Achieve3000®

MULTIPLE MEASURES

WHITEPAPER

How Achieve3000 Literacy Uses Multiple Measures

The Importance of Multiple Measures

American students are frequently assessed to determine their achievement status and their progress toward institutional and personal educational goals. For example, all states administer end-of-year assessments of reading and mathematics achievement in compliance with the Elementary and Secondary Education Act (PL-107-110), which in its current authorization is known as the No Child Left Behind Act of 2001 (No Child Left Behind [NCLB], 2003). However, these annual summative assessments are intended primarily for institutional accountability and occur too infrequently to provide useful feedback to educators. Consequently, states and school districts increasingly use interim assessments to gauge student progress and monitor short-term educational program effects. In addition, educators, parents, and students have discovered the value of differentiated instruction to obtain individually tailored educational experiences for students in K-12, as well as for adult learners. As the leader in online differentiated instruction, Achieve3000 Literacy® provides differentiated instruction that is integrated with individual measurement of reading ability so users of its instructional systems can continuously track individual progress. This increased use of assessment in education results in students having multiple measures—some of which come from repeated measurements over time (with the same test) and some of which result from using a variety of tests during a school year.

Why Multiple Measures Are Sometimes Seen as Problematic

Whenever a particular trait of an individual (e.g., reading ability) has been assessed multiple times, the individual has several scores to characterize his or her reading achievement. This is what we mean by multiple measures—multiple scores derived from being assessed multiple times, perhaps with different tests being used each time. Some people would assume that more information is better; therefore, multiple measures of reading ability would be a great thing for students, parents, and teachers. However, it is usually the case that scores from multiple measures differ for the same individual. When scores from multiple measures are reported on a common scale (e.g., The Lexile® Framework for Reading), score discrepancies seem even more obvious. When a student is learning and the scores go up, no one worries. However, when scores go down, or when scores from multiple sources differ even though the tests occurred at approximately the same time, questions usually arise. Larger discrepancies typically generate more concern. In these cases, multiple measures may be perceived as a problem.

How this Paper Can Help

This paper is intended primarily for educators and administrators who use Achieve3000 Literacy to accelerate literacy growth for their students. It may also be helpful to other adults, e.g., parents, caregivers, and adult learners, who share in the responsibility of helping students understand and track their reading achievement and progress.

The paper begins by introducing Achieve3000 Literacy, LevelSet™ and the Lexile Framework for Reading to create an appreciation for how multiple reading measures are generated and managed in relation to differentiated instruction and other interim or summative assessments that may be available. Second, the paper provides a brief synopsis of various strategies that can be used to make sense of multiple measures. Finally, the paper points to other resources that may be helpful. With this introduction, readers should be able to better understand the resources at their disposal and can use them to become more confident with regard to accurately reporting reading achievement and growth.

Context

Achieve3000 Literacy

Achieve3000 Literacy operates on the fundamental premises that literacy unlocks achievement and differentiated instruction is the key to improving literacy. Achieve3000 Literacy is the only web-based, differentiated instructional solution designed to reach a school's entire student population as well as adult learners. For students using Achieve3000, reading ability is enhanced through a series of literacy solutions, each designed for a particular portion of the developmental continuum. These solutions have been designed to closely align with key objectives of the Common Core State Standards (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) to give students the content-area literacy skills they need to succeed in school and prepare for college and career. Details about each of the products are available at www.achieve3000.com.

Achieve3000 Literacy is powered by the LevelSet assessment tool and proprietary software engine, which distributes lessons to an entire class, yet tailors them according to each individual student's reading level. The two main purposes of the Achieve3000 Literacy reading assessments are to initially measure student reading comprehension so reading materials can be appropriately targeted (i.e., matched with the student's reading ability) and to iteratively measure growth in reading comprehension throughout the school year. In order to meet these goals, a developmental scale must be used to report the results. The Achieve3000 Literacy assessments are reported on the Lexile scale, a scientifically-based scale of reading ability. The Lexile scale is applied to both readers and texts, making it possible to match readers with texts of appropriate difficulty to facilitate reading improvement. Importantly, the Lexile scale provides accurate feedback on a students' developing reading ability, helping measure progress and forecast future performance.

Completion of activities upon reading the differentiated nonfiction articles produces repeated measures of students' reading abilities as they learn. For students showing sufficient reading abilities, Achieve3000 Literacy uses a Bayesian scoring algorithm to provide continually updated measures that monitor progress in reading development. The Bayesian approach uses prior scores to refine each new estimate of achievement to improve the accuracy of measurement as students learn. In this way, Achieve3000 Literacy uses multiple measures over time to improve the assessment of reading ability, which in turn improves the ability to match students with appropriate texts.

