Every Student College & Career Ready: A Compendium of MetaMetrics' Collected Works

A collection of policy briefs, position papers, white papers and research briefs written by MetaMetrics researchers and psychometricians.

MALBERT SMITH III
JASON TURNER
STEVE LATTANZIO
TODD SANDVIK
MATT COPELAND
DAVID LIBEN
ANNE SCHIANO
ELIZABETH LATTANZIO

TABLE OF CONTENTS

INTRODUCTION

For the past few years, there has been great optimism among educators that the fog of uncertainty around ESEA (NCLB) would be lifted: the status, adoption, and implementation of the Common Core State Standards (CCSS) would be complete; the transition from state assessments to one of two national testing consortia (PARCC and Smarter Balanced) would occur; there would be marked improvement in the achievement gap; and more students would graduate college and career ready. But if anything has become crystal clear from the lessons of the past few years, it is that uncertainty and change are here to stay.

While there is uncertainty around the reauthorization of ESEA, the implementation of the CCSS, and the future of the testing consortia, there is no uncertainty associated with the goals of college and career readiness and eliminating the achievement gap. On both of these goals we have written extensively on how U.S. educators are faring versus public perception, on policies and practices that will help us achieve these two goals, and the importance of taking a longitudinal approach to these goals.

In this compendium we have attempted to provide our best work, reflecting our best thinking on educational policy in a volume that is meant to address the policy work now being done in reading and literacy, especially in relation to college and career readiness. We have split this collection into four sections:

Section 1: Educational Policy
Section 2: Looking to the Future
Section 3: The Lexile Framework for Reading and Text Complexity
Section 4: Assessment and Educational Reform

Within these sections, we cover a wide range of topics, from instructional strategies to close the readiness gap to global trends in education to the hallmarks of next-generation assessments. We have also written extensively on the impact that the CCSS is having in the field of education.

The intended audience for this volume is primarily educators—state administrators, district and school leaders, teachers, literacy coaches, and related practitioners—but also policy makers and laypersons concerned with the state of education in America. It is our hope that, looking through the lens of the Lexile Framework for Reading, these essays serve as a guide for educators and policy makers who are working tirelessly to make a difference, to help eliminate the achievement gap, and to promote college and career readiness for all students.

For more research and technical information on the development of the Lexile Framework, as well as an extensive collection of papers, visit www.Lexile.com.

www.MetaMetricsinc.com

www.Lexile.com

www.Quantiles.com

SECTION 1: Educational Policy

MetaMetrics® Policy Brief | 1420L
March 3, 2011

Bridging the Readiness Gap:

Demystifying Required Reading Levels for Postsecondary Pursuits

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D.

Graduation from high school no longer guarantees that students are prepared for the postsecondary challenges that await them. This reality—combined with disheartening trends in graduation rates and an increasingly global economy— informs and underscores the current national educational reform agenda. In January 2010, President Obama announced an extension of the Race to the Top program, already the most ambitious reform effort in history. And the initiatives and programs comprising this movement continue to gain momentum. The Common Core State Standards, released this past June, have been adopted by nearly all states and the District of Columbia. Two state consortia have formed to develop "next-generation" assessments to measure student achievement against the Standards as part of the $350M Race to the Top Assessment Program. And more recently, The American Association of State Colleges and Universities, the Council of Chief State School Officers and the State Higher Education Executive Officers partnered to promote broad implementation of the Standards.

A top priority of these initiatives is the adoption and use of internationally benchmarked standards. Why are consistent, measurable standards important? According to Education Secretary Arne Duncan, the Holy Grail of education is to ensure that all high school students are adequately prepared to meet the challenges of postsecondary endeavors; that they are "college- and career-ready." For this reason, both Race to the Top requirements and Common Core criteria advocate standards that build toward and ensure readiness. If we do not adequately prepare our secondary students, we set them up for certain frustration and a higher likelihood of future failure.

Unfortunately, several indicators suggest that many high school graduates are ill-equipped to meet the challenges and seize the opportunities that await them. First, an alarmingly high percentage of high school graduates need to take remedial courses in reading, writing and mathematics upon entering technical schools, community colleges and four-year universities. According to Alliance for Excellent Education, 42 percent of freshmen at community colleges— and 20 percent of freshmen at four-year institutions —enroll in at least one remedial course (Alliance for Excellent Education, 2006). In addition, surveys of educators across the P–20 landscape reveal a schism between high school and college expectations. ACT recently reported that 91 percent of high school teachers believe that they are adequately preparing their students for college, whereas only 26 percent of college instructors believe that their students have been sufficiently prepared (ACT, 2009). These startlingly divergent perceptions throw into question the very concept or nature of "readiness."

So how, exactly, should we define "ready?" Numerous elements comprise readiness. However, one of the most important is the ability to read and comprehend complex texts. Whether a student is applying to a community college, attending an elite four-

year university, or entering the workplace or military, grappling with high-level texts is likely to be a major component of the experience. And a student's ability to understand said texts is one of the key predictors of success in these domains.

Clearly, in order for standards to be meaningful—and readiness to be achieved—focused attention must be paid to the text complexity continuum across the P–20 landscape. This imperative is being recognized as a top priority by reform programs and initiatives. The Common Core State Standards, for example, affirm that "by the time they complete the core, students must be able to read and comprehend independently and proficiently the kinds of complex texts commonly found in college and careers" (NGA Center and CCSSO, 2010).

The Common Core State Standards reveal two alarming trends in terms of text complexity across the P–20 continuum. First, over the last 50 years the text complexity of K–12 texts have trended downward (Chall, Conrad & Harris, 1977; Hayes, Wolfer, 1996; Williamson, 2008). Second, the text complexity of reading demands in college, careers, and citizenship have held steady or increased over this same time period (Hayes, Wolfer, & Wolfe, 1996).

With this in mind, educators must continually assess whether the reading demands placed upon secondary students are rigorous enough to equip them for the texts they will encounter in their postsecondary endeavors. One study in particular already has demystified reading requirements across the P–20 environment by quantifying the difficulty of representative text collections in various domains.

In his investigation of postsecondary text demands, Williamson (2008) analyzed broad samples of texts from the college, military, citizenship and workplace domains. Detailed information about the specific texts included in the study can be found in Williamson's paper. The median Lexile® measure for military texts is 1105L, while the median Lexile measure for citizenship texts is 1230L. As might be expected, workplace texts—with a median difficulty of 1260L—have the broadest range of difficulty, extending both below and above all of the other text collections. Further, though the median text demand is fairly uniform across the workplace, military and citizenship domains, it is higher for postsecondary education: 1295L for two-year institutions and 1395L for four-year institutions.

Williamson's examination of K–12 texts reveals a gap of 65L to 230L between the demands placed on high school seniors and the difficulty of postsecondary texts. To put this gap in perspective, a 250L difference between reader ability and text difficulty can cause a drop from 75 percent comprehension to 50 percent comprehension. This means that a successful high school senior confidently reading twelfth grade texts may enter

college several months later and encounter texts that result in less than 50 percent comprehension. Fifty percent comprehension causes confusion, frustration and feelings of inadequacy in most readers.

The sizable breach between high school and college text complexity explains the high percentage of students in remedial courses, as well as the different perceptions of readiness reported by high school teachers and college instructors. Educators in each domain assess student readiness based on the customary texts at their level, unaware that—in order to successfully comprehend postsecondary reading—students must take a giant leap instead of an incremental next step.

