomitantly used during instruction; b) interpret a reading growth trajectory in light of future (e.g., postsecondary) reading requirements; c) forecast individual reading comprehension rates relative to both contemporary and future text complexity requirements; d) interpret developmental growth over longer periods of time by combining developmental and content area scores in ELA and mathematics on a common scale; and e) create growth velocity norms for average academic growth in reading and mathematics.

Keywords

reading; mathematics; longitudinal data; growth; norms

During the 1980s two measurement companies in the United States introduced a fundamental innovation in the scaling of student reading ability that moved the world closer to an absolute framework for the measurement of reading comprehension. The strategy entailed combining the Rasch measurement model with an operationalized reading construct theory. As with other item response theory (IRT) models, the Rasch model makes it possible to place persons and tasks (items) on a common scale, but certain scale properties (location, unit size) are arbitrary. The key innovation involved two steps that anchored the scale and defined its unit size in terms of a real-world context.

The first step was to define and validate a construct model that operationalized the reading difficulty of texts in terms of specific surface features of texts that are effective proxies for the cognitive demand experienced by readers while reading. Secondly, it was demonstrated that the empirical difficulties of a well-defined text-based item type could be nearly perfectly predicted by the complexities of the texts associated with the items. Once students and items were measured via the Rasch model, the construct theory was used to calibrate the items to the text difficulty continuum. This effected a direct correspondence between the person measures and the text measures in terms of a real-world text complexity continuum.

The company now known as Questar Assessment, Inc. was the first to use this type of approach. They developed the Degrees of Reading Power® (DRP®) Program, which reports student reading measures from criterion-referenced tests on a proprietary DRP Scale of Text Complexity, which it uses to measure the reading difficulty of printed material (Bruning, 1985). Nelson, Perfetti, D. Liben, and Liben (2012) described the scale as follows:

> DRP text difficulty is expressed in DRP units on a continuous scale with a theoretical range from 0 to 100. In practice, commonly encountered English text ranges from about 25 to 85 DRP units, with higher values representing more difficult text. (p. 11)

Questar defined a DRP *prose comprehension model* based on the application of the Bormuth (1969) readability formula to measure text complexity. Their reference item type was a text-embedded cloze item administered according to a specific protocol. The unit size of the DRP scale was specified in terms of a transformation of the Bormuth text complexity measure, R. Research has shown that the DRP scale places both student reading ability and text complexity on a common well-defined, unidimensional scale that remains invariant over time. Thus, research supports the claim that the DRP tests "are like measures in the natural sciences." (B. L. Koslin, Zeno, & Koslin, 1987, p. 171)

At nearly the same time, a second company pursued the same fundamental idea. MetaMetrics® developed The Lexile® Framework for Reading to measure both readers and texts on a common scale. They independently developed a *construct-specification equation* to operationalize text complexity and predict item difficulties (Stenner & Smith, 1982; Stenner, Smith, & Burdick, 1983). They also developed a well-defined reference item type (consisting of a text passage followed by a cloze-like sentence-completion stem) and demonstrated that the empirical difficulties of such items could be nearly perfectly predicted by the difficulties of the associated texts (Stenner, D.R. Smith, Horabin, & Smith, 1987). They coupled this construct model with a Rasch measurement model to place both a student's reading ability and a text's readability on a common invariant scale.

In order to define a logical unit for the Lexile scale, MetaMetrics chose to explicitly anchor its scale at two points on the text complexity continuum. Based on its anchoring, a Lexile scale unit equals 1/1000 of the difference between the readability of certain specific basal primers and the readability of an online adult encyclopedia (Stenner, H. Burdick, Sanford, & Burdick, 2007). This approach provided a well-defined unit of measurement that retains its absolute size across different applications of measurement. It may be noted that this method is directly analogous to the way the meter was standardized by Legendre (1805) through his work to determine the length of the meridian quadrant through Paris. It is also precisely analogous to the way that temperature scales are anchored.

Both the Lexile Framework and the DRP can be utilized to generate student scores that are reported on a text difficulty continuum, giving the scores supplemental interpretability anchored in a real-world context. Since their creation, both systems have been widely implemented in the United States. The primary use of both these systems to date appears to have been the matching of students with texts of appropriate difficulty.

In 2004, MetaMetrics launched The Quantile® Framework for Mathematics, a measurement system for mathematical understanding, which uses Rasch measurement to conjointly scale both persons and items and anchors the resulting scale in a real-world task continuum. The Quantile Framework uses a quantified mathematics lesson continuum as the real-world context for anchoring the developmental scale (Sanford-Moore et al., 2014). As a companion scale to the Lexile Framework, the Quantile Framework demonstrates that the strategy of combining Rasch measurement with a construct theory and anchoring the resulting scale in a real-world task continuum is a viable method for measurement which generalizes to multiple constructs. As was the case with the Lexile Framework, the Quantile Framework was primarily designed to link assessment with instruction (MetaMetrics, 2009).

The purpose of this paper is to demonstrate, through several examples, that interpretations of student academic growth benefit from the use of Rasch-based measurement scales that have been anchored in a real-world task continuum by means of construct theory. These examples benefit from the fact that one state had the foresight and commitment to utilize such scales over a long period of time. The state of North Carolina (NC) began linking its reading assessment scales to the Lexile Framework starting with the first edition of its End-of-Grade assessments (introduced in 1993) and continuing with subsequent editions of reading tests up to the current day. Similarly, the state began linking its mathematics assessments to the Quantile Framework starting with the third edition (introduced in 2006) of their mathematics End-of-Grade tests and continuing to the present day. In addition, the state began linking its high school content area tests in 2008, providing a basis to extend the longitudinal measurement of reading and mathematics achievement on common scales beyond the elementary and middle school years.

These measurement innovations adopted by North Carolina have several advantages for the interpretation of academic growth. As demonstrated in the examples, the benefits include: a) conjointly interpreting student reading growth in the context of reading materials concomitantly used during K-12 instruction; b) interpreting a reading growth trajectory in light of future (e.g., postsecondary) reading requirements; c) forecasting individual reading comprehension rates relative to both contemporary and future text complexity requirements; d) interpreting developmental growth over longer periods of time by combining developmental and content area scores on a common scale; and e) creating growth velocity norms for average academic growth in reading and mathematics.

Data

The “data” for the subsequent examples consist of the parameter estimates from multilevel growth models estimated for various panels of students who participated in the North Carolina assessment program. The parameter estimates are adopted from previous work (e.g., MetaMetrics, 2011; Williamson, 2014). The original student-level data, which were the basis for the fitted growth models, consisted of Lexile® or Quantile® measures that were obtained through linking the North Carolina assessment scales to the Lexile Framework and the Quantile Framework.

North Carolina assessments have well-documented technical characteristics (Bazemore & Van Dyk, 2004; North Carolina Department of Public Instruction, 2009; Sanford, 1996) and have successfully satisfied the requirements of the Elementary and Secondary Education Act (No Child Left Behind, 2002). In general, panels were comprised of longitudinal data spanning

Grades 3-8, where the assessments were administered once a year at the end of each grade. For some of the following examples, extended waves of data were available as explained in the related descriptions.

Examples

The first three examples are based on a multilevel unconditional quadratic growth model, which was fit to the longitudinal data from a North Carolina panel spanning Grades 3-8 in 2000-2005. Based on data from every student who had at least one reading measure during the six-year time frame, this curve provides a historical summary of average student reading growth for 98,515 students, representing 92.8% of the Spring 2005 eighth-grade cohort that defined the panel. The estimates of the intercept, velocity, and curvature parameters for the average reading growth curve were 670.2L, 119.6L/year, and -6.1L/year^2, respectively. In Figure 1, I provide a visual summary of the statewide average reading growth curve, the corresponding velocity curve, and the acceleration curve for reading growth based on the multilevel analysis.

