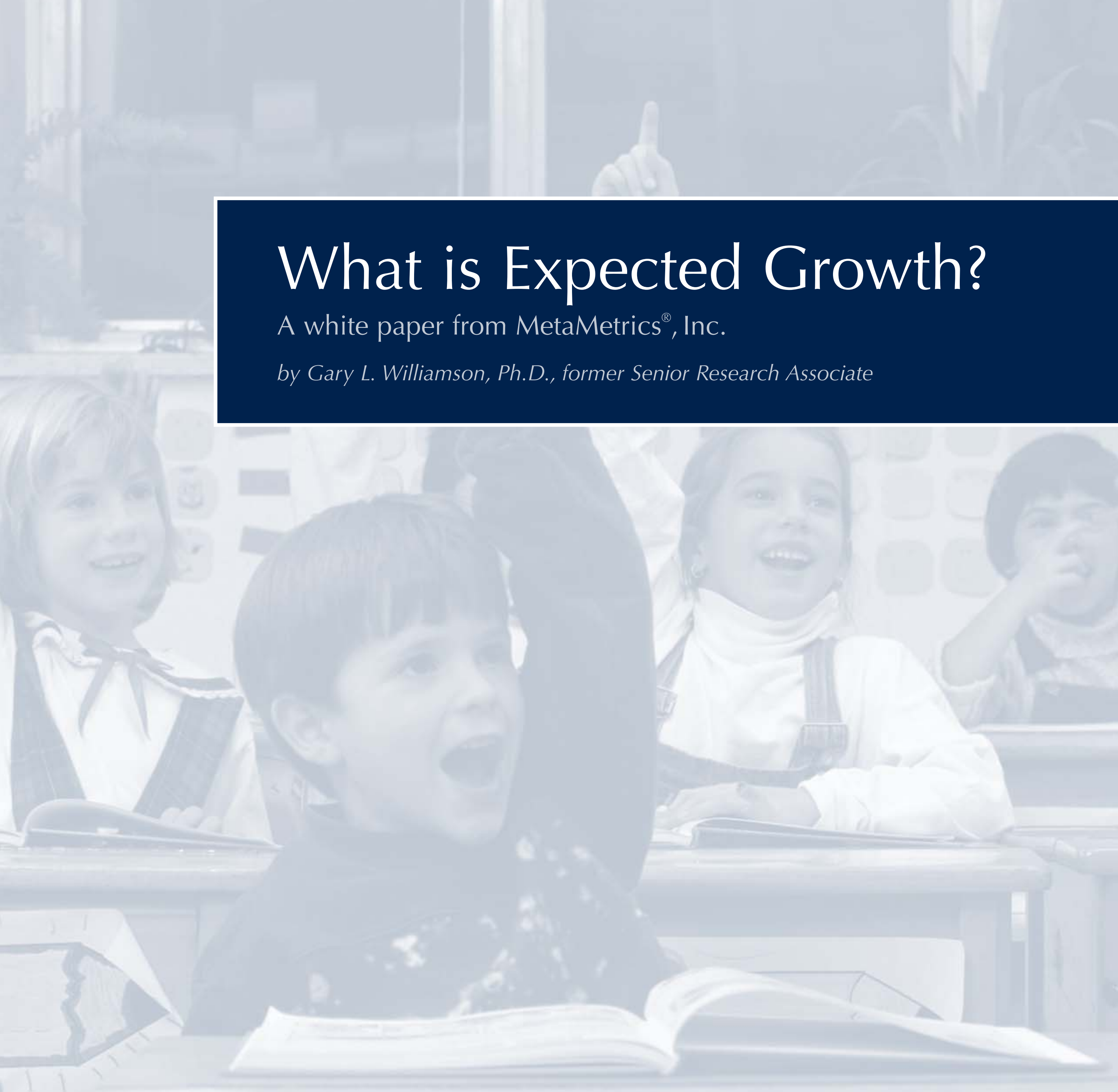


What is Expected Growth?