The consequences of this gap in reading preparedness are significant. Colleges suffer the economic burden of providing remedial instruction, and training programs have trouble recruiting suitable trainees. Even worse, many struggling students become disillusioned with postsecondary pursuits and fail to reach their potential.

The good news is that this unfortunate situation can be remedied. Progress already has been made in reconsidering the entire scope of the P–20 educational landscape and fostering cooperation between K–12 and postsecondary educators. Further, we have the tools to evaluate, reconsider and re-map desired reading growth trajectories. Education Secretary Duncan believes that educators concerned about the readiness gap will see data as a "boon," not a burden. Indeed, as the aforementioned studies indicate, we can begin to supplant vague and inconsistent labels like "proficient" with objective, empirical evidence of whether reading standards and goals are being met.

In short, we need to begin with the end in mind. If we adjust the desired endpoint of secondary reading growth, students will no longer blindly follow a path only to find that they are ill-equipped to handle the challenges that await them at their destination. Quantifying the reading demands on high school students and comparing them to the text demands of the postsecondary world is a necessary first step toward better conceptualizing reading requirements over the span of an education.

The next challenge for policymakers and educators is to "bend" the curve of student reading growth by elevating expectations in each grade to better align with future demands. Table 1 (on the following page) from the Common Core State Standards' Appendix A provides a text continuum by grade bands so that educators have a reliable road map to make sure students graduate high school with the reading skills they need to succeed in their postsecondary endeavors.

Table 1: Text Complexity Grade Bands and Associated Lexile Ranges (in Lexile measures)

Text Complexity Grade Bands	Previous Lexile Ranges	Lexile Ranges Aligned to CCR Expectations*
K–1	N/A	N/A
2–3	450L–725L	420L–820L
4–5	645L–845L	740L–1010L
6–8	860L–1010L	925L–1185L
9–10	960L–1115L	1050L–1335L
11–CCR	1070L–1220L	1185L–1385L

**Common Core State Standards for English, Language Arts, Appendix A (Additional Information), NGA and CCSSO, 2012*

Measuring student progress in reading empowers parents and educators by allowing them to track whether students are on the proper path toward their goals and quickly address any deviations from the desired rate of growth. Akin to retirement planning tools, reading measures help students project what they have to do to get where they want to be. By forecasting deficiencies in reading comprehension through empirical studies of reading growth, we can demystify the "readiness gap," raise the bar for reading achievement, and better prepare our students for success in their postsecondary endeavors.

REFERENCES

Alliance for Excellent Education. (2006). Paying Double: Inadequate High Schools and Community College Remediation. Washington, DC: Author.

ACT. (2009). ACT National Curriculum Survey ® 2009. Iowa City, IA: Author.

Chall, J.S., Conrad, S., & Harris, S. (1977). An analysis of textbooks in relation to declining SAT scores. Princeton, NJ: College Entrance Examination Board.

Hayes, D.P., Wolfer, L.T., & Wolfe, M.F. (1996). Sourcebook simplification and its relation to the decline in SAT-Verbal scores. American Educational Research Journal, 33, 489-508.

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

Williamson, G. L. (2008). A Text Readability Continuum for Postsecondary Readiness. Journal of Advanced Academics, vol. 19 (4), 602-632.

MetaMetrics Policy Brief | 1560L
March 30, 2011

Bending the Reading Growth Trajectory:

Instructional Strategies to Promote Reading Skills and Close the Readiness Gap

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D.

An *Education Week* article summarizes the postsecondary readiness gap in unequivocal terms: "High school completion does not equal college readiness" (Gewertz, 2011). In particular, research findings that quantify how well students need to read in order to succeed in postsecondary pursuits have finally dispelled the false assumption that graduating high school is sufficient to prepare students for the challenges of college and careers. In his investigation of postsecondary reading demands, MetaMetrics Research Scientist Gary L. Williamson, Ph.D. discovered a substantial leap in reading expectations and text complexity from high school to postsecondary domains— a gap large enough to help account for high remediation rates and disheartening graduation statistics (Williamson, 2008).

These new understandings serve as the cornerstone of the revised reading standards articulated in The Common Core State Standards, an unprecedented set of national guidelines intended to raise educational standards and ensure postsecondary readiness. Appendix A of the Standards—which have already been adopted by nearly all states—emphasizes that, by the time they graduate high school, "students must be able to read and comprehend independently and proficiently the kinds of complex texts commonly found in college and careers" while also noting "a serious gap between many high school seniors' reading ability and the reading requirements they will face after graduation" (NGA Center and CCSSO, 2010). Ambitious new reading goals reflecting the importance of text complexity are set forth in Appendix A for each grade (Table 1).

Now that the need for revised reading growth patterns has been acknowledged as a top national priority—and newly adopted standards have raised the bar for reading achievement— the great challenge for policy makers and educators in the coming decade is clear: we must "stretch" current reading growth curves and elevate expectations to better reflect and align with postsecondary demands.

Table 1: Text Complexity Grade Bands and Associated Lexile Ranges

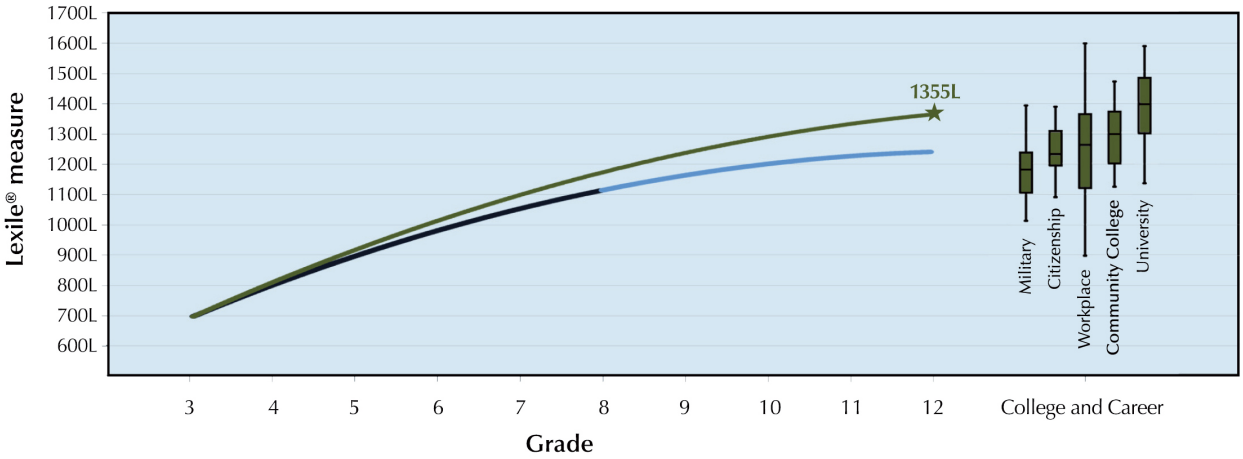
Text Complexity Grade Bands	Lexile Ranges Aligned to CCR Expectations*
K–1	N/A
2–3	420L–820L
4–5	740L–1010L
6–8	925L–1185L
9–10	1050L–1335L
11–CCR	1185L–1385L

**Common Core State Standards for English, Language Arts, Appendix A (Additional Information), NGA and CCSSO, 2012*

But *how* can we bring these new growth patterns into reality? What instructional strategies can educators employ to guide all students onto higher reading growth trajectories?