The Lexile Framework for Reading

The Lexile Framework for Reading is a unique resource for accurately matching readers with text. Unlike other measurement systems, the Lexile Framework evaluates reading ability based on actual assessments, rather than generalized age or grade levels. The true power of the Lexile Framework is its ability to measure both a person's reading ability and the complexity of a text (i.e., a book or magazine article) on a single developmental scale. The Lexile measure is shown as a numeral with an "L" after it—880L means 880 Lexile.

A student gets his or her Lexile reader measure from a reading test or program. For example, if a student receives an 880L on her end-of-grade reading test, she is an 880 Lexile reader. Higher Lexile measures represent higher levels of reading ability. A Lexile reading measure can range from below 200L for emergent readers to above 1600L for advanced readers. Readers who score below 0L receive a BR for Beginning Reader. In some cases, a BR code for readers is followed by a numeral and L (e.g., BR150L). A Lexile reading measure of BR150L indicates that the Lexile measure of the reader is 150 units below 0L. The smaller the number following the BR code, the more advanced the reader is. For example, a BR150L reader is more advanced than a BR200L reader.

The Lexile Framework is based on more than 20 years of research funded by the National Institute of Child Health and Human Development. Its distinct approach to measuring readers and texts has resulted in adoptions by departments of education in nearly half the states and by school districts in all 50 states. Tens of millions of students worldwide now receive a Lexile measure that helps them to find targeted text from a universe of more than 100 million articles, books, and websites. Major book retailers, test companies, and text publishers provide Lexile measures. The Common Core State Standards (CCSS) for English/Language Arts cite Lexile measures as key indicators of text complexity. The CCSS provide recommended Lexile bands for reading comprehension development by grade levels to ensure students are on track for college and career text demands.

Today, the Lexile Framework for Reading is recognized as the most widely used reading metric. It is a powerful tool for connecting learners of all ages with materials at the right level of challenge and monitoring their progress toward goals and standards. You can find out more about the Lexile Framework at www.Lexile.com.

Strategies for Dealing with Multiple Measures

People often ask how a Lexile measure can be different for the same student as measured by different assessments. For example, a student might take a computer-based adaptive assessment such as the Measures of Academic Progress (MAP®) or the Scholastic Reading Inventory™ (SRI) and also a state assessment such as the Smarter Balanced Assessment (SBA) in reading. Each of those assessments can produce a Lexile measure; and, when a student takes two or more such assessments, the student's scores will vary from one assessment to another. Yet when a student's scores vary from one assessment to another, people often ask, "Which one is correct?" It's one of the top questions being asked today.

As strange as it may seem, the answer to the question is "none of the scores is correct," because every score a student receives is subject to some degree of measurement error. It doesn't matter which tests were used or how many tests were used. All educational measurements are subject to measure error. However, there are ways to deal with this fact, as explained in the next few sections, which provide an overview of various ways to make sense of multiple measures. Although the answer is not as simple as people may wish, understanding the situation is an important first step toward understanding how Achieve3000 Literacy's LevelSet assessment and the Lexile Framework for Reading help control measurement error for improved estimation of student reading achievement.

Averaging

Let us suppose that a student has taken two assessments, only a couple of weeks apart. One score might come from a benchmark test and perhaps the other comes from a LevelSet assessment. Let's further suppose that the two measures are 1000L and 1100L. There are many factors that can influence the interpretation of the scores. One factor stands out because of its ubiquity. All measurements are subject to measurement error—that is, random fluctuations due to idiosyncratic factors that can influence a student's performance on any single occasion.

Tests are designed to control measurement error through standardization of their construction and administration; however, tests that are designed for different purposes may measure with different precisions. The consistency with which a test measures an attribute is called its reliability, and its precision is called its standard error of measurement. In a sense, the standard error of measurement is an indication of how well a particular test controls measurement error when the test is administered under specific, well-controlled conditions. However, test administrations cannot control every factor that might possibly influence a student's performance on any particular occasion. When measurement error and test precision are involved, a relatively simple solution for managing multiple measures is to average the available scores. However, there are two situations to consider:

Simple averages. When two assessments measure with similar precision, the main influence on score differences is usually random measurement error. When that is the case, a simple arithmetic average of the scores produces a single estimate of the student's achievement that can be used for educational purposes. So in our example, we may compute an average from the two assessments—namely, (1000L + 1100L)/2 = 1050L as the single best estimate of the student's reading ability.

