

Expressing Academic Growth on a Conjoint Developmental Scale with General Objectivity

G L Williamson

Senior Research Scientist, MetaMetrics, 1000 Park Forty Plaza Drive, Suite 120,
Durham, NC 27713, USA

E-mail: gwilliamson@lexile.com

Abstract. Individual growth curves yield insights about growth that are not available from any other methodology; and, developmental scales based on conjoint measurement models provide unique interpretive advantages for investigations of academic growth. Benefits are demonstrated in three examples. First, a series of fifteen statewide reading growth curves is annotated with historical policy actions related to assessment, accountability and early interventions. Second, a common measurement framework simultaneously addresses five interpretive perspectives—student reading growth; achievement level standards; K-12 text complexity standards; postsecondary reading demands; and, occupational reading demands. Third, incremental velocity norms are introduced for average reading growth based on a parametric mathematical model for individual growth curves.

1. Perspective

There are five essential properties for the optimal measurement of student growth. The measurement scale must be *unidimensional*; *continuous*; *developmental*; *equal-interval*; and, must have *invariance of scale unit*. The first four properties can be achieved through item response theory [1]. Invariance of scale unit can be attained when a Rasch-based measurement scale is anchored by means of a construct specification theory. See [2] for a detailed discussion of these principles.

As of this writing, there are only two conjoint measurement scales for reading ability that have all five properties mentioned above. One of them, The Lexile[®] Framework for Reading [3], was adopted in North Carolina (NC) around two decades ago as a supplemental scale for its assessment program. This has yielded unprecedented interpretations of reading growth featured in the Results section.

2. Approach

The “data” for the examples were drawn from multiple traditions of previous research. One tradition produced distributions of text-complexity measures for reading materials characteristic of the public schools [4] and postsecondary life in the USA [5, 6]. Another tradition produced reading achievement measures for individuals who attended public schools in the state of North Carolina [7, 8, 9]. The third tradition of research consisted of multilevel growth models applied to longitudinal data [2, 10].

3. Results

3.1 Student growth with historical policy annotations

Aggregate growth curves for 15 successive panels of students are featured in Figure 1 using a common scale for reading achievement. The growth curves are ordered chronologically from left (earlier) to



right (more recent). Each growth curve spans Grades 3-8 in different years, noted along the horizontal axis. Each panel consists of students who progressed from grade to grade without repeating a grade and who had complete data across all six occasions.

In addition to the displayed growth curves, Figure 1 is annotated schematically with historical policy actions that took place during the time frame (1993-2014). There are three categories of operational policy initiatives: a) assessment editions; b) accountability programs; and, c) early intervention reading initiatives. A few comments about the three categories will facilitate understanding the subsequent discussion of results.

There have been four editions of NC End-of-Grade (EOG) reading assessments administered statewide in Grades 3-8—Edition 1 (1993-2002); Edition 2 (2003-2007); Edition 3 (2008-2012); and a fourth edition, called NC READY EOG (2013-present). These editions are annotated across the top of Figure 1. Each edition had its own unique vertical scale. Three of the four scales were directly linked to the Lexile scale via symmetric linking functions. However, Edition 2 was not directly linked to the Lexile scale, though it was linked to Edition 1 (equi-percentile method) and Edition 3 (Stocking-Lord method). These linking studies provided the means to translate all scores to the Lexile scale.

Six waves of accountability initiatives occurred during the time frame and are annotated from left to right in the mid-section of Figure 1. From 1993-1996 institutional accountability was focused on cross-sectional analyses of school district average performance (Senate Bill 2). In school year 1995-96, the state piloted a new accountability system in ten local education agencies using the school building as the unit of analysis for institutional accountability. The program, dubbed the *ABCs*, was first implemented statewide in elementary schools in 1996-97 and additionally in high schools the following year. The signal feature of the *ABCs* was that unique growth expectations were set for each school and growth was based on the performance of the same students over time. There were two generations of the *ABCs*, distinguished by different methods (gain-score versus status-projection) for calculating growth. The current generation of school-level accountability, NC READY, was initiated in 2012-13. Under NC READY, the Education Value-Added Assessment System (EVAAS) provides school value-added growth results. Finally, individual accountability became effective for students in Grades 3 and 8 in 2002. Student accountability standards required that individual students must score at or above a grade-level cut score in order to be promoted to the subsequent grade.