In order to evaluate and implement effective instructional strategies to support reading growth, it helps to understand the nature and characteristics of growth curves as they have existed historically. Extensive longitudinal research by Williamson over a period of fifteen years describes normative growth curves for five cohorts of students in North Carolina (Williamson, Thompson & Baker, 2006). In Figure 2, the solid blue line represents the average growth curve for a group of these students from grades 3-8. It is apparent that the impressive reading gains and high velocity of growth students demonstrate by the end of third grade are replaced by increasingly slower growth rates in the middle school years. If students remain on this trajectory, they can be expected to experience further deceleration through high school (as shown by the light blue line) and reach an average reading ability of 1275L at the end of twelfth grade. This level lies between the median complexity of workplace and community college texts, but falls short of the 1355L required for college and career readiness. (NGA Center and CCSSO, 2010).

Figure 2: Bending the Reading Growth Trajectory



This characteristic pattern of growth—strong in the early years and decelerating over time—is key to understanding which instructional responses will be most effective in helping students follow a more ambitious path: one that will prepare them for the higher text demands required for college and career readiness. Raising average reading scores in the early grades has long been a successful focus of educational research and reform. Many states, for example, have capitalized on the fast growth exhibited by young children by adopting early intervention initiatives. Strategies to raise early reading scores even further might include reviewing the K-3 curriculum to ensure that instruction is properly aligned with curriculum and assessment; and using a formative assessment system, which helps educators check for understanding and deliver instruction.

A greater challenge for reformers has been increasing the velocity or speed of student growth in reading. To accelerate growth, educators must first review and adjust the scope, sequence and pace of instruction: i.e., the breadth and depth of content to be covered, the order in which content is presented, and the speed at which material should be introduced.

Reading growth can also be addressed by exposing students to more complex text—especially in the middle and high school years—so that they have increased opportunities to stretch their skills. Unfortunately, as Appendix A of the Common Core Standards laments, "K-12 texts have actually trended downward in difficulty" and have become "less demanding" over the past fifty years (Chall, Conrad, & Harris, 1977; Hayes, Wolfer, & Wolfe, 1996). Intended to remove barriers to content with more accessible texts, this trend has had the unintended effect of hampering students' ability to tackle more challenging texts as they progress toward graduation. It should be noted that exposing secondary students to more demanding text no longer has to result in

discomfort, strain or frustration. With measurement tools like Lexile measures that help students determine their "just-right" reading range to enhance reading growth and lead to readiness, students can challenge themselves with success and a resulting sense of accomplishment.

Educators can also use benchmark assessments—tests given at prescribed intervals throughout the school year—to supplement end-of-year tests (which are more "summative" in nature). Benchmark assessments, like those administered by many districts to yield students' Lexile measures, are intended to measure progress along the way. They not only provide data on how to adjust instruction, but are also useful for providing information for district decision-making and evaluating student growth trajectories toward the textual demands of college and careers.

Perhaps most importantly, deliberate reading practice by students has shown promise in increasing the velocity of reading growth. Research suggests that the move from novice to expert in almost any domain (whether musical, athletic or intellectual) involves an intricate process in which practice must be targeted, intensive, distributed, self-directed, and be followed by feedback on one's performance (Ericsson, 2006; Glaser, 1996; Kellogg, 2006; Shea & Paull, 1996; Wagner & Stanovich, 1996). Utilizing a metric, like the Lexile measure, allows educators to match students to the appropriate level of challenging text and facilitates the sort of deliberate and targeted practice that promotes expertise.

Furthermore, rapid advances in technology and the emergence of personalized learning platforms have allowed educators to take individualized learning to scale, and facilitate personalized instruction by targeting learners at their current level to promote reading growth. These innovative web-based programs provide students with the targeted, intensive, self-directed practice and immediate feedback that are proven to support skill development.

As speed of reading growth is supported, the deceleration or slowing of growth rates over time must also be addressed. How can deceleration be mitigated so that the pace of growth is better maintained over time and robust throughout the later grades? For one thing, better vertical articulation of curriculum—supported by vertical systems of measurement— can eliminate gaps or unnecessary repetition in the learning sequence. Along the same lines, educators must rethink the outdated and potentially harmful belief that students *learn to read* until a certain point in their schooling and then switch to a model of *reading to learn.* This faulty philosophy—which underlies a diminished focus on reading skills in the later grades—results in increasingly less measurement, monitoring and explicit instruction in reading comprehension. A continued emphasis on *learning to read* even as students begin *reading to learn* is key to stemming deceleration

in reading growth. Why? Because comprehension remains a challenge and skills must continually be honed as students confront more difficult text.

Lastly, and perhaps most importantly, mitigating "summer loss"—the loss of skills experienced by most students over the summer months, especially in reading—is another very promising way to slow deceleration and maintain a higher pace of reading growth. The reading loss that occurs over the summer months has been well-documented and adversely affects students primarily from low-income households (Cooper, 1996). Research by James Sangil Kim, Ed.D., assistant professor of education at Harvard University, has shown that students who select books of interest to them and read steadily at appropriate levels over the summer months exhibit gains in reading ability (Kim, 2006). This kind of beneficial summer reading can be supported in numerous ways, including "Find a Book," a free, online search utility that allows students to build customized reading lists based on Lexile level and personal interests.

In order to ensure students graduate prepared for the reading demands of college and careers, we must bend current reading growth curves so that students attain higher skills by the end of high school. Since the effects of velocity and deceleration of growth accumulate over time, small but consistent modifications in either or both can have tremendous impact on reading achievement over the course of several years. The green curve in Figure 2—which represents a more ambitious, "aspirational" growth trajectory—makes clear that if students had a growth rate only 5L higher at the end of third grade, and could reduce their deceleration by less than 1L per year, their reading skills would reach the median text demand for college and career readiness.

By taking a longitudinal growth model perspective as Figure 2 illustrates, we can rationalize the reading growth demands over the lifespan of learners. This perspective is helpful because:

1. Reading skills development is a shared responsibility across the K-12 landscape, as opposed to only in the early elementary years.

2. Policy makers, educators, parents and students have consistent, objective metrics to monitor progress and forecast growth.

3. It demonstrates that even modest gains sustained over time can have a huge long-term impact and lessens the need for "Herculean" efforts as students get closer to graduation.

Research has already pointed us in the direction of the aforementioned instructional strategies, all of which can address the velocity and deceleration of reading growth in order to enhance comprehension skills and support students on higher trajectories. As

idealized growth trajectories are envisioned and adopted in response to the Common Core Standards—and states continue to collect more and better longitudinal data—we will be even better positioned to think strategically about how we can modify reading instruction.

Never has there been a better time to adopt a long-term perspective on reading growth in order to enhance student achievement and ensure college- and career-readiness.

REFERENCES

Chall, J.S., Conrad, S., & Harris, S. (1977). An analysis of textbooks in relation to declining SAT scores. Princeton, NJ: College Entrance Examination Board.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996).
The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. Review of Educational Research, 66, 227-268.

Ericsson, K., Charness, N., Feltovich, P., & Hoffman, R. (Eds.). (2006). The Cambridge handbook of expertise and expert performance. Cambridge: Cambridge University Press.

Gewertz, Catherine (2011). Success of College-Readiness Intervention Hard to Gauge. *Education Week,* January 26, 2011.

Glaser, R. (1996). Changing the agency for learning: Acquiring expert performance. In K.A. Ericsson (Ed.), The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games (pp. 303-311). Mahwah, NJ: Lawrence Erlbaum Associates.

Hayes, D.P., Wolfer, L.T., & Wolfe, M.F. (1996). Sourcebook simplification and its relation to the decline in SAT-Verbal scores. American Educational Research Journal, 33, 489-508.