Figure 1

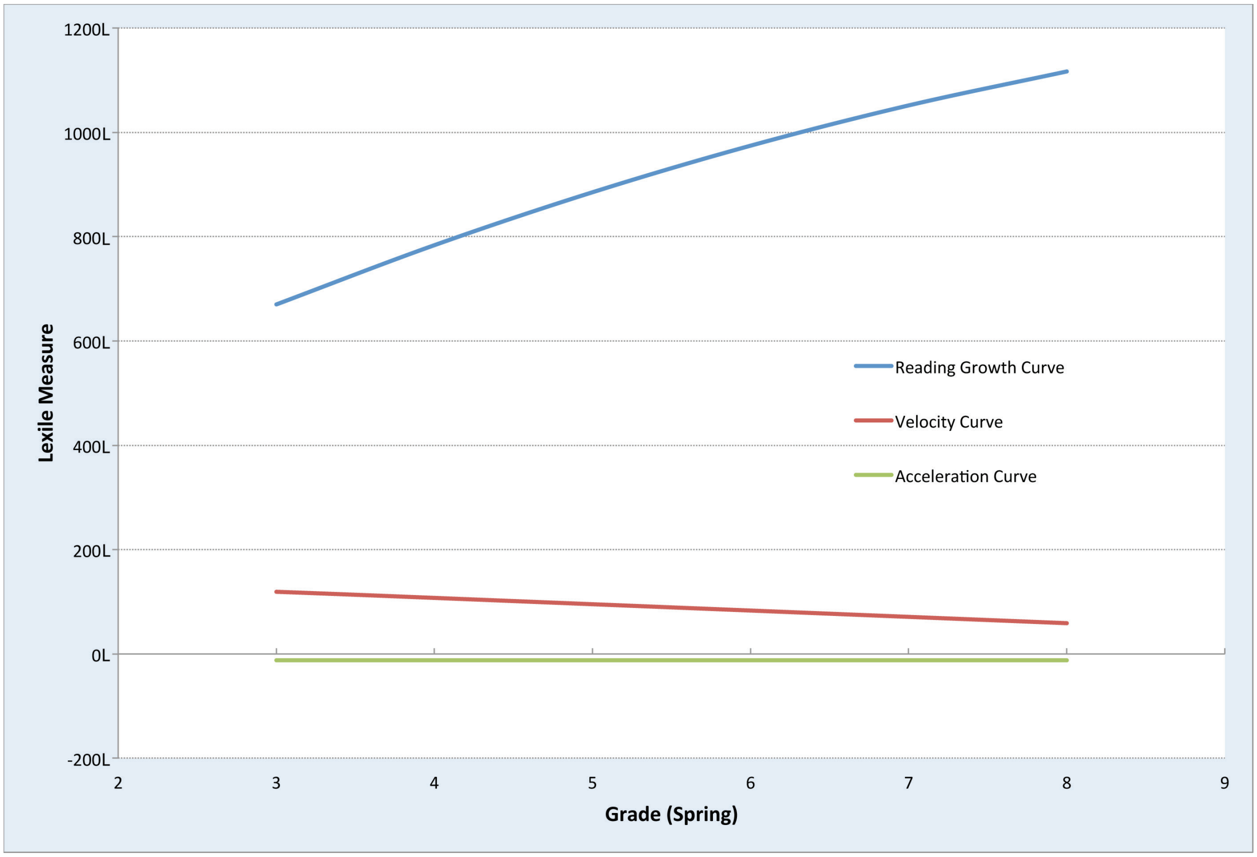


Figure 1. Average reading growth, velocity, and acceleration curves for the 2000-2005 North Carolina panel (n = 98,515). The left vertical axis graduates the growth curve in Lexile measures, the velocity curve in Lexile units/year, and the acceleration curve in Lexile units/year^2.

Note in Figure 1 the horizontal scale is graduated by grade, where the coding refers to the end of the respective year. So for example, the numeral 3 on the grade scale refers to the end of Grade 3. Furthermore, the time scale for the growth model was centered at the end of Grade 3; thus the velocity estimate refers to the velocity at the end of Grade 3. The vertical axis is denominated in Lexile units. The meaning of the Lexile scale unit was described earlier.

In Figure 1, notice that the growth curve begins around 670L at the end of Grade 3 and then rises quickly during the early grades; however, the curve decelerates across the Grade 3-8 time frame. The velocity curve in Figure 1 displays the fact that velocity is linearly related to time when the growth curve has a quadratic functional form. In this particular example, the velocity curve shows that velocity declines from approximately 120L/year at the end of Grade 3 to approximately 60L/year at the end of Grade 8. The slope of the velocity curve (-12.2L) is equal to the acceleration rate of the growth curve. Because the slope of the velocity curve is negative, growth is decelerating during the time frame. For a quadratic growth curve, the acceleration rate is manifested through the curvature parameter. Acceleration is constant and equal to twice the curvature parameter (i.e., -6.1L in this case). This is consistent with the constant negative elevation displayed for the acceleration curve in Figure 1. The growth, velocity, and acceleration curves are relatively simple for a quadratic growth model; nevertheless, it is useful to display them in the fashion of Figure 1 because it provides a convenient and readily understandable summary of the key features of growth.

Student Growth in Reading versus the Common Core State Standards

The Common Core State Standards (CCSS) Initiative (NGA Center & CCSO, 2010) established quantitative text complexity standards for specific grade bands in the public schools. The standards are expressed as text complexity ranges denominated in terms of six text complexity metrics in common use in the United States. One of those metrics is the Lexile measure, which makes it possible to compare the text complexity standards of the CCSS to actual student reading achievement measured with the Lexile Framework. The CCSS College and Career Readiness Anchor Standards for Reading require that by the end of specific grades that demark the end of the CCSS grade bands (i.e., Grades 3, 5, 8, 10, and 12), students must "read and comprehend literature, including stories, dramas, and poetry/poems, at the high end of the ... text complexity band independently and proficiently." (pp. 12, 37, 38) The upper end of the text complexity range for the Grade 11-12 grade band was labeled "CCR" by the CCSS to connote college and career readiness.

In Figure 2, I depict the 2000-2005 NC growth curve and the CCSS text complexity ranges for Grades 3, 5, and 8. The lower and upper boundaries of the CCSS text complexity ranges at the critical grades are represented by dots, which are connected by dashed lines to provide a visual reference as context for the growth trajectory. If student growth were commensurate with the CCSS text complexity standards, then one would expect to see the growth curve traversing a path that lies within the text complexity boundaries, rising near the upper end of the range by the specified grades, which denote the end of each grade band. In fact, the NC average growth curve approximates this behavior. Its intercept appears to be slightly above the mid-point of the text complexity range for the Grade 2-3 grade band and the curve rises nearer the upper boundary by the end of the Grade 6-8 grade band. If one imagines that the average growth curve is in fact the growth curve for an individual student, then it would seem that the student's growth is reasonably well aligned with the standards. Is it good enough? What does the growth curve imply about the actual reading experience that the student would have relative to the CCSS upper boundaries as he or she grows? I will come back to these questions in a subsequent example. First, I wish to introduce the idea that there are additional text requirements that characterize reading experiences which students may encounter after they graduate high school. Consequently, student growth during the K-12 years has implications for reading experiences that students will encounter later.