Two early-intervention initiatives were implemented during the time frame for Figure 1. They are annotated just above the horizontal axis. In 1993, SMART START was enacted to provide educational resources for four-year-olds to better prepare them to enter Kindergarten at age five. SMART START was gradually implemented. By 1998, it had been deployed in 55 of the state's 100 counties. By 2001, SMART START was fully deployed statewide. Also in 2001, a second early intervention program called *More at Four* was added to provide greater support for four-year-olds.

Each growth curve in Figure 1 visually manifests three features of growth—magnitude, velocity and acceleration. The overall elevation of each curve in relation to the vertical axis reflects achievement status on each occasion (i.e., the magnitude of growth). The slope of a tangent to a curve quantifies the instantaneous velocity of growth at a particular time. The curvature of each trajectory connotes the acceleration of growth. The general impression is that growth was fairly consistent from panel to panel, characterized by positive initial velocity and deceleration over time. Furthermore, there was systemic improvement in the magnitude of growth over time as evidenced by the general rise of the curves moving from left to right in chronological sequence. Other insights are revealed by considering the three categories of policy initiatives.

In relation to changing assessment editions, an important insight is that the general shape of the growth curves is robust and consistent across time, even though four successive editions of assessments were used. Capturing this consistency of growth is possible because all four editions of reading tests were calibrated to the same (Lexile) scale.