Kellogg, R.T. (2006). Professional writing expertise. In K.A. Ericsson, N. Charness, P.J. Feltovich, & R.R. Hoffman (Eds.), The Cambridge handbook of expertise and expert performance (pp. 389-402). New York, NY: Cambridge University Press.

Kim, J. (2006). Effects of a Voluntary Summer Reading Intervention on Reading Achievement: Results from a Randomized Field Trial. *Educational Evaluation and Policy Analysis,* Vol. 28, No. 4, 335-355.

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

Shea, J. B., & Paull, G. (1996). Capturing Expertise in Sports. In K.A. Ericsson (Ed.), The Road to Excellence: The Acquisition of expert performance in the arts and sciences, sports and games (pp. 321-335). Mahwah, NJ: Lawrence Erlbaum Associates.

Wagner, R. K., & Stanovich, K.E. (1996). Expertise in Reading. In K.A. Ericsson (Ed.), The Road to Excellence: The Acquisition of expert performance in the arts and sciences, sports and games (pp. 189-325). Mahwah, NJ: Lawrence Erlbaum Associates.

Williamson, G. L., Thompson, C. L., and Baker, R. F. (2006, March). North Carolina's growth in reading and mathematics. Paper presented at the 2006 annual meeting of the North Carolina Association for Research in Education (NCARE), Hickory, NC.

Williamson, G. L. (2006). *Aligning the Journey With a Destination.* A white paper from The Lexile Framework for Reading. Durham, NC: MetaMetrics.

Williamson, G. L. (2008). A Text Readability Continuum for Postsecondary Readiness. Journal of Advanced Academics, vol. 19 (4), 602-632.

MetaMetrics Policy Brief | 1450L
September 28, 2011

Fulfilling the Promise of the Common Core:

Innovative Solutions to Eliminate the Readiness Gap in the "New Normal"

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D

At no time in the history of American K-12 education have the challenges and consequences of our actions been so great. As students across the nation settle into a new school year, administrators and educators are simultaneously attempting to implement the new Common Core State Standards in an economic climate plagued by budget shortfalls—an environment many economists have dubbed the "New Normal." While no consensus exists on whether our economy will double-dip or will just be slow to recover, what is certain is that educators will need to do more with less. How can educators and students realize the promise and potential of the Common Core in the "New Normal" when projected federal and state funding cliffs are quickly becoming reality?

Education Secretary Arne Duncan has stated that the Holy Grail of education is to ensure all high school students are adequately prepared to meet the challenges of postsecondary endeavors; that they are "college- and career-ready." To this end, both Race to the Top and the Common Core criteria advocate standards that build toward and ensure readiness. But, the reality is that high school completion does not equal college readiness, and an alarming number of students graduate unprepared for the academic and professional challenges that await them.

What may be equally alarming is that this lack of readiness is not a new phenomenon. Nearly 100 years ago, researchers Wilson and Hoke (1921) wrote:

> *The college instructor blames the high school teacher, the high school teacher complains of the grade teacher, each grade teacher above first grade finds fault with the poor work of the teacher in the grade below, and the first grade teacher in turn is chagrined at the shortcomings of the home training. Must this go on indefinitely? Whose opinion shall prevail? Is it not possible to get away from personal opinion to an agreed-upon consensus of opinion? May we not replace the constantly conflicting subjective standards with definitely defined objective standards?*

The lack of defined standards has been an inherent impediment to student readiness. With near nationwide adoption and, ultimately, implementation of the Common Core, there is great hope that this limitation of our education system will finally be addressed.

Table 1: Text Complexity Grade Bands and Associated Lexile Ranges

Text Complexity Grade Bands	Lexile Ranges Aligned to CCR Expectations*
K–1	N/A
2–3	420L–820L
4–5	740L–1010L
6–8	925L–1185L
9–10	1050L–1335L
11–CCR	1185L–1385L

**Common Core State Standards for English, Language Arts, Appendix A (Additional Information), NGA and CCSSO, 2012*

As noted in past policy briefs, closing the gap means bending the reading growth trajectory toward college and career readiness for all students. According to Appendix A of the Common Core State Standards for English Language Arts, this starts by engaging students in increasingly complex texts as they progress from grade to grade (see Table 1) (NGA Center and CCSSO, 2010). With objective reading metrics and clear standards to help chart this journey, we can now take a more longitudinal perspective that allows reading growth to be rationalized over each student's educational lifespan. This perspective is helpful because:

1. Reading growth is a shared responsibility across the K-12 continuum, as opposed to only in the elementary years.
2. Policy makers, educators and families have consistent, objective metrics to monitor progress and forecast growth.
3. It demonstrates that even modest gains sustained over time can have a huge long-term impact and lessen the need for "Herculean" efforts as students get closer to graduation.

With nearly all states having adopted the Common Core, national attention has shifted to the more difficult task of implementing these higher standards. This begins, in part, with allocating more instructional time. If funding issues were nonexistent, school districts would simply hire more educators to reduce teacher-to-student ratios and possibly add much-needed days to the school calendar. Unfortunately, our legacy public education system has maintained a conventional calendar of 180 days. And while this

calendar made sense in an agrarian society more than 100 years ago, it is no longer viable.

The "New Normal" requires us to find innovative solutions to eliminate the readiness gap. There are two promising, cost-effective strategies that can help us achieve the Common Core within today's financial and time parameters: personalized learning platforms and summer reading. Both approaches support "blended learning," which Michael Horn defines as: "any time a student learns at least in part at a supervised brick-and-mortar location away from home and at least in part through online delivery with some element of student control over time, place, path, and/or pace" (Horn, 2011). Personal learning tools and metrics already in use—and others being developed and tested—promise to revolutionize literacy learning and facilitate the necessary upward bending of reading growth trajectories by supporting key instructional strategies in and outside of the classroom.

Lexile measures, for example, evaluate reading ability and text complexity on the same developmental scale, enabling the targeting of reader with text essential for growth. As depicted in Figure 1, Appendix A of the Common Core State Standards for English Language Arts offers recommended Lexile bands by grade levels that can be incorporated into instruction designed to prepare students for the reading demands of college and careers (NGA Center and CCSSO, 2010). Lexile measures are available from nearly 50 popular reading tests and programs, including more than twenty state assessments and the most commonly used norm-referenced and benchmark assessments.

Personalizing Learning with Technology

The idea of personalized learning systems has existed since the 1950s, but implementations have never fully realized the ideals of their creators. While pioneers such as B.F. Skinner, Robert Gagne and Fred Keller defined the original design principles, their ideas were only crudely implemented. Most early efforts toward personalization merely allowed students to move through a course of study at their own pace by following tightly scripted pre-programmed material. Today, these limitations no longer apply. With improved technology, better understanding of instructional principles, and the advancement of psychometric theory and tools, there is capacity to realize the promise of what these early learning scientists initially conceived.

Metrics and research tools have already transformed the way educators think about reading growth and the benefits of matching readers with texts. The Lexile Frameworks for Reading and Writing, for example, are currently being used to power more integrated, interactive utilities: technology-based personalized learning platforms.

These innovative learning systems—which harness the power of technology and recognize the value of "personalized learning"—are already engaging and empowering students and are being embraced by more and more teachers and administrators. Organizations, including Capstone Digital and Achieve3000, have utilized The Lexile Framework for Reading to offer programs and products that continue to spur the personalized learning movement.