A white paper from MetaMetrics®, Inc.

by Gary L. Williamson, Ph.D., former Senior Research Associate

LEXILE® QUANTILE®

Table of Contents

Overview

We are all familiar with children, either through knowing our own or through acquaintance with those of other people. Perhaps no other thing in life is as obvious as the dramatic way that human beings develop and grow. Our key social and political institutions devote a significant part of their resources to ensuring that children grow and learn to function as productive citizens. Growth and learning are central to the mission of our country's public schools.

In January 2002 the President signed into law a major reauthorization of the Elementary and Secondary Education Act (ESEA) that has become known as the No Child Left Behind (NCLB) Act of 2001. The law established sweeping new requirements for educational measurement and accountability for all schools. Not surprisingly, the focus is on the academic achievement and progress of students. These terms (achievement, progress) and related ones (learning, growth, development, performance, proficiency, etc.) occur over 1,660 times in the text of the 670-page law. Setting goals for student performance and monitoring the progress that students make toward those goals are at the heart of the new federal accountability requirements.

NCLB prescribed one way of setting goals and monitoring student progress. States have worked diligently since its enactment to comply with the law and to integrate their efforts within already existing accountability frameworks. In 2005, the U.S. Secretary of Education, Margaret Spellings, created an opportunity for some flexibility when she invited states to propose growth models as part of their strategy to address the requirements of NCLB.

Because there are a number of alternative ways to conceptualize student growth and to measure it, states face a challenge to design and implement accountability systems that address a variety of information needs and still comply with state and federal laws. In this context, there are naturally many viewpoints about how best to conceptualize and measure student growth and to set appropriate goals for growth. This makes it especially important for students, parents and educators to better understand student growth, how it is measured, and how growth expectations may be set in different contexts for different purposes.

What is growth?

In the simplest terms, growth is change over time. To study growth, we measure a thing repeatedly on successive occasions and draw conclusions about how it has changed. People may speak of growth in the context of a system (e.g., a population) or in terms of an organism (i.e., an individual). In the former, we may be concerned with how many individuals comprise the population, how they are dispersed and how rapidly their number increases. In the latter instance, we are generally concerned with how attributes of the organism (e.g., height, weight, reading ability) change over time. Although both notions of growth are interesting, in this paper we are mainly concerned with the second idea because it most closely relates to the concern we have for how individual students develop physically and cognitively.

Most people are familiar with physical growth and some of the ways in which it is measured. For example, one of the things doctors do with new babies is to weigh them and measure their length. Height and weight measurements are continued as the child matures. On any given occasion, specific measures of height (length when very young) and weight are obtained in terms of inches and pounds. Each year (or more often when very young) the measurements can be repeated and a history of development can be gathered for the individual. The change in these measurements over time tells us about the growth in height and weight of the individual, which in turn gives us clues about the child's general health and well-being. Similarly, when children become students in our public schools, their academic performance is measured, for example, in reading. On any given occasion a specific measure of their reading ability is obtained in some metric. Each year (or perhaps more often in some situations) the measurements can be repeated and a history of the student's reading achievement is possible. The change in these measurements over time tells us about the growth of the student's reading ability, which in turn gives us clues about the cognitive health and well-being of the child.

In the preceding paragraph it sounds as though the measurement of physical growth and the measurement of cognitive growth are very similar. In some respects they are, but there is actually a huge difference in practice. You may have noticed that in the preceding comments about height and weight, the measurements were in terms of inches or pounds. In contrast (and this is significant!) the measurement of reading ability was in "some metric." The difference is that whenever we measure height and weight we always use inches and pounds. (In Europe it would perhaps be centimeters and kilograms, but this is not a fundamental difference because there is a direct universal relationship between inches and centimeters and between pounds and kilograms.) In sharp contrast, for the majority of the last century there was no universally accepted metric for the measurement of reading achievement. For the most part, each reading test had its own proprietary metric and, unlike Fahrenheit and Celsius, the reading metrics were not "exchangeable," "convertible," or "translatable" from one to another.

Near the end of the twentieth century, MetaMetrics®, Inc. developed a common metric for reading called The Lexile Framework® for Reading, which is now the most widely used reading scale. However, other metrics still abound. This has huge implications for our understanding of academic growth, as we discuss next.