Weighted averages. However, when tests measure with different precisions, a better estimate takes the standard errors of measurement of the tests into account. In this scenario, a weighted average is computed, where the weights are based on the standard errors of measurement for the two tests. Fortunately, one does not have to rely on a hand computation to obtain the result. There is an online application that will provide a solution, which takes account of the appropriate factors. The application is available at: https://lexile.com/educators/understanding-lexile-measures/managing-multiple-measures-resource-center/

Bayesian Estimation and Forecasting

Sometimes assessments are repeated over time to monitor a student's progress. In some cases the time frame may be a few weeks or months. In such cases, students may exhibit some growth. Under such conditions, averaging the successive scores is not sufficient; more sophisticated techniques are required to estimate a student's performance and progress. Such is the case with students who use Achieve3000 Literacy. As mentioned earlier, Achieve3000 Literacy uses Bayesian statistical estimation to handle this added complexity.

Bayesian methodology provides a paradigm for combining prior information with current data (both of which are subject to uncertainty) to arrive at an estimate of current status (which is also subject to uncertainty). Uncertainty is modeled by mathematical probability. With Achieve3000 Literacy, when a student is administered the first LevelSet placement test, the results constitute the prior information for the following test administration—i.e., reading activities or another LevelSet assessment. Each subsequent assessment uses the accumulated prior information from all previous assessments. At each step, the current data come from the performance on the most recent assessment (i.e., activity score or LevelSet test).

Both prior information and current data are represented via probability models reflecting uncertainty. From the Bayesian point of view, uncertainty arises because test performance is associated with, but not determined by, the ability of the student; it is that ability, rather than the test performance per se, that we endeavor to measure. Any single performance may over- or underestimate a student's ability because of measurement error. Thus, we are always uncertain about the ability that produced the performance; the Bayesian approach handles that uncertainty by incorporating it into the statistical model for estimating student performance. In the Bayesian paradigm, if a substantial amount of time has passed since the last assessment, an allowance is also made for an uncertain amount of growth having occurred since the previous assessment. This allowance is accomplished by means of a growth model, which estimates as a function of elapsed time both student growth and the augmentation in uncertainty.

Achieve3000 LevelSet reading assessments are designed to measure a reader's ability to comprehend expository texts of increasing complexity. The results of the Achieve3000 LevelSet reading assessments and multiple-choice activities can be used to measure where students stand in the development of their reading ability.

One outcome of the Achieve3000 Literacy LevelSet reading assessments and multiple-choice activities is the location of the reader on the Lexile scale. Once a reader is measured, it is possible to forecast how well the reader will comprehend thousands of books and articles that have been measured with the Lexile Framework. Readers and texts are measured with the same Lexile scale making it possible to directly compare a reader and a text. When reader and text measures match, the Lexile Framework forecasts 75% comprehension for the student reading that text. That means if the reader were asked 100 questions based on the text, he or she should be able to correctly answer 75. When the text has a Lexile measure 250L higher than the reader measure, the Framework forecasts 50% comprehension. When the reader measure exceeds the text measure by 250L, the forecasted comprehension is 90%.

Growth

Educators and parents are interested in whether students' reading achievement improves as a result of schooling. Consequently, the Elementary and Secondary Education Act requires states to measure students' reading achievement annually from grades 3 to 8. Over a number of years, we expect students to grow substantially in their reading achievement. The growth will likely be great enough that the changes in scores over time outstrip the random fluctuations in scores due to measurement error. Furthermore, such changes in scores should be systematically increasing on average. Nevertheless, because of measurement error, sometimes it can be hard to discern growth from measurement error in individual cases. Consequently, statistical models have been devised to detect student growth in a longitudinal series of scores collected over time.

Multi-level models. When growth occurs over longer periods of time (e.g., years), a student may have four or more reading measures dispersed across time. When such measures are derived from a single test edition, a statistical model is typically used to estimate the student's changing performance as well as features of growth, such as the growth rate. A family of statistical models is available for such purposes. They go by various names—multi-level models, hierarchical linear models, random coefficient models, mixed linear models, etc. The common feature of these models is that each individual has a specific growth curve, and the individual growth curves may vary across individuals. The statistical models are constructed to detect individual growth as well as the variation in individual growth for a group of students (e.g., a class, school, district or state). Multi-level models comprise a class of advanced statistical techniques that require special expertise to use.

Latent growth models. An even more complicated situation arises when students have been measured many times with each of several assessment devices. For example, we can imagine that over the six years students spend in grades 3-8 they could accumulate a record of reading measures that consisted of six end-of-year reading assessments administered for the purposes of NCLB, perhaps 18 scores (three per year) from a progress-monitoring test used for program monitoring and perhaps as many as 54 (nine per year) monthly measures based on two LevelSet administrations per year and Multiple-Choice Activities, plus intervening monthly updates of reading ability. In this kind of data collection, multiple measures come both from repeating the measurement over time and from using multiple assessment devices to measure reading ability. To handle this increased complexity there is an advanced type of statistical analysis called latent growth analysis which can be used to understand the data. Such techniques are sophisticated and require special training to implement.