Another system that shows great promise is MetaMetrics' Learning Oasis™. This personalized learning platform uses the Lexile Frameworks to differentiate reading and writing practice and support overall literacy growth. The reading component of Learning Oasis immerses developing readers in text of high interest or content relevance that is targeted to each reader's ability. Articles from periodicals and newspapers are classified by Lexile measure and category. During their engagement with articles in Learning Oasis, students are periodically presented with auto-generated "cloze" items: passages in which selected words have been deleted. Students are prompted to fill in the missing words and the system uses these answers to generate an updated Lexile measure. In this way, without students experiencing the pressure of being "tested," growth is consistently measured and monitored. Learning Oasis automatically offers selections of appropriate and increasing complexity based on continuing growth.

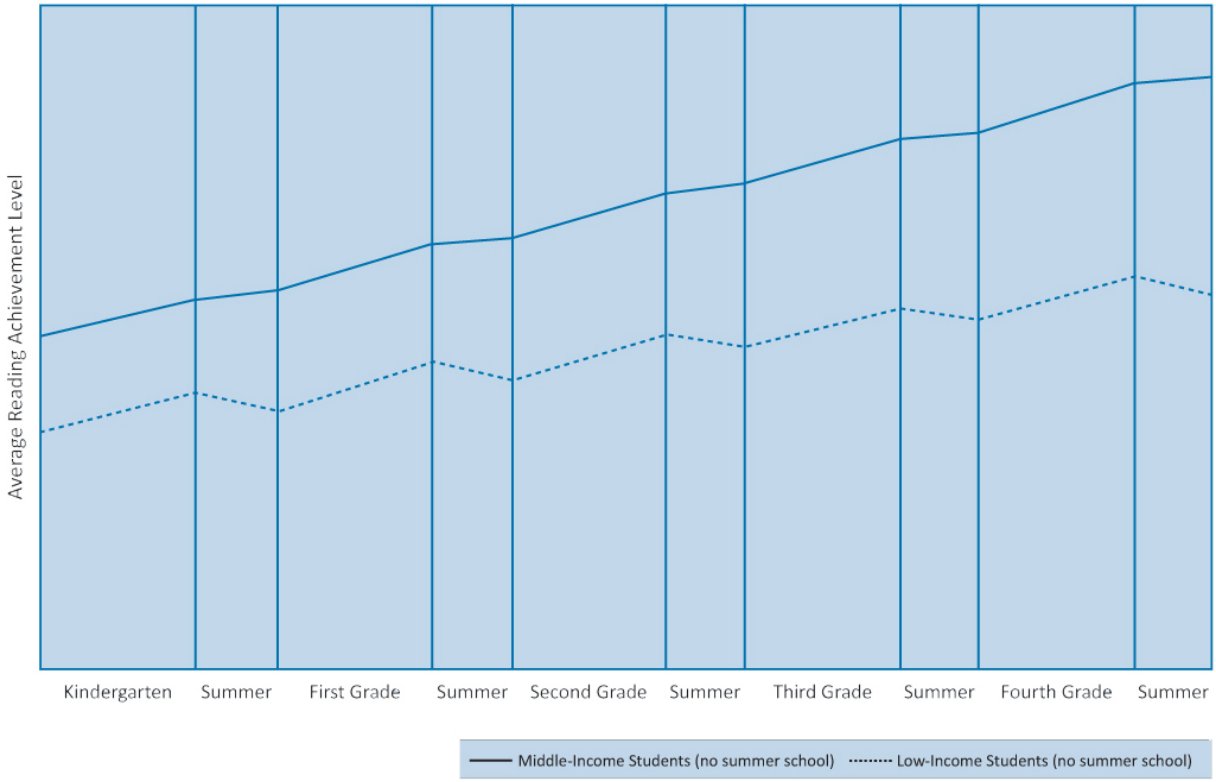
Computer-based personalized learning platforms promise to meet the needs of both students and teachers. By elegantly blending assessment with daily classroom instruction—and customizing the learning experience for each student— personalized learning platforms take differentiated instruction to a new level and will likely revolutionize and redefine the way we think about learning and assessment. While educational standards may be laudable, there is no "standard" student. Each has a unique profile of aptitudes, weaknesses, motivations, needs and interests, which is why many of the most influential leaders in education believe that the best way to ensure student success and achievement is through personalizing the learning process.

Extending Learning Over Summer

No matter what type of academic calendar a state or school district uses, U.S. students attend school every year for an average of 180 days. During that time, most students achieve some level of growth in their reading ability and mathematics achievement. Then summer break starts, the formal learning process ends, and, instead of progressing, students start to slip in their abilities. Research shows that all students experience some level of learning loss—commonly referred to as summer slide—when they do not engage in educational activities during the summer. While the slide in mathematics is universal, the summer reading slide disproportionally affects low-

income students—those who generally reside in low-literacy environments where daily reading is not encouraged or modeled and appropriate reading materials are not available. Research indicates that two-thirds of the reading achievement gap can be explained by the cumulative impact of these lost summers (Alexander, Entwisle & Olson, 2007). Figure 2 illustrates how summer loss significantly affects low-income students (adapted from McLaughlin & Brady, 2006).

Figure 2: Summary of Reading Achievement Trajectories



Based, in part, on the research of Harvard University's James Kim, MetaMetrics developed a free book search utility, called "Find a Book," that allows students, teachers and parents to build personalized reading lists based on interests and Lexile level, and to find their reading selections at the public library. Kim's research demonstrated that if children read high-interest, ability-appropriate books during the summer their reading skills can grow as much as their peers who attended summer school (Kim, 2005). With more than twenty states reporting Lexile measures from their year-end tests, many of these states have now bypassed the status quo of simply encouraging students to read over the summer and have opted instead to promote "Find a Book" for more personalized reading experiences. Other freely available tools for creating personalized reading lists include Barnes & Noble's Lexile Reading Level Wizard and Scholastic's Book Wizard.

The Potential for Revolutionary Change

According to the National Education Technology Plan (NETP), the use of technology in schools does not sufficiently reflect or build on the ways students use digital tools in their lives outside school or how technology is used in the professional world. This represents a "gap" as significant and detrimental as the postsecondary readiness gap—one that must be bridged with "revolutionary transformation rather than evolutionary tinkering" (U.S. Department of Education, NETP, 2010, p. 3). Supporting a similarly revolutionary approach, Susan McLester states in *District Administration* that "personalized learning represents a sweeping, systemic change to American education" (McLester, 2011, p. 45).

Technology-based learning platforms can serve as the cornerstone of this revolutionary change. By harnessing the power of technology, they have the potential to personalize the learning process; support teachers in enacting best teaching strategies; and help students meet ambitious and rigorous standards.

Perhaps, most importantly, by offering "anytime, anywhere" access, personalized learning platforms and customized applications can change our ideas about where and when learning should take place, and thereby answer the call of the NETP to embrace online learning as a way to "extend the learning day, week or year." The Plan's proposed model for an infrastructure for learning is "always on, available to students, educators and administrators regardless of their location or the time of day" and "enables seamless integration of in- and out-of-school learning" (U.S. Department of Education, NETP, 2010, p. xiii).

The great hope—and promise—is that technology can supplement teacher-directed instruction in such a way that what is learned from teachers is reinforced by technology; that students will finally enjoy ownership and responsibility for their own learning; and that teachers will be freed to do what they do best: guide, inspire and mentor their students.

REFERENCES

Alexander, K. L., Entwisle D. R., & Olson L. S. (2007a). Lasting consequences of the summer learning gap. American Sociological Review, 72, 167-180.

Fairchild, R. McLaughlin, B. & Brady, J. (2006). "Making the most of summer: A handbook on effective summer programming and thematic learning." Baltimore, MD: Center for Summer Learning.