How is growth measured?

A central question to be addressed when discussing growth is "growth in what?" What are we measuring on each occasion? What is changing over time? Underlying these questions is the assumption that it is the same thing on each occasion even though its magnitude might differ over occasions. (Indeed, we expect its magnitude to change. That is why we measure it more than once.)

For example, when we measure height or weight we fully expect measurements to increase from birth to adulthood. It is this change that interests us. But even though their magnitudes increase over time, it is always height and weight that we measure on each occasion. We do not measure height and weight on one occasion and arm length and girth on the next occasion. This seems trivially obvious when we measure physical attributes, but it is not so obvious when measuring cognitive attributes, like reading ability.

Measuring reading ability is more like measuring temperature. Although we can see a person's height or weight, we cannot directly observe the temperature of an object. We can see evidence of temperature by observing the height of a column of mercury in a thermometer. Similarly, we cannot see a person's reading ability. However, we can see evidence of a person's reading ability by asking them to respond to questions about textual matter they have read. For both temperature and reading ability, we construct an instrument that gives evidence of the unseen attribute. Unseen cognitive attributes are called constructs because we infer their existence from the behavior or performance of individuals. When performance changes, we understand this reflects a change in the underlying construct. Hence, we assume that changes in these unseen constructs are the primary causes of variation in the measurements we observe.

There is a challenge to measuring constructs that is not present when measuring physical attributes such as height and weight. How can we know that the construct that we measure on the second occasion is the same one that we measured on the first occasion? For example, if we ask the same questions on subsequent occasions that we asked the first time we measured the person's reading ability, they could have remembered the answers to some of the questions. The next time we ask the same questions, the student might be able to answer them without even reading the text. In that case we would certainly not be measuring the student's reading ability!

In the example above, the construct changed. On the first occasion we may have actually measured reading ability. But the next time we may have obtained a measure of reading ability contaminated by memory of prior questions and answers. That being the case, we cannot examine the change in the two measures and conclude that the reading ability has changed. We did not measure only reading ability on both occasions!

This points out a key requirement for measuring growth. If we are to measure growth in cognitive constructs there must be a fundamental constancy or invariance in the construct over time. Its magnitude may change but its nature must remain the same. We have to measure the same thing on each occasion in order to even talk about growth.

Psychometrics is a branch of psychology dealing with the design, administration and interpretation of quantitative tests for the measurement of psychological constructs such as intelligence, aptitude (e.g., reading ability) and various personality traits. Making sure that tests really measure what they are intended to measure is one of the fundamental jobs of psychometricians. When they do this, they are ensuring the construct validity of the test. But psychologists who develop measures of cognitive growth must go even further. They must assure that tests are constructed and administered in ways that result in the same construct being measured each time the test is administered. There must be invariance of construct in studies of growth.

There is another fundamental requirement for measuring growth. We must use an appropriate equal-interval scale consistently over time.

A scale is called equal-interval whenever a unit distance at one place on the scale indicates the same amount of change in the underlying construct as a unit distance at another place on the scale. For example, a two-inch increase in height means the same thing regardless of whether the increase was from 32 to 34 inches or from 70 to 72 inches. As long as we record the numbers in terms of inches each time, we have used the same scale (inches) consistently over time.

These fundamental requirements must also apply when we measure psychological constructs. Psychometricians must develop scales that behave in an equal-interval fashion. Furthermore, when we study growth we must use the same equal-interval scale consistently over time. One famous psychometrician coined a now well-known phrase to capture this notion: "If you want to measure change, don't change the measure."

When we design studies of growth, it is important to use a valid equal-interval scale. Furthermore we must be able to persuasively demonstrate that over time there is invariance of construct and consistency of scale. These are the fundamental underpinnings for measuring growth. If these conditions are not met, the study may be interesting but it is not about growth.