Figure 2

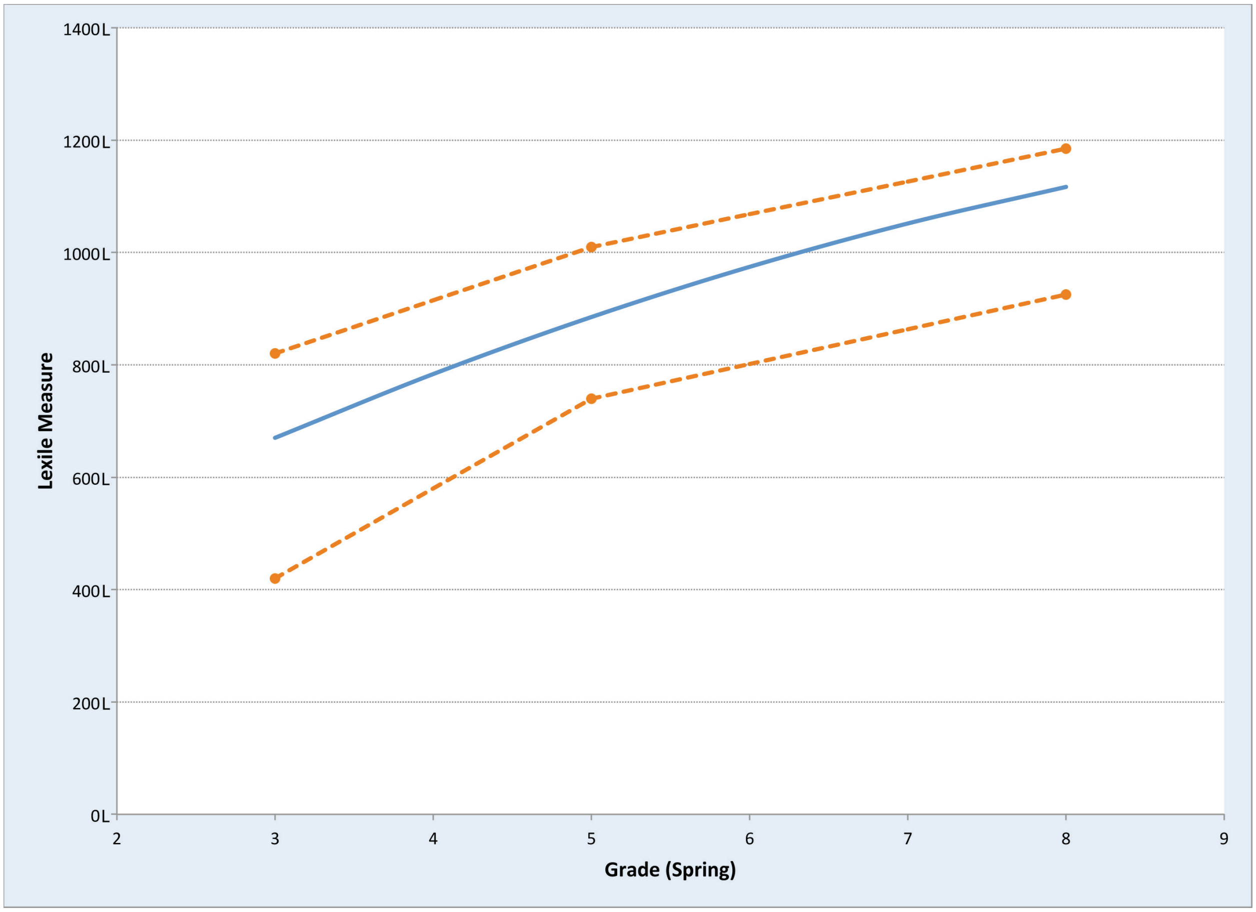


Figure 2. Reading growth relative to the Common Core State Standards (CCSS) text complexity ranges. The growth curve is the 2000-2005 North Carolina average growth curve (n = 98,515). The dots represent the CCSS Lexile range boundaries at grades 3, 5, and 8. The dashed lines provide a visual reference for the growth trajectory as it traverses the CCSS grade bands.

Reading Growth in Relation to Postsecondary Text Complexity

The objective of this example is to illustrate average student reading growth in relation to the text complexity of reading materials that students may encounter beyond high school. To accomplish this objective, I combine knowledge about the functional form of reading growth during K–12 with text complexity measures of postsecondary reading materials to construct an empirically-based model of student growth toward postsecondary performance aspirations. Such a model can be a useful first step toward understanding the possible long-term implications of growth.

Williamson (2008) elaborated a continuum of text complexity for reading materials associated with typical postsecondary endeavors (e.g., postsecondary education, the military, the workplace, citizenship). This work demonstrated substantial differences between the materials that high school students are expected to read and the materials they may encounter after high school. The latter reflect a substantially higher text demand, or correspondingly, require a higher reading ability from students in their postsecondary lives. The median Lexile measures for five postsecondary text collections summarized by Williamson are: 1395L (university), 1295L (community college), 1260L (workplace), 1230L (citizenship), and 1180L (military).

Once again, I use the statewide average reading growth curve of the 2000-2005 NC panel. Using the fixed effects estimates from the multilevel analysis, the average reading growth curve is expressed as a mathematical equation: $E(L|T) = 670.2 + 119.6\ T - 6.1\ T^2$. This equation quantifies the estimated average achievement in any grade.

The 2000-2005 panel is comprised of 98,515 North Carolina public school students who were third graders in the spring of 1999-2000 and who progressed to the end of eighth grade in the spring of 2004-05. These students progressed from Grade 3 to 8 without repeating a grade and were included in the analysis if they had at least one reading measure during the six- year time frame. Consequently, the average growth curve of these students should provide a good illustration of typical student growth toward postsecondary expectations. All of the relevant information about the growth curve is summarized in the three parameter estimates: 670.2L (initial status—end of third grade), 119.6L (initial velocity), and -6.1L (curvature).

Data were not available prior to the end of Grade 3 or after the end of Grade 8. However, with some caution, the quadratic equation that characterizes the curve through the range of observed data can be used to estimate average performance before Grade 3 and after Grade 8. Simply evaluating the growth curve at the other time points suffices.

It is important to use caution for at least two reasons. First, there are no actual data to check the assumption that growth from Grades K–2 and Grades 9–12 can be described by the same quadratic equation that describes growth from Grades 3–8. Second, the nature of a quadratic polynomial is that it has a maximum point or a minimum point, after which the curve reverses direction. When the curve is concave to the time axis (as is the case for the NC average growth curve), there will be a maximum point after which the curve turns downward. It is implausible that future performance will decline back to the third-grade level and below; this would be inconsistent with normal developmental growth.

There are (at least) three ways to address these concerns. The easiest way is to analytically check the quadratic equation to determine when the maximum point occurs. If it occurs outside the range of time to which one wishes to generalize, then there is less reason to worry that the depiction of growth may be inappropriate. As it turned out, the maximum for the 2000-2005 North Carolina growth curve occurred at Grade 12.9, almost a year beyond the end of twelfth grade, which is the last occasion for which average student achievement was projected.

A more direct way to avoid non-developmental behavior in a growth model is to adopt a different mathematical model for growth—e.g., one that cannot display a reversal in direction. A linear model with a transformed time scale is one possibility, such as: $r(t)=a+b \ln t$, which increases monotonically without bound. Another alternative is to select a model that is nonlinear in the parameters, such as the negative exponential: $r(t) = a - (a - b)\, e^{(-ct)}$, which increases monotonically to an asymptote. There are many possibilities (e.g., see Singer & Willett, 2003; or, Goldstein, 1979 for a variety of specific choices). Alternative models carry with them alternate interpretations of growth, may be more complex mathematically, and may require additional data to obtain satisfactory fit. Ultimately, the choice of most appropriate model is based on multiple considerations—e.g., substantive theory, available data, empirical fit, parsimony, and, perhaps, other requirements.

The third way to address the risks of extrapolation is to strategically collect more data to fill in the missing time points with student achievement information. Unfortunately, this is harder than it sounds for a variety of reasons, including the costs of collecting the information and the challenge of measuring the same construct over longer and longer periods of time. I offer one possible practical solution to this problem in a later example. For the present example, the results may be regarded as provisional, bearing in mind that extrapolations to lower and higher grades may need to be revised based on future information.

With those cautions in mind, Figure 3 shows the results of combining the information from the text analyses and the information from the NC reading growth curve. There are several important things to notice about Figure 3.