Horn, M. The rise of K-12 blended learning. The Huffington Post. January 1, 2008. http://www.huffingtonpost.com/michael-horn/the-rise-of-k12-blended-l_b_814951.html.

Kim, J.S. (2005). Project READS (Reading Enhances Achievement During Summer): Results from a randomized field trial of a voluntary summer reading intervention. Paper presented at Princeton University, Education Research Section, November 7, 2005.

McLester, S. (2011, March). Learning gets personal. District Administration, 47(3), 30-45.

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

U.S. Department of Education, Office of Educational Technology, Transforming American education: Learning powered by technology (National Education Technology Plan), Washington, D.C., 2010.

Wilson, G.M., & Hoke, K.J. (1921). How to measure. New York: The Macmillan Company.

Transitioning from Adoption to Implementation of the Common Core State Standards

by Malbert Smith III, Ph.D., MetaMetrics President and Co-founder

Since the release on June 2, 2010 of the Common Core State Standards (CCSS), there has been a great deal of attention devoted to the five Ws of reporting (NGA and CCSSO, 2010). That is, *who* created them, *what* they embody, *where* you can find them, *when* they were created, and *why* they were created. Unfortunately, as we have passed the two year anniversary of the CCSS, the *how* to implement them has been largely neglected. In the marketing and promotion of the CCSS, careful attention has been rightfully devoted to the five Ws and today one would be hard pressed to find an educator who, when asked, could not recite the answers to the five Ws. However, those same educators would likely be unable to tell you *how* they will implement these standards in a real classroom. When we conduct a Google search on the CCSS, one finds 13,400,000 hits[1], however upon closer examination, the first few hundred sites all focus on the five Ws—not the H.

The success of the roll-out campaign for the CCSS is reflected in the fact that 46 states have adopted them (Wilhoit, 2012). At this point, it is likely that the few remaining holdout states will eventually adopt the CCSS as the standards become a common touchstone among educators throughout the nation. But the ultimate success of the CCSS movement will now largely depend on the next critical phase—the shift from adoption to implementation—how to use them.

What will teachers do differently in a post-Common Core world from what they did in the pre-Common Core world? Are there actionable tools, resources, assessments, and curriculum material aligned and predicated upon the new standards? On the assessment side, two testing consortia (Partnership for Assessment of Readiness for College and Careers and SMARTER Balanced Assessment Consortium) have embarked on a mission to create new statewide assessment models that align with the new standards. The earliest timeframe for these two groups to launch their assessments is the 2014-15 school year. Likewise, the educational publishing industry is racing to create and refresh their resources in an effort to provide such alignment as well.

While we enjoy tremendous success in the acceptance and adoption of the CCSS, we now enter the much more challenging time of how to effectively implement them. At this point we should apply the historical lessons we have learned from previous adoption movements to ensure that we get this done right. One of the most ambitious, yet disappointing adoption movements in our country was the failure to implement the metric system. Even with congressionally authorized committees and legislation, our country never made the conversion to the metric system. In 1968, Congress authorized a three year study to examine the feasibility of adopting the metric system (NIST, 2002). This study culminated in the auspiciously titled report, "A Metric America: A Decision Whose Time Has Come" (NIST). In 1975, Congress passed the Metric Conversion Act

[1] Search conducted June 20, 2012. Google yielded 13,400,000 search results.

"to coordinate and plan the increasing use of the metric system in the United States" (NIST). Growing up during this period, I remember seeing new sign construction on the highways to reflect the transition to the metric system. When was the last time you saw a highway posted speed limit sign expressed in the metric system in our country? While the adoption of the metric system is not totally dead, in that our government has continued to promote the metric system as the "preferred system of weights and measures for United States trade and commerce," the acceptance and implementation by the overall public has been forfeited (NIST).

In every standard adoption movement there are critical "tipping point" moments and actions in which the movement takes hold or withers. From the unification movements associated with the measurement of temperature and standardized time, one can see the need for compelling and practical applications, the importance of underlying science, and the necessity of public and professional commitment. If the current movement of the CCSS is to realize its potential and promise, it is incumbent upon educators, policy makers, educational entrepreneurs, and the publishing industry to move as expeditiously as possible to build actionable tools and resources for educators and parents.

At MetaMetrics we have been actively engaged in the *how to* part of the Common Core Standards since its inception. On the English Language Arts side, our research on how well one needs to read for college and career readiness is one of the core components in the standards. Table 1 displays the associated Lexile range by grade band. To facilitate the practical application of text complexity, we have provided a free web-based tool (the Lexile® Analyzer) that allows educators to measure text. Nearly 145,000[2] educators and other individuals have registered to use this tool, and over 1.6 billion words have been analyzed. Publishers and digital content providers have also utilized the Lexile Analyzer so that hundreds of thousands of books and hundreds of millions of articles now have Lexile measures. Additionally, more than 135,000 titles with Lexile measures are available to search on the popular "Find a Book" tool. "Find a Book" enables users to build custom reading lists based on Lexile range and personal interests and to check the availability of books at the local library.

[2] Reflects number of users registered for the Lexile Analyzer as of July 15, 2012

Table 1: Grade Ranges for Lexile Measures

Text Complexity Grade Bands	Lexile Ranges Aligned to CCR Expectations*
K–1	N/A
2–3	420L–820L
4–5	740L–1010L
6–8	925L–1185L
9–10	1050L–1335L
11–CCR	1185L–1385L

**Common Core State Standards for English, Language Arts, Appendix A (Additional Information), NGA and CCSSO, 2012*

To truly support differentiated instruction in reading, it is not sufficient to only measure the text complexity of content. It is imperative that student reading ability be expressed on the same scale as text complexity so readers and text can be matched. Fortunately, there are over 50 publishers of reading assessments that elect to report Lexile reader measures. These tests range from standardized norm-referenced publishers to interim test publishers and even statewide No Child Left Behind assessments. For a complete list of assessments, please visit: www.Lexile.com.

On the mathematics side, educators, students and parents can go to www.Quantiles.com for access to an array of educational resources and tools that support differentiated instruction aligned to the Common Core. Built upon the same psychometric principles as the Lexile Framework, the Quantile® Framework allows educators to place a learner's mathematical knowledge and math content on the same developmental scale so a match can be affected.

Currently, there are six end-of-the-year state accountability assessments (Kentucky, North Carolina, Oklahoma, Virginia, West Virginia and Wyoming) and two publishing companies with norm-referenced assessments (ERB and The Riverside Publishing Company) that report Quantile measures of students. American Education Corporation, Measured Progress, Scholastic and Cambium Learning Group are four of MetaMetrics' partners that provide Quantile-based interim assessments for grades K-12.

More than 580 textbooks, 64,000 lessons and 3,100 downloadable resources that have been calibrated to the Quantile scale are available. Teachers can also utilize the following free tools:

1. The Math Skills Database: A great resource for teachers and parents. Search the QTaxon (math skills) database for math skills and aligned resources.

2. Math@Home®: Find family-friendly math activities that match your child's Quantile measure or math level.

3. Find your Math Textbook Tool: Offers teachers a Quantile measure for every lesson in their textbook that will help to inform their instruction and offer ideas for differentiated lessons.

4. Quantile® Teacher Assistant: This easy to use tool helps educators utilize the Quantile Framework for Mathematics to differentiate math instruction and to locate resources that can help identify those skills that are most relevant to the topic of the day.