The great advantage gained by employing stable constructs and consistent equal-interval scales is that we can perform mathematical operations (addition, subtraction, etc.) in sensible ways with the scale values that are recorded on each occasion of measurement. We can add the amounts of growth in consecutive time periods to establish the amount of growth over the whole time-span, for example. More importantly, we can mathematically model the growth over time and look at its functional form mathematically. For example, does the individual grow in a steady fashion with a constant rate of growth? Or does the individual grow faster when young and more slowly as he or she grows older? Do different individuals exhibit different patterns of growth? What is the most typical pattern of growth? How much variation should we expect to see across individuals? Once construct invariance and scale consistency have been demonstrated, it becomes possible to address questions like these.

What is "normal" growth?

When we ask, "What is normal?" whether it pertains to performance, height, reading ability, or growth in these attributes, we generally assume that we can make a judgment about what occurs most frequently in the general population of individuals. Usually this is accomplished by gathering information about the general population so that we have a frame of reference (data) against which to make comparisons. Such reference data are called norms.

In theory there are two types of norms for growth—cross-sectional norms and panel norms. In cross-sectional norms for growth, a sample of people representing the ages of interest are studied at a single point in time; or perhaps comparable samples of people are studied on multiple occasions, but not the same individuals each time. For panel norms, the same individuals are followed and studied on multiple occasions (as many as necessary to reflect the ages of interest.) In practice, cross-sectional norms are more common because panel norms are expensive and time-consuming to construct. Cross-sectional norms are useful for seeing how an individual compares to the general population at any given point in time. Panel norms are preferable if we want to examine the rate of growth of the individual in relation to that of the population.

Probably the most familiar cross-sectional norms for growth are the Centers for Disease Control and Prevention (CDC) Growth Charts: United States published by the National Center for Health Statistics (NCHS), one of several centers under the umbrella of the Centers for Disease Control and Prevention (CDC) of the U.S. Department of Health and Human Services. The CDC Growth Charts are used by doctors everywhere in the United States as the frame of reference for evaluating the physical development of children.

The CDC Growth Charts are based on surveys of representative samples of people of different ages at specific points in time (but not the same people each time). The NCHS examined the distribution of height (also weight and selected other physical characteristics) across all individuals of a given age in their samples. In essence, they plotted selected percentiles of the distributions for every age from 2 to 20 years and created cross-sectional "growth curves" by connecting the corresponding percentiles from the distribution at each successive age. (It was considerably more complicated than that in reality. Sophisticated curve fitting and smoothing techniques were used to assure that the curves best described the data.)

For education, test companies construct cross-sectional norms by grade, rather than age, to be more applicable to the way public schools are organized. Test companies periodically test a nationally representative sample of students in each grade and construct norms tables to show how the academic performance of students is distributed in each grade. However, these norms are usually limited to a specific point in time and to a specific edition of a test. As a result they are not really growth norms, but achievement norms. Most test companies provide such norms for reading and mathematics, and often for other subjects as well.

The CDC Growth Charts show how the sizes (heights and weights) of individuals in the population vary at different ages. However, this is different from showing how the size of any specific individual changes as he or she ages over time. To do that you must follow the same individual over time and make measurements on the same individual at each successive occasion. If you do that with a representative population of individuals, then you have the basis for panel norms.

The CDC makes this distinction in their report 2000 CDC Growth Charts for the United States: Methods and Development. In the report, they use terms like "growth (or size)" and "size-attained" and "growth progress" to characterize the information about physical growth obtained from their cross-sectional norms. In contrast, they use the term "growth velocity" to distinguish the kind of information about growth that could be provided by panel norms. They say:

> There is a difference between growth (or size) charts and growth velocity charts. The 2000 CDC Growth Charts for the United States are based primarily on cross-sectional national survey data that were statistically smoothed to create percentile curves. … Therefore, these charts more appropriately may be considered size charts. When serial values for an individual are plotted, assessments can be made of that individual's growth progress over time.
>
> Growth velocity charts are constructed from incremental data obtained from longitudinal observations. (p. 14)

In education, the publishers of achievement tests do not publish growth curves comparable to those displayed in the CDC Growth Charts. However, cross-sectional norms tables are published independently for each grade and academic subject. In the past, some educators have used such tables for successive grades to see if students maintain the same percentile rank in the norms. Maintaining the same percentile rank was considered "normal growth." (Earlier evaluations of federally funded Title I programs were based on such a model.)