Figure 3

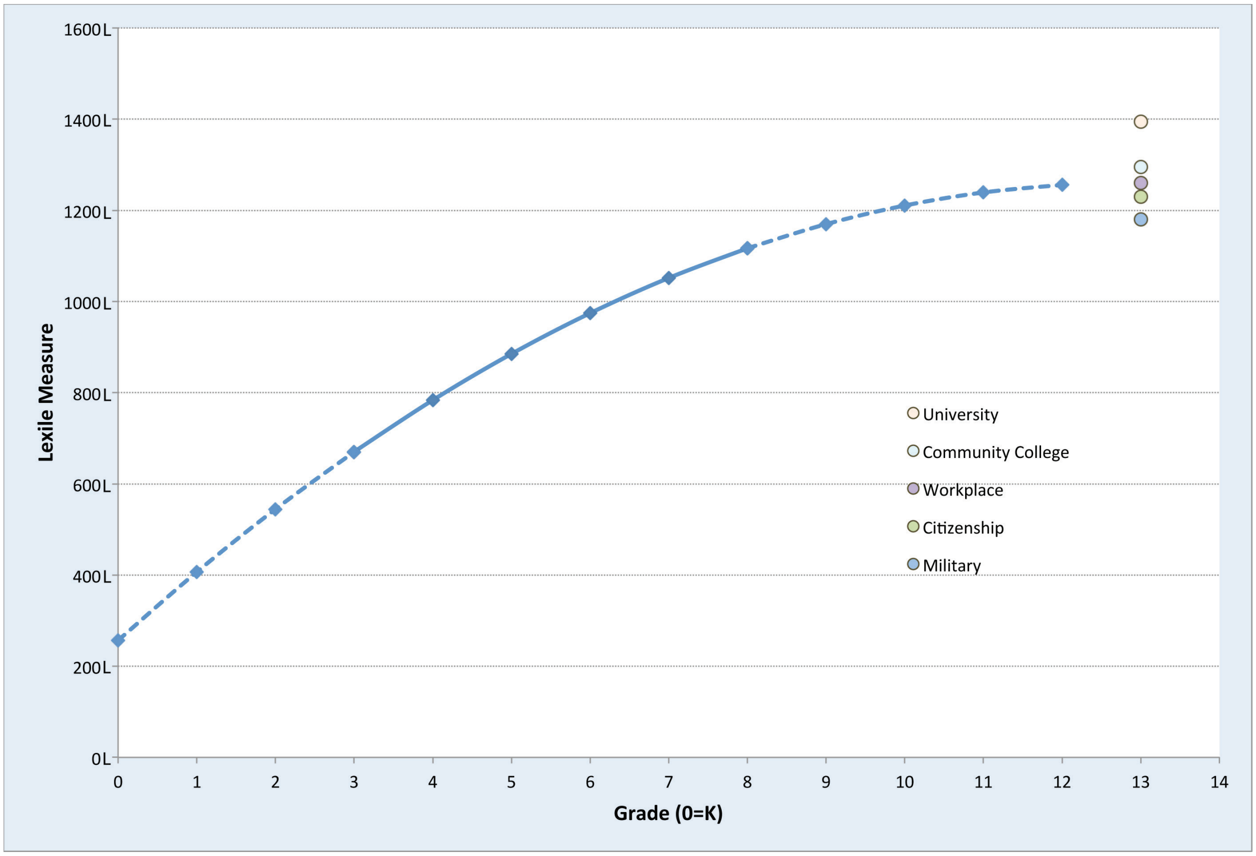


Figure 3. Average student growth in relation to postsecondary text complexity. The solid curve represents the 2000-2005 North Carolina average growth curve (n = 98,515). The dashed portions of the curve are mathematical extrapolations based on the quadratic equation for the average growth curve. The shaded dots in the upper right represent the median text complexities for the respective text collections listed in the legend (Williamson, 2008).

Once again, the horizontal scale represents grade in school. On this scale, zero stands for kindergarten. Subsequent Grades (1–12) are denoted as usual. Then the numerals 13 through 14 are used to denote the next two years of postsecondary experience. The vertical scale displays the Lexile measure, which is used to quantify *both* the students' average reading achievement and the median text difficulty of each text collection.

In the graph, diamonds are used to indicate the estimated average reading ability of students at each point in time. The estimates for Grades 3–8 are connected by the solid empirical growth curve to represent the fact that they are based on the available data. The estimates for Grades K–2 and 9–12 are connected with dashed curves to represent the fact that they are theoretical extrapolations determined analytically from the quadratic equation for the empirical growth curve. As such, the dashed portions of the curve are only reasonable guesses based on the observed data, subject to future revision based on more complete longitudinal records. The farther one goes from the observed data (Grades 3–8), the more one has to bear in mind the provisional nature of the projections. Finally, in the figure, the median text difficulty of the postsecondary text collections are arrayed vertically at Grade 13 to indicate that students face these expectations in the year following their exit from Grade 12.

The primary feature of the chart is the alignment of the projected twelfth-grade reader measure in conjunction with the postsecondary text measures.[i] It appears that the average growth trajectory of these students, if unaltered, will carry them to a reading level (1256L) that lies near the median text requirements of the workplace (1260L). Students with higher postsecondary aspirations (e.g., the community college, the university) need to be on a higher trajectory that tracks above the average growth curve depicted in the figure.

One must remember, however, that individual growth is variable and that students vary in their individual parameters of growth. That is, students have different beginning points, different initial velocities, and different degrees of deceleration. Each of these features of growth results in a slightly different trajectory compared to the average growth trajectory. Thus, there are many possible ways to reach a given end point. For example, one student might begin at a higher level and exhibit modest but steady growth with little deceleration over time. Another might start out lower in reading ability but progress very rapidly with some deceleration over time. Both students might reach the same twelfth-grade reading ability through different individual growth curves. Williamson, Fitzgerald, and Stenner (2014) discussed alternate growth trajectories in terms of the pedagogical and educational policy implications of directly targeting key features of growth (status, velocity, and acceleration). For example, early-intervention reading programs can successfully influence initial reading status; increased deliberate practice might impact velocity; and, systematic exposure to summer school could be a viable strategy to moderate deceleration.

Forecasted Comprehension Rates Based on a Growth Curve

For this example, I return to the question of what kind of reading experience students are likely to have with particular levels of text complexity—e.g., the CCSS text standards or postsecondary text requirements. Again using the 2000-2005 NC average growth curve and supposing that the curve might describe the trajectory of a particular individual, it is possible to estimate the individual's comprehension rate relative to texts the individual may encounter. To do this, it is necessary to have a general idea of how the Lexile Framework can be used to forecast reading comprehension given a reader of a particular reading ability and a text of a particular difficulty. Stenner, H. Burdick, Sanford, and Burdick (2007) described the approach. In essence, one forecasts the comprehension rate by using the Rasch model equation, which expresses the reading outcome (comprehension) as a function of the exponentiated difference between the reader's ability and the text's difficulty. The Lexile Framework is designed so that an exact match between reader and text (i.e., reader ability equals text complexity, and so the difference between the two is zero) results in a comprehension rate of 75%. A comprehension rate of approximately 75% seems to be associated with successful reading experiences; whereas, a comprehension rate of 50% or lower results in frustration for the reader (Scholastic, 2007). MetaMetrics typically advises educators to choose texts that lie in a proximal zone ranging from 100L below the reader's ability to 50L above it when using the Lexile Framework to match readers with texts of appropriate difficulty. This proximal zone corresponds to comprehension rates that range from approximately 70% to 80%.

Consider a reader whose growth curve is equal to the 2000-2005 NC average growth curve. What happens when such a student reads a book that has text complexity equal to the upper end of the CCSS text complexity ranges? What happens when such a student reads a book that has a text complexity equal to the typical text complexity of postsecondary reading materials (1300L)? In the first scenario, the CCSS text demand changes from grade to grade as the student's reading ability (reflected by the growth curve) changes. In the second scenario, there is a fixed future, postsecondary target toward which the student is progressing. I address both situations in Table 1.

For each of the Grades 3, 5, 8, 10, and 12 (i.e., the transition grades between the CCSS grade bands), I tabulate in the first four rows of Table 1: a) the average student performance (estimated from the growth curve); b) the CCSS text complexity upper bound; c) the difference between the two; and, d) the resulting forecasted comprehension rate at the end of the grade.