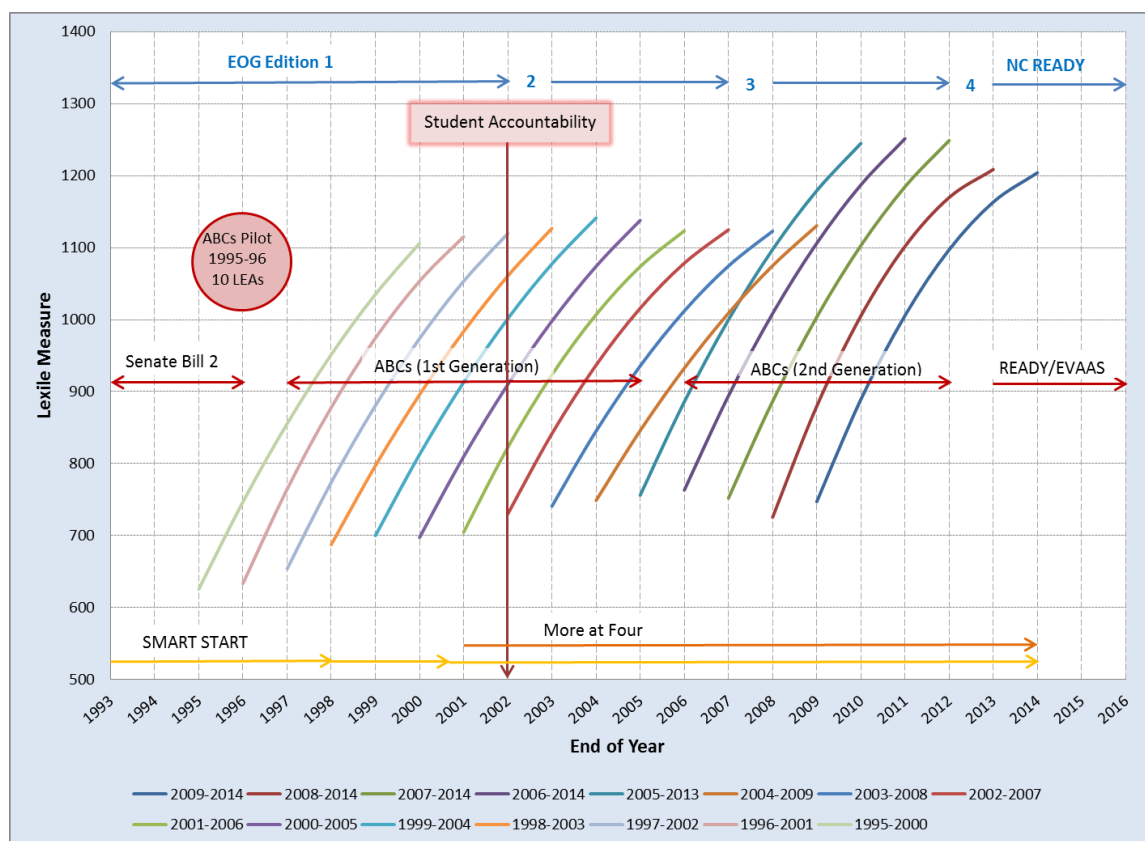


Figure 1. North Carolina (NC) growth curves ($n = 1,012,118$) annotated with assessment, accountability and early intervention initiatives

When growth curves and policy initiatives are viewed together in Figure 1, there are interesting correspondences that suggest how policy actions *may* have influenced changes in growth patterns over time. Although, I offer some thoughts on possible causal hypotheses, they are limited to North Carolina accountability and early intervention initiatives. For this reason, I must emphasize that what I present is not a demonstration of cause and effect. Rather, I am merely illustrating how one might identify potential causal hypotheses that could be relevant to interpreting growth and performance.

Let the reader scan Figure 1 from left to right and notice the beginning points of the growth curves. There seems to be an uptick in Grade 3 performance in 1997 and initial performance continues to rise for several years before leveling off around 2000. This uptick in the starting point for growth is interesting because 1997 was the first year that ABCs growth expectations applied to all elementary schools and it is also the year when the first students to experience SMART START reached Grade 3.

There seems to be another uptick in initial status of growth in spring 2002 and again several years when initial status continued to rise. This is interesting because student accountability standards first applied to Grade 3 in 2002. Also, the panel that began in 2002 would have contained larger numbers of students who benefited from SMART START because the four-year-olds from 1998 (when 55 of 100 counties had the program) would have reached the end of Grade 3 in spring 2002.

Now, observe the end-points of the growth curves (i.e., student performance at the end of Grade 8) for the first six panels. Eighth-grade performance was gently rising for the first three panels, possibly due to the ABCs. Then eighth grade performance appears to increase more rapidly across the next three panels after the introduction of student accountability standards for eighth-graders in 2002.

Eighth-grade achievement peaked in 2004 and 2005. The subsequent four panels exhibited lower Grade 8 performance due to increased deceleration of growth. This was likely a compound result of

slowly declining reading growth in middle schools (particularly sixth grade) and the subsequent adoption of revised growth standards for the ABCs (2nd generation) effective in 2006.

Other interesting scenarios can be constructed for Figure 1. Due to space limitations, I mention only one more—the fact that the last five growth curves seem to be different from the first ten growth curves, primarily in terms of their overall length. This can be traced, in part, to a methodological artifact. For the first ten panels, Edition 2 scores were translated to the Lexile scale after applying the equi-percentile link between Edition 2 and Edition 1. For the last five panels, Edition 2 scores were translated to the Lexile scale after applying the Stocking-Lord link between Edition 2 and Edition 3. Though this may not be the only explanation, it does account for some of the observed difference in the length of the growth curves. Whether a methodological artifact or some combination of factors is ultimately the explanation, one point cannot be ignored. Placing all scores on a common scale may help identify anomalies as well as intentional changes in the features of growth.

3.2 Five perspectives for growth

In Figure 2, I display the earliest growth curve from the previous example and a recent growth curve spanning Grades 3-11 along with four additional background contexts: a) student reading achievement level standards; b) K-12 text-complexity standards; c) postsecondary text complexity; and, d) reading demands associated with specific occupations. Figure 2 originally appeared in *The Education Standard* [11] with detailed discussion; the authors used it to illustrate how states can assure that student growth is well-aligned with achievement standards and postsecondary expectations associated with college and career readiness. I incorporate the figure here to illustrate how conjoint measurement provides unique insights and interpretive perspectives for student growth and academic performance.