This approach provided for academic achievement an analog to the "growth size" information available from the CDC Growth Charts for height or weight. But as discussed above, this information is cross-sectional in nature. Such an approach does not produce "growth velocity" information of the kind that could be obtained from panel norms.

Panel norms for growth would have significant advantages over cross-sectional norms for growth. Foremost, they focus on the same individual over time, so they provide a more true-to-life perspective on intra-individual change. Second, they can better detect growth. NCHS points out that "growth velocity charts are more sensitive indicators of small changes in growth status than the size-attained charts …." Third, they are more accurate. Williamson (1988) demonstrated that a cross-sectional approach to constructing norms for a growth distribution could produce distortion, particularly at the extremes of the growth distributions.

Given these advantages, it may be surprising that panel norms for growth generally have not been constructed or widely used in practice. One can surmise several reasons. Among the possibilities, consider the following:

- It would take a long time to gather the data necessary for panel norms. Imagine that we desire to have norms for growth during the public school years. To be able to have norms for grades K–12, it would take at least 13 years to follow a group of students from Kindergarten to graduation. (That ignores the fact that some students will not actually finish in 13 years due to having to repeat one or more grades.)
- Over such long periods of time, people can move away and become unavailable for follow-up measurements.
- The attrition of students in turn compromises the validity of the norms because highly mobile students would not be included.
- There would have to be adequate measurement over the time period (invariance of construct, consistent measurement with an equal-interval scale).
- Curricular changes across grades may make it difficult or irrelevant to focus on the same construct throughout the entire timeframe.
- The cost of collecting the data for panel norms for growth would be much larger than for cross-sectional norms at a single point in time.
- Before such a study could be completed, most test companies would have published new editions of their tests. Norms for the older editions would be of less interest and not highly marketable.

Still, with the wide-spread proliferation of accountability programs in the United States, data bases are being expanded and improved. With a little work, some states may be in a good position to retrospectively create panel norms with data that already exist. Others have the opportunity to incorporate the right design features in systems that they are creating now to operate over the next 10-15 years. There would be several motivations for exploring this possibility.

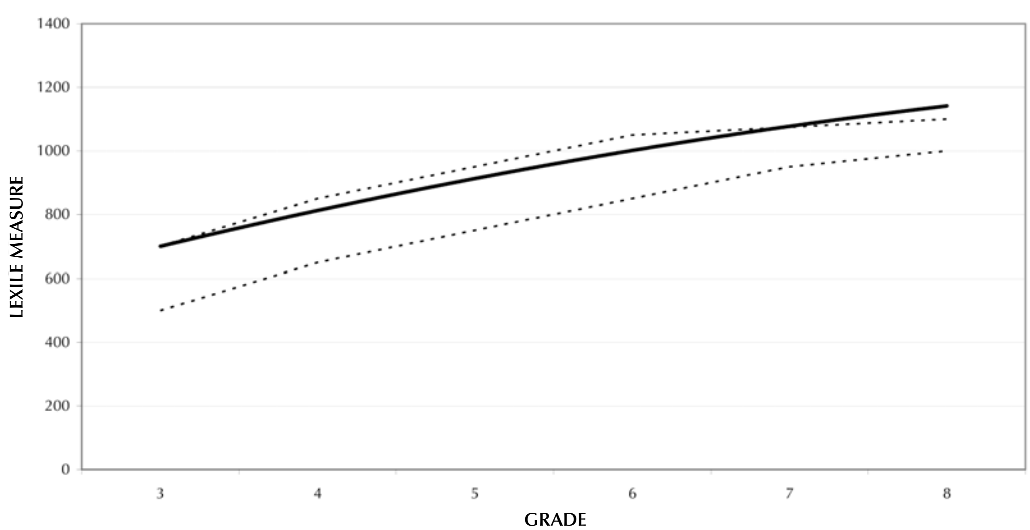
- A developmental approach to growth is possible with panel data. Analysts can mathematically model individual growth curves for students using such data.
- A focus on developmental growth curves allows one to study not only achievement, but also change in achievement over time (its rate and/or magnitude), and the functional form of growth (acceleration, deceleration, etc.).
- The relation between growth and initial achievement is more easily investigated. For example, do initially lower-performing students grow faster over time than those who start out with better performance?
- A wide variety of new analytical techniques have arisen in recent years to take better advantage of panel data.
- A longitudinal focus on student growth would be consistent with the current statutory emphases on student progress.
- New measurement scales (e.g., Lexiles®) provide a more nearly universal measurement framework for such long-term studies.