Table 1

Forecasted Comprehension Rates Implied by the 2000-2005 North Carolina Average Growth Curve Relative to a) the Common Core State Standards (CCSS) Grade Bands and Text Complexity Ranges and b) Median Postsecondary Text Complexity

	GRADE 3	5	8	10	12
	Average Student Achievement Summarized by the Longitudinal Growth Curve				
NC (2000-2005)	670L	885L	1117L	1211L	1256L[a]
	CCSS Text Complexity Requirements				
CCSS	820L	1010L	1185L	1335L	1385L
Difference	-150L	-125L	-68L	-124L	-129L
Forecasted Comprehension	61%	63%	69%	63%	63%
	Median Postsecondary Text Complexity				
Postsecondary Texts (Median)	1300L	1300L	1300L	1300L	1300L
Difference	-630L	-415L	-183L	-89L	-44L
Forecasted Comprehension	15%	32%	57%	67%	71%

Note. A multilevel growth analysis (n = 98,515) was used to estimate the average reading achievement at the end of each respective CCSS grade band. The upper boundaries of the CCSS Lexile ranges associated with the respective grades are given in the row labeled CCSS. A reader who is well matched with a text at his or her Lexile measure is forecasted to have a 75% comprehension rate.

[a] The empirical data spanned Grades 3-11. The estimated average achievement at the end of Grade 12 is extrapolated from the growth curve.

In general, we expect a reader to have 75% comprehension of a well-targeted text (i.e., a text at the students' reading level). Because the CCSS text complexity standards represent a series of increasing aspirational goals, we can ask how well the average reader in our example might do relative to the changing text complexity standards as he or she grows. That is, what would be the student's comprehension rate when confronted with a text with the higher text complexity prescribed by the CCSS? Table 1 provides the answer. The forecasted comprehension rates rise from 61% (in Grade 3) to 69% (in Grade 8) during the empirical time frame for the panel. However the comprehension rate is forecasted to drop back to 63% during the high school years, if the individual continues on the same trajectory traversed during Grades 3-8. Although, the CCSS grade bands and text complexity ranges are designed to provide flexibility to accommodate readers with a wide range of abilities, this example suggests that the average student in the 2000-2005 panel may experience some challenge relative to texts at the upper ends of the CCSS text complexity ranges (because all of the forecasted comprehension rates are less than 75%).

In the bottom half of Table 1, we can see that the hypothetical average student experiences increasing rates of comprehension while growing toward the fixed postsecondary text complexity target. Although forecasted comprehension of the median (1300L) postsecondary text is understandably low (15%) when the student reads as a typical third grader, the forecasted comprehension rate steadily climbs to 71% by the end of Grade 12, based on the estimated average reading growth curve.

A nice feature of this analysis is that it can be replicated with any estimated growth curve, whether for an individual or for a group of students. One only needs estimates of reading ability at each desired point in time, which can easily be determined from the mathematical equation for growth.

Incorporating Content Area Measures into Developmental Growth Analyses

Prior to 2010, North Carolina lacked an empirical (as distinct from extrapolative) basis for extending its average reading and mathematics growth curves into the secondary years. End-of-Grade (EOG) reading and mathematics tests were designed for Grades 3-8. The state had several End-of-Course (EOC) tests designed to measure student achievement in high school courses. However, EOC tests used different scales than EOG tests.

Beginning around 2008, the North Carolina Department of Public Instruction (NCDPI) psychometrically linked the English I EOC test with the Lexile Framework and linked the Algebra I, Geometry and Algebra II EOC tests with the Quantile Framework for Mathematics. This effectively brought the existing EOG and EOC reading and mathematics tests onto common scales (one for reading, one for mathematics).

Further realizing that several other content area tests had evolved to incorporate substantial amounts of reading material into the test items, the NCDPI questioned whether some measure of reading ability might be created from them by attempting a psychometric link with the Lexile Framework. MetaMetrics (2011) confirmed the feasibility of this strategy, effectively expanding measurement on a common scale well into the high school years. The next step was to investigate whether combining the high school EOC measures with the EOG measures would improve the estimation of statewide average growth curves for reading and mathematics.

Interest centered on knowing whether the NC average reading and mathematics growth curves could be empirically extended beyond Grades 3-8 by combining linked EOC measures with EOG measures in the estimation process. The success of the endeavor was judged by comparing the estimated average growth curves under two scenarios: a) the growth curve derived from jointly using EOG- and EOC-based measures; and b) the growth curve based on EOG-based measures alone. If successful, the addition of EOC-derived measures should have the following consequences:

1. The functional form should be maintained—i.e., the unconditional quadratic growth model should still be tenable.
2. The intercept should be relatively stable. For example, new intercepts should be within +/- 75L or +/- 75Q of the intercepts based on data from Grades 3-8 only.
3. Initial velocity should show modest decrease.
4. Curvature/deceleration should be reduced (trajectory is flattened somewhat).
5. Ideally, the estimated average quadratic growth curves should reach their maxima after Grade 12.
6. The extrapolation of the empirical growth curves to the end of Grade 12 (i.e., the predicted value of the average achievement at the end of Grade 12) should be aligned with terminal outcomes.

Postsecondary texts have a median difficulty of 1300L and an interquartile range that stretches from 1200L to 1380L. Algebra II/Integrated Mathematics 3 lessons have a median difficulty of 1220Q and an interquartile range extending from 1100Q to 1350Q. These values provide approximate targets for the last criterion above.

A suitable data set was created for the cohort of students who were eighth graders in 2006-07. First, the historical EOG reading scores from Grades 3-8 were assembled to produce an initial six-wave longitudinal panel spanning Grades 3-8; subsequently, EOC Lexile measures for English I, Biology, U.S. History, and Civics & Economics from 2008-2010 were also merged into the panel. The final reading panel contained ten waves of data which spanned the nine school years from spring 2002 to spring 2010. Similarly, EOG mathematics Quantile measures from Grades 3-8 were assembled to produce a six-wave panel of longitudinal data; subsequently, this panel was augmented with Quantile measures for Algebra I, Geometry and Algebra II for the same students. The final mathematics panel contained nine waves of data spanning 2002-2010.

Based on the results in Table 2 (on the following page), the strategy to improve the mathematics growth curve is completely successful. Compared to the six-wave analysis, the nine-wave analysis is superior because all six of the pre-specified criteria for improvement were met. The reading analysis is moderately successful, clearly meeting five of the six criteria. Although the maximum of the reading growth curve based on the ten-wave analysis is clearly improved relative to the maximum of the growth curve based on the six-wave analysis, the maximum occurs about half way between the end of Grade 11 and the end of Grade 12.

The two reading growth curves are displayed in Figure 4. The two mathematics growth curves are depicted in Figure 5. Both figures visually confirm the improvements obtained by incorporating additional longitudinal measures into the estimation of the respective growth curves.

Table 2

Fixed Effects Parameter Estimates: Six-Wave Growth Analyses vs. Growth Analyses That Incorporate Content Area Scores to Increase the Number of Waves of Longitudinal Data

	READING		MATH	
Parameter	Six Waves	Ten Waves	Six Waves	Nine Waves
N	101,530	101,610	75,601	101,650
Intercept	650.4	663.8	583.2	586.0
Velocity	167.7	148.0	136.9	100.6
Curvature	-12.7	-8.7	-11.3	-3.0
Grade @ Max	9.6	11.5	9.1	19.9
Predicted Grade 12	1133L	1287L	902Q	1250Q

Note. Except for the six-wave mathematics analysis (which was based on complete longitudinal records), the analyses included all students with at least one wave of data and any sequence of available End-of-Course (EOC) content-area measures. Six-wave analyses were based on End-of-Grade test scores in Grades 3-8. Additional waves of reading measurements were derived from English I, Biology, U.S. History, and Civics & Economics End-of-Course tests. Additional waves of mathematics measurements were derived from Algebra I, Geometry, and Algebra II End-of-Course tests.