The rectangle located in the middle of Figure 2 represents the interquartile range of postsecondary text complexity and the diamond inside the rectangle represents the median [5]. The dots on the right-hand side of Figure 2 represent the median text complexity of reading materials associated with 59 *Bright Outlook Occupations* [12]. The dashed line segments on the left-hand side of the figure represent text complexity standards widely adopted by states for use in Grades K-12. Finally, the hash marks at Grades 3-8 represent rising student achievement standards adopted in North Carolina.

Looking at Figure 2, it is immediately obvious that the state raised achievement standards substantially over twenty years. As standards were raised they were increasingly aligned with the text requirements associated with postsecondary contexts. Most importantly, the average student growth curves also rose over time, keeping track with increasing policy and postsecondary reading requirements. These disparate perspectives are unified by the use of a common scale with conjoint measurement of persons and texts.

3.3 Incremental velocity norms

For my final example, I reproduce a table which appears in [13]. I used a parametric growth model to construct incremental velocity norms for academic growth from the end of Grade 3 to the end of Grade 11. The results are displayed in Table 1. Along the diagonal, the table contains the spring-to-spring gains in reading achievement between adjacent grades. Thus a fifth-grade teacher who wishes to know how much growth is typical during Grade 5 only needs to refer to the entry at the juncture of the row for Grade 4 and the column for Grade 5 to discover the answer—specifically, 122L. The table also contains gains between any pair of grades in the time frame. So for example a middle school principal serving Grades 6-8 can ascertain that the amount of growth exhibited by students during those three years was 260L.

Cross-sectional status norms have been the standard in education for the last 100 years. Table 1 provides growth norms based on longitudinal panel data—the same students are reflected across the entire time frame from the end of Grade 3 to the end of Grade 11. Table 1 thus fills a void for educators who wish to have defensible growth standards as a reference for interpreting aggregate student growth over time. Table 1 provides estimates of the amount of developmental reading growth between any pair of years spanned by the normative growth curve.