Cross-sectional norms and panel norms provide performance-based frameworks for determining what normal growth is. There is another useful perspective based on an analysis of curricular demands that are applicable to students in school. A good example is the text demand analysis enabled by The Lexile Framework for Reading (MetaMetrics, Inc., 2006). This approach looks at typical texts used in each grade, and places them on the Lexile scale. By depicting the range of Lexiles corresponding to texts typically used in each grade one can see what range of reading ability students must exhibit to be able to comprehend their texts. This provides a curricular-based rather than a performance-based perspective that might be used for defining "normal" growth.

An Example

The figure below provides an illustration of how more than one normative perspective may be combined. In the figure, the average growth curve for a longitudinal panel of readers is displayed in conjunction with the changing text demand for each successive grade. The growth curve represents the average growth of approximately 68,000 students who were third graders in the Public Schools of North Carolina in 1999, and who were followed until the end of eighth grade in 2004. Each year the students took the North Carolina End-of-Grade Test in Reading. Their scores were converted into Lexiles to facilitate the comparison with grade-level text demand. Longitudinal data analyses were used to estimate each individual student's growth curve. The average growth is displayed in the Figure by the solid curved line, and the range of reading difficulty for typical textbooks (approximately the middle 50%) at each grade is portrayed between the two dotted lines. The chart shows that on average these North Carolina students were reading near the upper difficulty levels of typical grade level texts.

1999-2004 North Carolina Average Growth Curve Relative to Grade Level Text Demand



Expected growth

In the previous sections we discussed normal growth and described three ways to characterize it (through cross-sectional norms, panel norms, or text demand). An example was given illustrating how to combine two normative perspectives (a panel norm for average growth and changing grade level text demand) into a single picture. In this section we raise a slightly different question—how much growth to expect of a given student or group of students.

There are at least two fundamental ways to address this question. We can base the decision on some reference standard regarding what growth has been in the past (norms) or we can base our decision on what we think is desirable for the future (aspirations).

If we base the decision on norms, we will look at past and present performance and ask:

- How much growth do we expect?
- How much growth is typical?

When we use past performance as our guide, then the answer to these questions is "whatever is normal" for the students given their age, grade, background, etc. For instance, we could use the average growth curve in the example from the previous section to calculate the year-to-year growth that occurred in North Carolina. This is one expression of what may be "normal" given the performance of the students in the panel.

Average Year-to-Year Growth Exhibited by the 1999-2004 Panel of North Carolina Students

Grades	Average Growth in Lexiles
3 to 4	113L
4 to 5	100L
5 to 6	88L
6 to 7	76L
7 to 8	64L

On the other hand, if we base our decision about what growth is expected on our aspirations for the future, we are asking questions such as:

- How much growth should the student(s) exhibit?
- How much growth is desirable?

To answer these questions, we have to look to sources other than (or in addition to) past performance. Subjective judgments must be made about how much growth is needed for some identified purpose, or consistent with some agreed upon values.

Even though norms describe what is or has been, they can still be a useful reference for what should be, or to what heights students should aspire. For example, by referring to achievement norms, educators can assess what the typical performance has been in the past and so set a baseline for future performance comparisons. Exceeding the achievement norms of the past has often been the goal of educators.