Figure 4

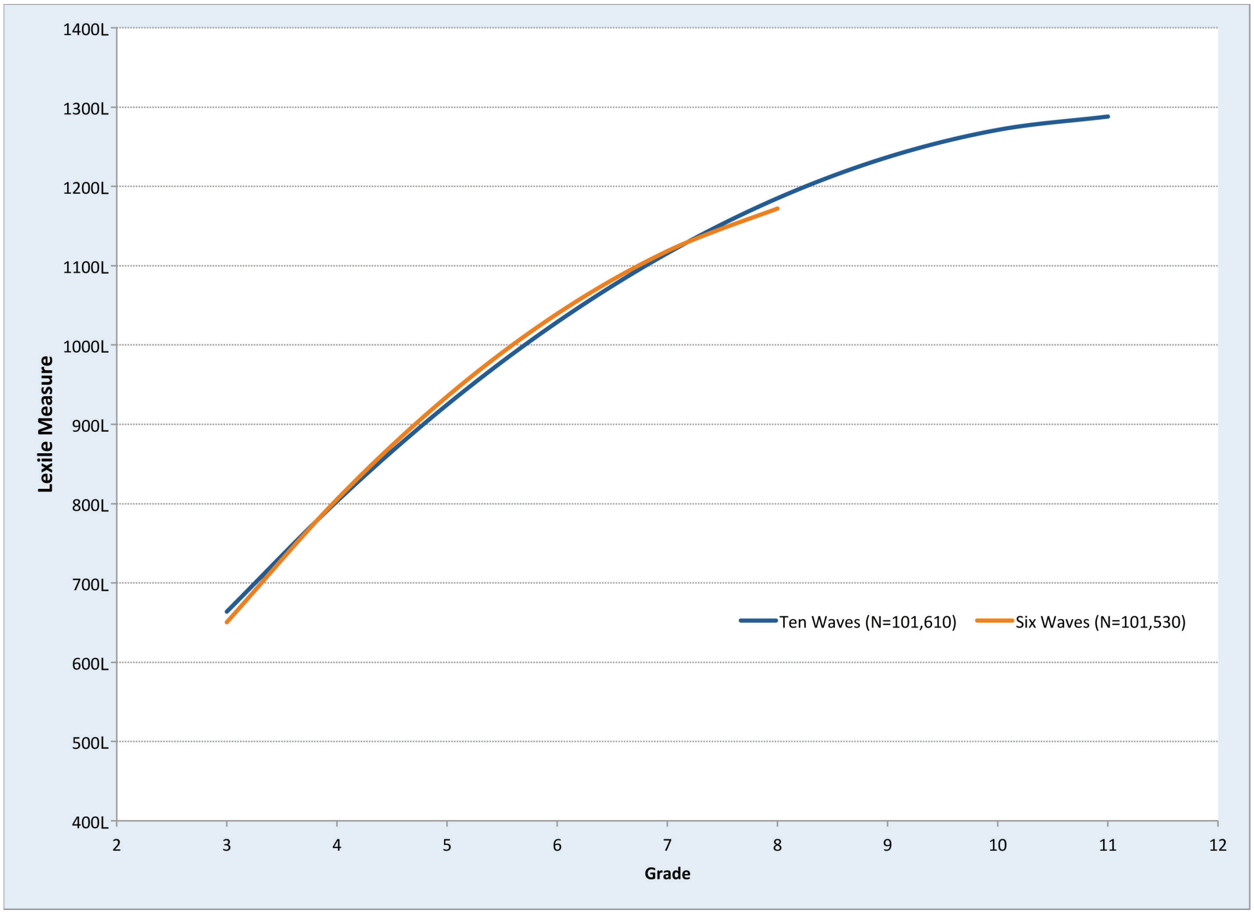
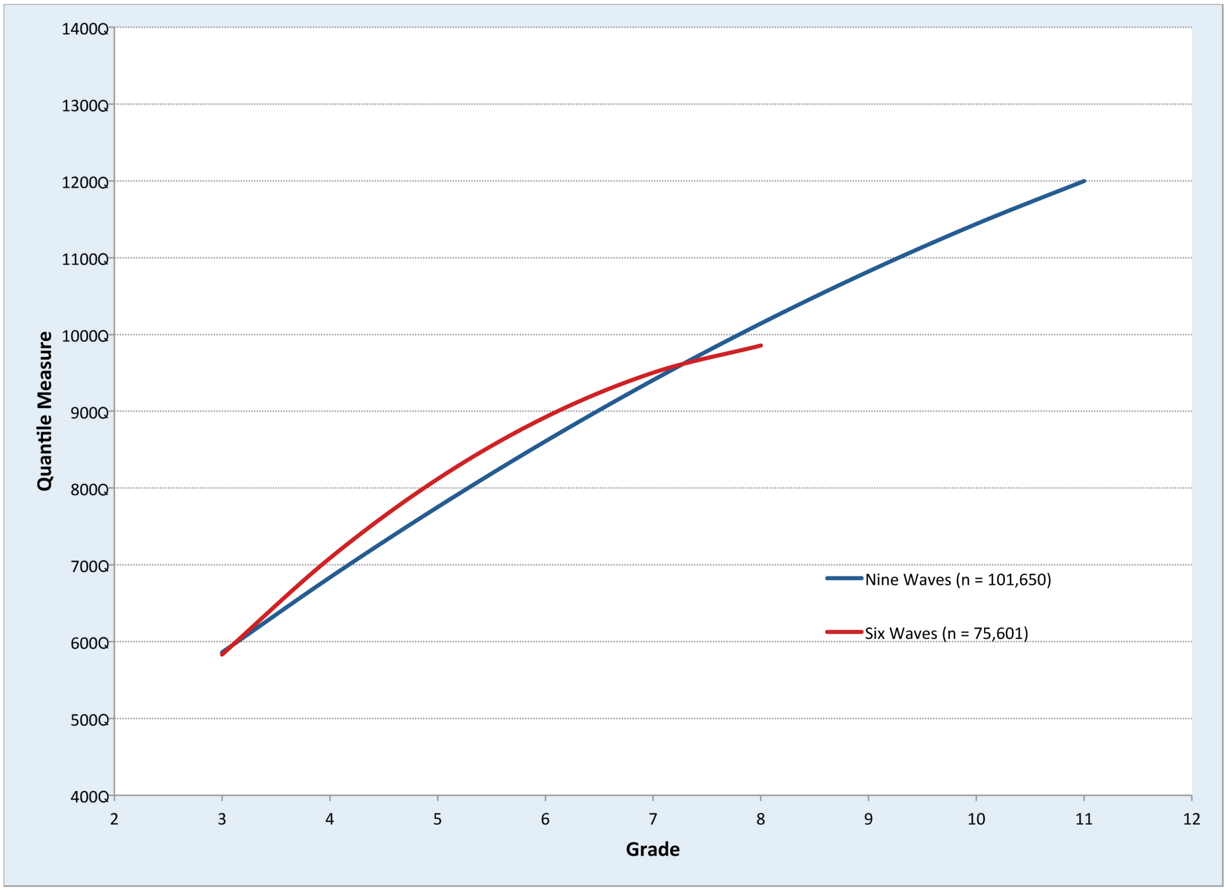


Figure 4. North Carolina reading growth curves based on End-of-Grade tests only (six waves) versus End-of-Grade tests and selected End-of-Course tests (English I, Biology, Civics & Economics, and U.S. History) (ten waves).

Figure 5



Panels 07/07E

Figure 5. North Carolina mathematics growth curves based on End-of-Grade tests only (six waves) versus End-of-Grade tests and End-of-Course tests (Algebra I, Geometry, Algebra II) (nine waves).