Our organization is committed to building resources and tools that support differentiated instruction in a Common Core world. To realize the full potential and to gain traction with educators it is imperative that we not only discuss and analyze the Common Core State Standards, we must provide educators with readily accessible tools, modules, and instructional resources that translate the theory into practice.

As Marc Tucker pointed out in "Standing on the Shoulders of Giants," of all the reform efforts we are implementing in our own country, only the Common Core State Standards is consistent with practices of the highest performing countries (Tucker, 2011). As Secretary Arne Duncan stated, the "Holy Grail" of education is to ensure that students graduate from high school college- and career-ready. The first step in capturing this Holy Grail was the creation and adoption of the CCSS. However, this is only the first, and likely the easiest step.

In many respects, the failure of the metric movement in America was the latency of conversion. If we follow the same path of that movement and proceed without creating the actionable CCSS-aligned tools and resources, educators, out of necessity, will resort to teaching and assessing the same content in the same way. The window of opportunity will have closed and we will have lost the momentum driving us through this historic moment of transformation. The opportunity to create lasting educational change will go the way of the metric system in America. Moving forward, it is my recommendation that we focus on the H—*how* to implement. If we get the *how to* right, not only will we restore our educational leadership in the world, but we will go a long way in maintaining our position of economic leadership around the world.

REFERENCES

National Governor's Association, Council of Chief State School Officers. (2010). *www.corestandards.org*. Retrieved June 21, 2012, from Common Core State Standards Initiative: http://www.corestandards.org/news

National Institute of Standards and Technology. (2002). *The United States and the Metric System: A Capsule History*. Retrieved June 21, 2012, from Toward a Metric America: http://www.nist.gov/pml/wmd/metric/upload/1136a.pdf

Rowlett, Russ. (2000). *The Metric System in the United States*. Retrieved June 21, 2012, from www.unc.edu: http://www.unc.edu/~rowlett/units/usmetric.html

Tucker, Marc. (2011). *Standing on the Shoulders of Giants: An American Agenda for Education Reform*. Retrieved June 21, 2012, from Learning Matters: http://www.learningmatters.tv/images/blog/Standing.pdf

Wilhoit, Gene. (2012). *States Raise the Bar With Standards Implementation*. Retrieved June 21, 2012, from Education Weekly: http://www.edweek.org/ew/articles/2012/06/22/36wilhoit.h31.html

MetaMetrics Policy Brief | 1490L
February 27, 2013

Looking Worse Before We Look Better: Student Performance in a Common Core World

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D., and Director of Professional Development Jason Turner

While political leaders and pundits wrangle over the various impending fiscal cliffs, it is important to call attention to a lesser known, but just as urgent, impending cliff in education. First, it is widely hoped that in 2013 we will finally see the reauthorization of the Elementary and Secondary Education Act (ESEA). The failure of Congress to reauthorize ESEA has led the Department of Education to circumvent issues of compliance by granting state waivers. However, these waivers are largely temporary, allowing states some breathing room until the reauthorization of ESEA can be addressed. The second critical event regards the final development and release of the common core assessments by two testing consortia, the Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced Assessment Consortium (SBAC). Thoughtful and careful deliberation of both of these factors will be required if we are to avoid a storm of uncertainty and confusion.

It has been almost a decade since Congress last reauthorized the Elementary and Secondary Education Act (ESEA) with the passage of the No Child Left Behind Act (NCLB). President Bush signed it into law on January 8, 2002, and at that time it represented a sweeping overhaul of federal efforts to support elementary and secondary education in the United States. Undertaken as a major effort to enhance the academic performance of *all* children, NCLB had many important and promising features that Congress hopefully retains when addressing the reauthorization—features such as disaggregation of data, annual assessments and expanded parental options. However, there are at least three flaws in the current bill that need to be corrected.

The first flaw in NCLB is its allowance for each state to define and set their own internal standards for proficiency. Fifty states determining their own "proficiency" standards, combined with the "proficiency" standards associated with the National Assessment of Educational Progress (NAEP), has resulted in an academic Tower of Babel when it comes to measuring progress in reading and math in our country.

The failure to establish clear, national standards of proficiency yielded highly disparate definitions of "proficiency" and, in practice, led to radically different ideas on what it meant to be proficient in reading and math. The unintended consequence of allowing states to set their own proficiency standards was that a "proficient 4th grade student in reading in Mississippi" was not the same as a "proficient 4th grade student in reading in Connecticut." Even worse, the proficiency level set by the majority of states was not consistent with the rigor of NAEP. This lack of a common scale led to a general confusion among the public about whether our academic performance actually improved or not. In a 2004 white paper, *The Need for Objective Measurement Under the No Child Left Behind Act*, Malbert Smith wrote:

> The serendipitous benefit of the high-stakes consequences of NCLB is that it will expose one of the most profound limitations of measurement in the social sciences: The lack of unification of metrics (universal and standard scales). Without universal, exchangeable scales in the social sciences, our assessment systems across states may employ the same labels (advanced, proficient, basic and below basic), but may vary dramatically in the achievement implied by these labels.

The second inherent flaw in NCLB is that it permitted states to back load their growth into the outer years. In other words, as Chester Finn of the Thomas Fordham Institute noted, this feature essentially allowed states to assume a balloon mortgage-type model when it came to ensuring that every student was proficient by school year 2013-14. Instead of requiring an equal amount of annual growth (like a fixed mortgage), much of the growth was deferred until the later years. An unintended consequence of this approach is that it became increasingly difficult for states to hit annual yearly progress (AYP) targets with each passing year.

A final and fundamental flaw of NCLB is that it led states to build accountability models that focused on status measures of student performance as opposed to growth measures. Within the 670 pages of NCLB, terms like "achievement," "progress," "learning growth," and "development" are used over 1,660 times. Unfortunately, these terms are used interchangeably, as if there is no distinction between growth and status. The bill could have been more accurately titled No "Cohort" Left Behind. The real and substantive concern of educators and parents involves growth, not status. Obviously, the reauthorization of a new bill presents an opportunity for correcting these previous flaws.

Around the same time that Congress will address these issues, we will be approaching the release of the common core assessments by PARCC and SBAC. While there is a healthy sense of optimism about these assessments, there are a number of nagging concerns that need addressing. The first and most obvious issue concerns cost. In today's economic environment, all states are seeking to reduce the amount they spend on their annual assessments. Though most educators are optimistic about the promise of these long-awaited assessments, many privately worry that ballooning costs—costs far greater than many current state assessment programs—will render their state unable to afford the high price tag. While those fears are completely reasonable, a thoughtful analysis by Linda Darling-Hammond of Stanford University indicates that while the two assessments are likely to be more expensive than current assessments, there is good reason to believe that a number of creative solutions, including cost-sharing among states, may offer ways to significantly lower cost.

Assuming that cost can be brought in line with current state assessment budgets and that the assessments pass basic psychometric assumptions of reliability and validity, a major remaining issue will be how to communicate the assessment results to the public. If one subscribes to the notion that the "proficiency" levels on PARCC and SBAC assessments will likely be more rigorous than what most states have previously set, then great care and attention must be paid to how we communicate these results. To paraphrase Warren Buffett, only when the tide goes out can you tell who has been swimming without a bathing suit. The release of two "national assessments" will take the tide out! For example, assume a state reports that 75 percent of their third graders are proficient in reading in 2013. Then, in subsequent years, the state reports that (under the new assessment program) only 50 percent are proficient. To the uninitiated, it appears that the state has suffered a precipitous decline in reading growth, when, in fact, it is the benchmark for what constitutes "proficiency," along with the introduction of a common scale, that has been raised and is responsible for a perceived drop in student performance. It is easy to imagine how, divorced from context and history, these reports will provide headline fodder and can fuel an inaccurate picture of state educational progress.