Another normative method of setting an expectation for growth is to use the curriculum as a reference and specify what students should know and be able to do at each grade. Such curricular standards are often further operationalized as proficiency standards for tests that have been developed to align to the curriculum. When a universal scale is available, such as the Lexile Framework, the texts used at each grade can be mapped on the scale to create a graph of the text demands over time. This is effectively a textual norm against which student performance and progress may be plotted, as illustrated in the example above.

Although "whatever is normal" is one answer to how much growth we should expect, it is not the only answer. It is generally not the answer in most accountability systems today, which are motivated by a desire to see students perform at higher levels than has been the case in the past. In this context, more than "normal" growth is generally desired. But how much?

One way to address this issue is to focus on a performance that may be required of students in the future. For example, using the Lexile Framework to study postsecondary text demands, Williamson

(2006) showed that a large gap exists between the difficulty level of texts used in high school (eleventh and twelfth grades) compared to texts used in the first two years of university work. Furthermore, there is a systematically increasing continuum of text difficulty that spans the most typical postsecondary activities of students—citizenship, the military, the workplace, community college and the university. Given the knowledge that students may encounter more difficult books in their postsecondary endeavors, we can ask how much growth in reading ability must occur during school to allow students to reach the desired reading ability by the end of twelfth grade. This would form the basis for a more demanding growth expectation in reading.

An example of another strategy for setting growth standards is the approach used by the Public Schools of North Carolina from 1996-97 until 2004-05. In North Carolina, as they developed the ABCs accountability program, they began by analyzing the year-to-year growth of the same students over time. Then the North Carolina State Board of Education (NCSBE) set a target of exceeding the statewide growth of previous years by 10% as one level of expectation for future growth. The decision to choose 10% was based on considerations of the potential payoff in terms of student proficiency after several years of implementation and the potential impact on schools under the accountability system of rewards and sanctions.

When setting expectations based on aspirations as North Carolina did in constructing the ABCs, a guiding philosophy is important. This is used to construct a framework for deciding how to balance the desire for challenge and rigor with a competing desire for realism and attainability. The NCSBE wanted to see improvements in proficiency and closing of achievement gaps within a short but realistic timeframe. They used input from educators across the state, analyses of data from previous years and the state's performance on other measures to provide the context for their decision. North Carolina used both norms and aspirations to set their expectations (growth goals) for the ABCs.

Sometimes aspirations are self-imposed, such as when a student decides to compete in the county spelling bee, or a state aspires to be "first in education" by some date. Sometimes aspirations are externally imposed, such as when a state establishes accountability targets for schools, or when the federal government passes laws such as NCLB to ensure that states incorporate new standards into their accountability programs.

However, rarely are expectations based purely on norms or purely on aspirations. Usually it is some combination of the two, or an intellectual interplay that involves the considerations of what is possible and what is reasonable.

Consequences of measurement error

All measurements are subject to errors that result from influences that have nothing to do with the thing being measured. For example, if we're measuring height, but the individual slouches, we'll get an inaccurate measurement of their erect stature. If we give an individual a test of reading ability, but the room where the test is administered is filled with noisy distractions, then it is unlikely that we will get an accurate indication of the person's reading ability. A more complete introduction to measurement error is presented in Williamson (2004).

Just as single measurements are subject to measurement error, repeated measurements that are used to estimate growth are each subject to measurement error. Consequently, estimates of growth produced from measurements on multiple occasions are also affected by measurement error. Knowing how dependably we have measured growth is thus complicated. There are a number of factors to consider.

One factor that affects the precision with which we can estimate growth is the precision with which we measure status on any given occasion. Two additional factors that affect the precision of growth estimates are how many replications of measurement are available and how they are distributed across time.

In general, having more data is better. It is easy to understand that we cannot measure growth if we only measure the individual on one occasion. Many studies of change have been based on two successive measures, and there has been much debate about the adequacy of this approach. Rogosa and Willett (1985,

1983) describe the limitations on what can be learned when only two measures are available. They have helped to clarify why more than two measures are needed when growth is the focus.

In addition, when the measurements are made affects the accuracy with which we can estimate growth. In essence, the more spread out the measures are in time, the more precisely we can estimate the amount (or rate) of growth. In addition, when the measurements are more spread out, we are better able to see the functional form (changes in or shape of) growth over time.