MetaMetrics

Incremental Velocity Norms for Average Reading and Mathematics Growth

Replicating or exceeding some specified previous student achievement level was the basis for educational expectations throughout most of the 20th century. Similarly, replicating or exceeding previous growth rates eventually emerged as a basis for student growth standards (NCDPI, 2000). Even so, the best implementation of educational growth standards to date has been based on year-to-year gains, without the benefit of an underlying longitudinal growth curve. Growth velocity norms did not emerge even for height or weight until the work of Tanner, Whitehouse, and Takashi (1966) in the United Kingdom and later in the United States (Roche & Himes, 1980; Baumgartner, Roche, & Himes, 1986). In this next example, I use the parametric models for growth summarized in the preceding example to create incremental growth velocity norms for average reading and mathematics growth. The approach yields not only estimates of year-to-year gain, but estimates of growth between any two points within the design time frame.

The starting point is the realization that an historical aggregate growth curve provides a long-term summary of observed growth for a group of students. As such, it may be regarded as a norm for growth. If this norm were treated as a growth expectation for future panels of students, the implicit policy goal would be that future students *should* grow in a manner that is similar to previous historical growth. When regarded as a set of expectations for future growth, the growth curve represents a *growth standard*. Perhaps the easiest way to operationalize such a growth standard is by generating incremental growth velocity estimates from the average growth curve. It is relatively easy to do this. One needs only the parameter estimates for the average growth curve. I use the ten-wave analysis for reading growth and the nine-wave analysis of mathematics growth from the previous example. These two growth curves are salient because they each span Grades 3-11, the grades during which accountability assessments are most often implemented in the United States and the grades most often the object of state accountability systems.

The estimated average reading growth curve is a function of time, $r(T) = 663.8 + 148.0\,T - 8.7\,T^2$. I can use it to estimate the expected amount of growth from one time point to another. For purposes of the example, let us interpret the time scale in terms of *grade in school* with the understanding that the gains so calculated will represent the growth from one spring to another because testing took place at the end of the school year.

When I calculate the gain between adjacent grades, I have calculated the amount of change per unit of time—i.e., the incremental velocity. When I calculate the gain between any two grades more than one year apart, it produces an incremental estimate of the amount of growth that took place between those two grades.

In Table 3, I have tabulated the values of $r(k) - r(j)$ for all pairs of grades (j,k) such that $k > j$ where $j = 3, 4, \ldots, 10$ and $k = 4, 5, \ldots, 11$. The resulting values are displayed in matrix form. Quantities along the diagonal represent the expected gain for each year-to-year transition: Grade 3 to Grade 4, Grade 4 to Grade 5, and so on. These are the incremental yearly, spring-to-spring growth velocity norms based on this sample of 101,610 students. The off-diagonal elements of the table display the amount of growth between every possible pair of grades. This information is useful because it captures longer-term growth expectations, spanning multiple grades.

To illustrate the interpretation of growth using Table 3, first consider the annual yearly growth expectations displayed along the diagonal. A fourth-grade teacher might reference the entry at the intersection of the row for Grade 3 and the column for Grade 4. The entry conveys the expectation for average reading growth between the end of Grade 3 and the end of Grade 4—namely during the fourth grade year. It is 139L. Similarly, the fifth-grade teacher would reference the entry at the intersection of the row for Grade 4 and the column for Grade 5 and learn that the average growth expected of fifth graders is 122L. The principal of a middle school serving students in Grades 6-8 would be interested in the total gain expected between the end of the fifth grade and the end of the eighth grade. Referring to the intersection of the row for Grade 5 and the column for Grade 8, the principal learns that the expectation for average reading growth for students who spend all three years at the middle school is 260L.

Similarly, the average mathematics growth curve from the preceding example can be expressed as: $m(t) = 586.0 + 100.6\ T - 3.0\ T^2$. Having evaluated the average mathematics growth curve at all grade-pairs, I displayed the results in Table 4. The interpretation of average mathematics growth in Table 4 follows in the same manner as for reading growth (Table 3).

In both Table 3 and Table 4 it is obvious that historical growth is typically greater in earlier grades and tapers off as grade increases. This is apparent as one scans along the diagonal from upper left to lower right. This pattern reflects the deceleration of growth and quantifies it in practical terms for educators. However, the off-diagonal entries in the table reinforce the realization that long-term growth is the result of a cumulative growth process that endures across the developmental life span.

Table 3

Incremental Velocity Norms for Average Reading Growth in Grades 3-11 Denominated in Lexile Scale Units

Student Achievement Estimated from the Average Reading Growth Curve									
	664L	803L	925L	1029L	1116L	1185L	1237L	1271L	1288L
Grade	3	4	5	6	7	8	9	10	11
3		139L	261L	365L	452L	521L	573L	607L	624L
4			122L	226L	313L	382L	434L	468L	485L
5				104L	191L	260L	312L	346L	363L
6					87L	156L	208L	242L	259L
7						69L	121L	155L	172L
8							52L	86L	103L
9								34L	51L
10									17L

Note. The table is based on an average reading growth curve (ten waves of measurement) for North Carolina students (n = 101,610), spanning Grades 3-11 during the years 2002-2010. The fitted model is summarized by the equation: $E(L|T) = 663.8 + 148.0\ T - 8.7\ T^2$ where the time scale is centered at grade 3 (i.e., T = Grade -3). Velocity increments for adjacent grades (i.e., year-to-year gains) are shown in the shaded diagonal.

Table 4

Incremental Velocity Norms for Average Mathematics Growth in Grades 3-11 Denominated in Quantile Scale Units

Student Achievement Estimated from the Average Mathematics Growth Curve									
	586Q	684Q	775Q	861Q	941Q	1014Q	1082Q	1144Q	1200Q
Grade	3	4	5	6	7	8	9	10	11
3		98Q	189Q	275Q	355Q	428Q	496Q	558Q	614Q
4			92Q	177Q	257Q	331Q	399Q	460Q	516Q
5				86Q	165Q	239Q	307Q	369Q	425Q
6					80Q	153Q	221Q	283Q	339Q
7						74Q	142Q	203Q	259Q
8							68Q	130Q	185Q
9								62Q	118Q
10									56Q

Note. The table is based on an average mathematics growth curve (nine waves of measurement) for North Carolina students (n = 101,650), spanning Grades 3-11 during the years 2002-2010. The fitted model is summarized by the equation: $E(Q|T) = 586.0 + 100.6\ T - 3.0\ T^2$ where the time scale is centered at Grade 3 (i.e., T = Grade -3). Velocity increments for adjacent grades (i.e., year-to-year gains) are shown in the shaded diagonal.

Summary

In this paper, I presented a series of simple examples to illustrate novel interpretations of student growth that may have relevance to educators. In three examples, I illustrated how to interpret student reading achievement and growth in light of the text complexity associated with reading materials that students may encounter during schooling or in the postsecondary world. In another example, I demonstrated a practical strategy for educators to lengthen empirical growth trajectories by using a common scale for developmental tests and content area assessments. In the final example, I proposed a strategy to create incremental velocity norms for average academic growth and provided examples of velocity norms for reading growth and for mathematics growth, each based on over 100,000 students.

The first three examples highlighted the power of conjoint measurement when combined with the longitudinal perspective of student growth curves. We first saw how to compare student growth to changing text complexity requirements such as those expressed in the CCSS. Then, we saw a student growth curve juxtaposed with postsecondary text requirements and I suggested that alignment between the two is desirable. Next, we saw how the first two examples lead us to forecasted comprehension rates for readers who are themselves growing in their reading ability. Although these three examples featured reading ability relative to text complexity requirements, it is possible to provide similar examples for growth in mathematics ability relative to the complexity of mathematical skills and concepts.

The findings from the fourth example confirmed that North Carolina could improve its estimation of average reading and mathematics growth curves by combining the Lexile (or Quantile) measures from EOC tests with the Lexile (or Quantile) measures from EOG tests to create longer panels of longitudinal data. The Lexile measures derived from the EOC tests were consistent with developmental expectations in the sense that: a) they generally reflected the higher reading comprehension that results from additional instruction and study; and b) their inclusion modified the average growth curve accordingly. The fact that additional waves of longitudinal measures facilitated the extension of the average growth curve is consistent with the assumption that reading comprehension and mathematical understanding can be treated as unidimensional constructs[ii] manifested from Grade 3 through Grade 11.