Managing Multiple Measures is Easy with Achieve3000 Literacy

Achieve3000 Literacy generates multiple measures for students via its LevelSet assessment and monthly monitoring. In most cases, LevelSet is administered twice a year. It starts with an initial assessment to establish a baseline Lexile score at the beginning of the school year or when the Achieve3000 Literacy program is first implemented at a site. A post-assessment is given at the end of the academic year. Between the pre- and post-assessments, for students with independent reading abilities, the system monitors performance on the multiple-choice activities and, if warranted, updates individual Lexile levels on a monthly basis. Students identified as Developing Readers receive sustained practice at the text level associated with their LevelSet score, and they complete a mid-year, interim LevelSet assessment. In this context, Achieve3000 Literacy uses its Bayesian estimation and forecasting procedures to manage the multiple measures that are generated by its assessment and monitoring system.

To make it easier to handle multiple measures that emanate from sources other than Achieve3000 Literacy, educators may use the previously described online application for computing simple and weighted averages of scores. In addition, a variety of other resources are described in the next section and are available for understanding and using multiple measures.

Resources for Educators

MetaMetrics Whitepapers

MetaMetrics provides a trilogy of whitepapers to help educators understand issues related to measurement error, growth, and multiple measures. Students in the United States may be more widely measured now than at any previous time in history. With more frequent measurement, parents and teachers have access to more information about their students' performance than at any previous time. With the increased availability of information, parents and teachers are better informed than in the past. Ironically, they may also find themselves having more questions about the results than at any time in the past. *Why Do Scores Change?* (Williamson, 2004) explores the nature of measurement error and how it can affect our interpretation of student achievement.

There are many viewpoints about how best to conceptualize and measure student growth. This makes it especially important for students, parents, and educators to better understand student growth, how it can be assessed, and how growth expectations may be set in different contexts for different purposes. *What is Expected Growth?* (Williamson, 2006b) explores issues related to student growth.

With increasing variety and frequency of assessments, combined with more frequent use of a common scale for assessments of the same construct, it is increasingly routine to have multiple, comparable assessment measures for reading and/or mathematics available for each student. *Managing Multiple Measures* (Williamson 2006) provides a more detailed introduction to the issues important to an understanding of how to use, interpret, and communicate information about student achievement derived from multiple measures. It is recommended as supplemental reading that is most directly related to the topic of this paper.

Achieve3000 Resources

Achieve3000 makes extensive resources available on its website. They include videos, whitepapers, webinars, literature on educational solutions, and networking opportunities for educators. These resources can be accessed at https://www.achieve3000.com/impact/research/.

Conclusion

Multiple measures of student achievement are a fact of life in the 21st century. Multiple measures are used to help understand student growth and to provide auxiliary information for many educational decisions. However, the interpretation of multiple measures can be complicated. There are a variety of strategies available, ranging from simple arithmetic averages to advanced statistical models. This whitepaper has provided a very brief introduction to the world of multiple measures and points the reader to various available resources for further exploration.

References

National Governors Association (NGA) Center for Best Practices, Council of Chief State School Officers (CCSSO). (2010). *The Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Washington, DC: Author.

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 et seq. (West 2003).

Williamson, G. L. (2004, April). *Why do scores change? A white paper from the Lexile Framework® for Reading*. Durham, NC: MetaMetrics, Inc. Retrieved from: https://metametricsinc.com/research-publications/why-do-scores-change/

Williamson, G. L. (2006a). *Managing multiple measures. A white paper from the Lexile Framework® for Reading.* Durham, NC: MetaMetrics. Retrieved from: https://metametricsinc.com/research-publications/what-is-expected-growth/.

Williamson, G. L. (2006b). *What is expected growth? A white paper from the Lexile Framework®* for Reading. Durham, NC: MetaMetrics. Retrieved from: https://metametricsinc.com/research-publications/managing-multiple-measures/.

To learn more about **Achieve3000 Literacy**®,
please contact **1-800-838-8771** or visit **achieve3000.com**

About Achieve3000

Achieve3000 delivers a comprehensive suite of digital solutions that significantly accelerate and deepen learning in literacy, math, science, social studies, and ELA. Using personalized and differentiated solutions, Achieve3000 enables educators to help all students achieve accelerated growth. For more than five million students in grades PreK-12, Achieve3000 improves high-stakes test performance and drives college and career readiness.

Achieve3000®