Imagine a student's growth in reading ability over six years, say from the end of second grade to the end of eighth grade. Suppose that growth is fast during the early years, but by grade five it starts to slow down and tapers off by grade eight. If we took only a few measures, what we would see would depend greatly on when we made the measurements. If they were all in the early grades, we would estimate the growth to be very fast; whereas if we measured only in the later years, we would conclude that growth was relatively slow. In order to get the whole story, we have to space the measurements out over the whole time period, with some measurements near the beginning, some in the middle and some near the end. Even when growth is uniform over the period, we can still estimate the growth more precisely when the measurements are spread out than when they are all near the middle of the timeframe.

Another thing that affects the accuracy with which we can estimate growth is the choice of growth model. Consider again the previous example. In order to correctly capture the growth trend, we must adopt a curved trajectory to describe the growth. If we chose to model the growth with a straight line, we would correctly capture the average change over the whole time frame, but we would miss the early acceleration and the slowing down in later years.

Our ability to assess the collective growth among a group of individuals is similarly affected by considerations related to precision of measurement, amount of data, and the data collection design. In addition, if we are interested in inter-individual (between persons) variation in growth, there must be real variability in growth for us to detect it. If everyone is growing in the same manner then there is no inter-individual variation to detect, and measures of change may appear to be unreliable. Similarly in such situations there may appear to be no relation between change and initial status for that group of individuals. However, these same conclusions are sometimes reached because of the way data were collected or analyzed. Rogosa & Willett (1983, 1985) provided an extensive analysis of the implications of various methodological choices that researchers have made in the measurement of change.

Although the notion of measuring growth seems simple, the details of obtaining dependable measurements and valid inferences about growth can be quite complicated. A vast literature about growth in cognitive constructs appeared during the twentieth century with an explosion of methodological advances during its final decades. Taking full advantage of these advances requires carefully constructed databases of longitudinal data, which in turn require patience and care to maintain. Among the payoffs are the capabilities to better understand growth and to set reasonable expectations for growth. These capabilities seem particularly relevant now in the context of the educational accountability requirements of the 21st Century.

References

National Center for Health Statistics. (2002). 2000 CDC growth charts for the United States: Methods and development. Vital and Health Statistics, 11(246). Washington, D.C.: National Center for Health Statistics.

MetaMetrics, Inc. (2006) Figure 1: Typical reader and text measures by grade [On-line]. Available: http://www.lexile.com/DesktopDefault.aspx?view=ed&tabindex=6&tabid=18#18

Public Law 107-110. The No Child Left Behind Act of 2001.

Rogosa, D. R. & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. Psychometrika, 50(2), 203-228.

Rogosa, D. R. & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change. Journal of Educational Measurement, 20(4), 335-343.

Williamson, G. L. (2006, April). Student readiness for postsecondary endeavors. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Williamson, G. L. (2004). Why do scores change? Durham, NC: MetaMetrics Inc.

Williamson, G. L., Appelbaum, M. I., & Epanchin, A. (1988). Analyses of achievement data from Durham County (NC) schools. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED294923.)

About the Author

Gary L. Williamson, Ph.D., is a former senior research associate with MetaMetrics, Inc. With more than 30 years of experience in educational research on the academic, state and school district levels, Williamson's specialty is quantitative methodology encompassing psychometric, mathematical and statistical applications to educational data. He has written and spoken extensively on the subjects of educational assessment and accountability. Williamson earned both a doctorate of philosophy in mathematical methods for educational research and a master's of science in statistics from Stanford University. He also holds a master's of education in educational research and evaluation from The University of North Carolina at Greensboro, and a bachelor's of science in mathematics from The University of North Carolina at Chapel Hill.

MetaMetrics®
LINKING ASSESSMENT WITH INSTRUCTION

MetaMetrics, Inc.
1000 Park Forty Plaza Drive, Suite 120
Durham, North Carolina 27713

Phone: 919–547–3400/1–888–LEXILES
Fax: 919–547–3401
Web site: www.Lexile.com