Perhaps the most remarkable aspect of the growth analyses is that measurement of student growth was improved by combining a Rasch measurement model with construct theory to effect a scale with general objectivity. In practical terms, the implication is that the study of academic growth can be facilitated by the use of both developmental and non-developmental measures that are linked to a common developmental scale.

Finally, we saw how improving the estimation of growth curves can strengthen the basis for setting growth standards based on longitudinal panel data, rather than the usual practice of setting year-to-year growth standards based on non-developmental (e.g., status projection) or short-term growth (e.g., gain score) formulations. Velocity norms such as those presented here are an indispensable complement to traditional cross-sectional norms for interpreting student achievement and growth because they a) base year-to-year gains on a longer-term growth curve and b) they make it possible to construct expectations of growth between any pair of grades.

Although, the growth velocity norms provided in this paper are for statewide average growth, they are easily extended to sub-populations. To briefly elaborate, one possibility for expanding growth standards is to disaggregate an historical average curve into multiple growth curves conditioned on initial status. For example, by grouping students into deciles based on initial performance, average growth curves can be estimated for each of the ten deciles. Once decile growth curves have been determined, incremental velocity norms can be established for each decile group simply by replicating Table 3 (or 4) for each group's aggregate growth curve. Conditioning growth standards on initial performance is a feature that has been desired in some accountability systems.

Similarly, if common scales were universally used for educational constructs and longitudinal data were routinely collected and analyzed, then growth velocity standards could have even greater generalizability. Individual state norms, national norms, perhaps even international norms for academic growth velocity would become possibilities.

Endnotes

[i] The median difficulty (1130L) of texts used near the end of high school (i.e., Grades 11 and 12) is not shown in the figure, because it does not represent a postsecondary aspiration. High school texts are significantly easier to read on average than are citizenship materials, workplace materials, community college texts, or university texts (Williamson, 2008).

[ii] By incorporating the End-of-Course (EOC) content area measures into the estimation of the growth curve, I assumed that I could treat the EOC measures as if they were interchangeable with each other and also as if they were interchangeable with the End-of-Grade (EOG) measures, in the sense that they are measures of a common construct, either reading comprehension or mathematical understanding. A priori there was no guarantee that I could expect this behavior because the EOC tests were developed independently of the EOG tests and were not designed to be on a developmental scale. Placing EOG and EOC tests on common scales (either the Lexile Framework for Reading or the Quantile Framework for Mathematics) made it possible to use the linked scores in this manner.

References

Baumgartner, F. N., Roche, A. G., & Himes, J. H. (1986). Incremental growth tables: Supplementary to previously published charts. *The American Journal of Clinical Nutrition, 43,* 711-722.

Bazemore, M., & Van Dyk, P. B. (2004). *North Carolina reading comprehension tests: Technical report* (Ed.2). ("Citable Draft"). Raleigh, NC: North Carolina Department of Public Instruction.

Bormuth, J. R. (1969). *Development of readability analyses*. Final Report, Project No. 7-0052, Contract No. OEG-3-7-070052-0326, Office of Education, Bureau of Research, U.S. Department of Health, Education and Welfare.

Bruning, R. (1985). Review of Degrees of Reading Power. *The Ninth Mental Measurements Yearbook, Vol. I,* 443-444.

Goldstein, H. (1979). *The design and analysis of longitudinal studies*. New York: Academic Press.

Koslin, B. L., Zeno, S., & Koslin, S. (1987). *The DRP: An effectiveness measure in reading*. New York: The College Entrance Examination Board.

Legendre, A. M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris: Courcier.

MetaMetrics. (2009). *The Quantile Framework for Mathematics: Linking assessment with mathematics instruction*. Durham, NC: Author. Retrieved from https://s3.amazonaws.com/quantile-resources/resources/downloads/static/BriefingDocument.pdf

MetaMetrics. (2011). *NC EOC Feasibility Studies Research & Development Report*. Durham, NC: Author.

National Governors Association Center for Best Practices (NGA Center) & the Council of Chief State School Officers (CCSSO). (2010). *Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects*. Washington, DC: Author. Retrieved from http://www.corestandards.org/wp-content/uploads/ELA_Standards.pdf

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. Washington, DC: Council of Chief State School Officers (CCSSO). Retrieved on June 11, 2012 from: http://www.ccsso.org/Documents/2012/Measures%20ofText%20Difficulty_final.2012.pdf

No Child Left Behind Act of 2001, PL 107-110 (2002).

North Carolina Department of Public Instruction. (2000, June). Setting annual growth standards: "The formula" (Accountability Brief, Vol. 1, No. 2). Raleigh, NC: Author.

North Carolina Department of Public Instruction. (2009). *North Carolina Reading Comprehension Tests: Technical report* (Ed. 3). Raleigh, NC: Author. Retrieved from http://www.ncpublicschools.org/docs/accountability/testing/reports/eogreadingtechman3.pdf

Roche, A. F., & Himes, J. H. (1980). Incremental growth charts. *The American Journal of Clinical Nutrition, 33*, 2041-2052.

Sanford, E. E. (1996). *North Carolina End-of-Grade tests technical report # 1: Reading comprehension, mathematics*. Raleigh, NC: Department of Public Instruction.

Sanford-Moore, E. E., Williamson, G. L., Bickel, L., Koons, H., Baker, R. F., & Price, R. (2014). *A quantitative task continuum for K-12 mathematics* (MetaMetrics Research Brief). Retrieved from https://s3.amazonaws.com/quantile-resources/resources/downloads/static/Quantile+Framework+Math+Continuum+Research+Brief+2014.pdf

Scholastic, Inc. (2007). *Scholastic Reading Inventory™ technical guide*. New York: Author.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Stenner, A. J., Burdick, H., Sanford, E., & Burdick, D. S. (2007). *The Lexile Framework for Reading technical report*. Durham, NC: MetaMetrics, Inc. Retrieved from http://www.lexile.com/research/9/

Stenner, A. J., & Smith, M. (1982). Testing construct theories. *Perceptual and Motor Skills, 55*, 415-426.

Stenner, A. J., Smith, M., & Burdick, D. S. (1983). Toward a theory of construct definition. *Journal of Educational Measurement, 20*, 305-316.

Stenner, A. J., Smith, D. R., Horabin, I., & Smith, M. (1987). Fit of the Lexile theory to sequenced units from eleven basal series. Durham, NC: MetaMetrics. Retrieved from https://lexile-website-media-2011091601.s3.amazonaws.com/m/resources/materials/Stenner_Smith_Horabin_Smith-_Fit_of_the_Lexile_Theory_to....pdf

Tanner, J. M., Whitehouse, R. H., & Takaishi, M. (1966). Standards from birth to maturity for height, weight, height velocity and weight velocity: British children, 1965. *Archives of Disease in Childhood, 41*, 454-471, 613-635.

Williamson, G. L. (2008). A text readability continuum for postsecondary readiness. *Journal of Advanced Academics,19*(4), 602-632.

Williamson, G. L. (2014). *Measuring and modeling individual academic growth: Methodological foundations for educational applications in the 21st century*. Manuscript in preparation.

Williamson, G. L., Fitzgerald, J., & Stenner, A. J. (2014). Student reading growth illuminates the Common Core text-complexity standard: Raising both bars. *Elementary School Journal, 115*(2), 230-254.

About The Author

GARY L. WILLIAMSON, PH.D., is Senior Research Scientist at MetaMetrics, Inc. where he engages in psychometric research and development in support of public education. Williamson received his Ph.D. from Stanford University in Mathematical Methods in Educational Research. In addition, he holds degrees in Statistics (MS, Stanford University), Educational Research & Evaluation (MEd, University of North Carolina at Greensboro), and Mathematics (BS, University of North Carolina at Chapel Hill). His academic specializations include research methodology, statistics, and psychometrics. Currently, his research interests are focused on promoting a life-span developmental approach to setting academic growth standards for public education.

www.MetaMetricsinc.com
www.Lexile.com
www.Quantiles.com