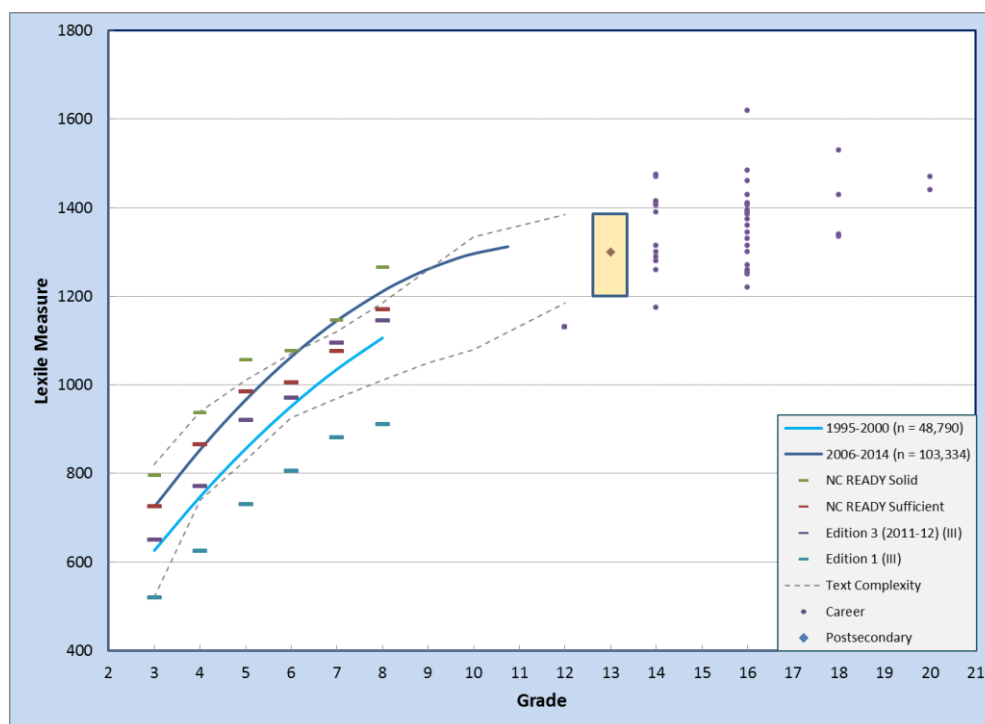


Figure 2. North Carolina reading growth and achievement standards (1995-2014) relative to K-12 text-complexity standards, postsecondary text requirements and text complexity for 59 Bright Outlook Occupations. Adapted from “Aligning the K-12 academic journey with a postsecondary destination,” by M Smith and G L Williamson 2016 *The State Education Standard* **16**(1) p 16. Copyright 2016 by the National Association of State Boards of Education (NASBE).

Table 1

Incremental Velocity Norms for Average Reading Growth 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. Based on average reading growth for North Carolina students ($n = 101,610$), 2002-2010. The fitted model is given by the equation: $E(L|T) = 663.8 + 148.0 T - 8.7 T^2$ (where $T = \text{Grade} - 3$). Velocity increments for adjacent grades (i.e., year-to-year gains) are shown in the shaded diagonal. Adapted from “Novel interpretations of academic growth,” by G L Williamson 2016 [Manuscript submitted for publication]. Copyright 2016 by G L Williamson.

4. Summary

The three examples presented here serve to illustrate the power of educational measurement scales that are designed to measure student growth with general objectivity. Developmental analyses of growth benefit from maintaining such scales over long periods of time because growth norms are improved and relationships between features of growth and policy contexts become more apparent. Additionally, with conjoint measurement, student reading growth and performance are easily interpreted in relation to the complexity of texts that students read.

References

- [1] Yen W M and Fitzpatrick A R 2006 *Item response theory* (Educational measurement 4th ed) ed R L Brennan (Westport, CT: American Council on Education and Praeger Publishers)
- [2] Williamson G L 2015 Measuring academic growth contextualizes text complexity *Pensamiento Educativo: Revista de Investigación Educativa Latinoamericana* **52**(2) pp 98-118.
- [3] Stenner A J, Burdick H, Sanford E and Burdick D S 2007 *The Lexile Framework for Reading technical report* (Durham, NC: MetaMetrics) Retrieved from <http://www.lexile.com/research/9/>
- [4] Williamson G L, Koons H, Sandvik T and Sanford-Moore E 2012 *The text complexity continuum in grades 1-12* Research Brief (Durham, NC: MetaMetrics)
- [5] Stenner A J, Sanford-Moore E and Williamson G L 2012 *The Lexile® framework for reading quantifies the reading ability needed for "College & Career Readiness"* Research Brief (Durham, NC: MetaMetrics)
- [6] Williamson G L 2008 A text readability continuum for postsecondary readiness *J. Advanced Academics* **19**(4) pp 602-632 doi: 10.4219/jaa-2008-832
- [7] Bazemore M and Van Dyk P B 2004 *North Carolina reading comprehension tests: Technical report* (2nd ed) (Citable draft) (Raleigh, NC: North Carolina Department of Public Instruction)
- [8] North Carolina Department of Public Instruction 2009 *North Carolina Reading Comprehension Tests: Technical report* (3rd ed) (Raleigh, NC: Author) Retrieved from <http://www.ncpublicschools.org/docs/accountability/testing/reports/eogreadingtechman3.pdf>
- [9] Sanford E E 1996 North Carolina end-of-grade tests technical report # 1: *Reading comprehension, mathematics* (Raleigh, NC: Department of Public Instruction)
- [10] MetaMetrics 2011 *NC EOC feasibility studies research & development report* (Durham, NC: Author)
- [11] Smith M and Williamson G L 2016 Aligning the K-12 academic journey with a postsecondary destination *The State Education Standard*, **16**(1) pp 14-18
- [12] Williamson G L and Baker R F 2013 *Enriching the concept of career preparedness by examining text complexity associated with Bright Outlook Occupations* Research Brief (Durham, NC: MetaMetrics)
- [13] Williamson G L 2016 Novel interpretations of academic growth [Manuscript submitted for publication]