When it comes to the implementation of the new assessments, it is helpful to take a page from the book of Wall Street. Publicly traded companies know the importance of managing expectations, and it is critical that education departments proactively shape expectations. Education departments would do well to take the lead in informing the public about what to expect from the reauthorization of ESEA and what the adoption of our new 'national assessments' will mean for the measurement of growth. Central to that effort will be the need to communicate the reason why state-specific multiple scales simply will not suffice as tools to measure our students' progress in reading and mathematics. As Smith wrote in 2004:

> Consequently, the real reason that the multiple measures requirement is on such a slippery slope is that our instruments do not have exchangeable scales. Without standard objective scales, like those employed in the hard sciences, educators will be left with less-than-satisfactory methods and very confusing, complicated schemes for reporting such data.

In fact, given the lack of exchangeable, highly disparate scales, the public should *expect* to see dips in student performance as we transition to a common metric and a unified way of measuring student growth.

While most of the nation's attention has been focused on the financial cliffs (debt ceiling, sequestration, etc.), educators should pay special attention to the fog of uncertainty surrounding the reauthorization of NCLB and the release of the PARCC and

SBAC assessments. A recent Gallup Poll indicates that the public's confidence in our schools is at an all-time low. We analyzed over 40 years of Gallup data on this issue, along with empirical assessment data, to see if such a dismal public view was justified. We concluded our analyses with the steps that we think are needed to restore public confidence in education. In this transition into what could be a transformative moment, we are concerned that no child be left behind and that each student graduates college and career ready. During this time of transition—as we move toward a higher level of academic rigor and proficiency standards—we may look worse before we look better. Getting this message right and helping the public understand this important transition will be necessary if we are to avoid further misperceptions and misunderstandings about the great work our nation's educators are doing.

REFERENCES

Darling-Hammond, L. & Adamson, F. (2010). "Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning." Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Finn, Jr., C. (2009). "Can we get to national standards, considering the pitfalls?" *The Education Gadfly*, The Thomas B. Fordham Institute.

Smith, M. (2004). "The Need for Objective Measurement Under the No Child Left Behind Act." MetaMetrics, Inc. Smith, M., Turner, J., & Lattanzio, S. (2012). "Public Schools: Glass Half Full or Half Empty?" *Education Week*.

MetaMetrics Policy Brief | 1390L
March 13, 2013

Looking Worse Before We Look Better Part II: Communicating Student Performance in a Common Core World

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D., and Director of Professional Development Jason Turner

It is widely accepted that the adoption and implementation of the Common Core State Standards will raise the academic bar of rigor and demand for U.S. students. Most have reasonably assumed that the accompanying common core assessments—Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced Assessment Consortium (SBAC)—will reflect that increased rigor. While there is a healthy sense of optimism building around the release of these two assessments, there is a general concern that what constitutes "proficiency" on these two assessments will be far more rigorous than what most states have previously set. One likely, unintended consequence of the release of the new assessments is that states around the country will appear to have suffered a precipitous decline in academic progress (Smith & Turner, 2013). Divorced from context, however, those assessment scores threaten to paint an inaccurate picture of student growth.

As an example of how assessment results may be misconstrued, consider a state that reports 75 percent of their third graders were proficient in reading in 2013. Then, in 2014, the state reports that (under the new assessment consortium program) only 50 percent are proficient. To the uninitiated, it appears that the state suffered a marked decline in reading growth, when in fact, it is the benchmark for what constitutes 'proficiency,' along with introduction of a common scale, that has been raised and is responsible for the perceived drop in student performance.

Fear that the transition to a new set of assessments may initially present an inaccurate picture of state educational progress, a picture that makes for easy headline grabbing, is nothing new. In fact, Wayne Camara and Emily Shaw made a compelling case that when it comes to reporting on educational test data, mainstream media sources, more often than not, cite the opinions of "testing opponents as the sole commentary in their articles on testing policy" (Camara & Shaw, 2012). In terms of the general public, a 2010 survey found that 75% receive their news on education topics from families and friends (Camara & Shaw, 2012). Meaning, perception matters. The perception that a state's academic performance is suddenly sliding in the wrong direction can be devastating to educators at all levels, from the classroom to the district office to the state department of education. Educational policy makers are well aware of this public perception issue. As a recent article, The Schoolmaster, in *The Atlantic* makes clear, even one of the primary authors of the Common Core, David Coleman, admits the transition to PARCC and SBAC will initially be perceived as an academic stumble:

> Coleman admits that the Core will probably lead to "a short-term reduction in [test] scores," but he seems to have made peace with this reality as a necessary hardship on the road to his academic utopia.

In that same article, Jeb Bush, a well-known advocate for education reform, admits the terrible political pressure that state and educational leaders will face in the coming years:

> "The big fight will be coming in 2014, when we begin to implement and assess these standards. If a third are ready, what will the response of states be then? Will they do what they historically have done, which is to pull back and say 'Oh my God, that's not fair, excuse, excuse, excuse'? Or will they accept responsibility to say 'That's the fact, that's where we are now. Maybe 40 percent of our kids are ready if we benchmark them against the world'?"

Bush's tough-love position is easy for a former politician to take, but less so for a current elected official. After all, no governor or legislator wants to preside over plummeting test scores. Pressure to roll back the Common Core or to relax the tests may be intense.

This perception, of course, is unfair. Most educators *expect* student assessment measures to decline under the new assessment scheme. This is not necessarily because student *performance* is actually falling or because students are actually performing worse. It is merely because the educational bar used to measure student performance, the benchmark for proficiency, has risen. However, these more sophisticated and nuanced explanations make for lousy sound bites. For the uninitiated public (including mainstream media), one unfamiliar with basic principles of assessment and psychometrics, a much more attention-grabbing story is that the educational sky is falling. Thus, educators nationwide are bracing for the impending backlash.

Of course, it does not have to be that way. Stephen Sireci has argued that even abstract philosophical and educational issues can be communicated to mainstream media in more palatable ways. Sireci recommends employing three strategies: the use of plain language to convey assessment information, the use of visuals, and the use of stories and examples (Sireci & Forte, 2012). Camara and Shaw demonstrate a similar understanding in their call for education departments to engage with the mainstream media:

> It would also be worthwhile to consider holding a pre-session with the media if you are introducing a new assessment or if there will be major changes to the test. If the media are not familiar with the assessment or there have been major changes in the assessment it is likely they will have preconceived ideas about the results, which are easier to address before results are released and reporters are on a deadline.

The challenge for educators is to proactively manage public perceptions *before* the release of the assessment results, and to communicate how the transition to the common

core creates a higher bar for U.S. students. When it comes to the implementation of the new assessments, it is helpful to take a page from the book of Wall Street. Publicly traded companies know the importance of managing expectations. Similarly, it is critical that education departments proactively shape expectations. Education departments would do well to take the lead in informing the public about what to expect from the adoption of our new 'national assessments' and what that adoption will mean for the measurement of growth.

Following Sireci's recommendations, part of that state effort to effectively communicate should involve the use of clear analogies and examples. Analogies are rhetorically powerful and, when used properly, can be instructive and illuminating. Consider a simple analogy taken from track and field. Imagine a track coach who sets a proficiency benchmark of three feet for her hurdle jumpers. Using a simple process, she is able to identify the percentage of athletes who are able to clear three feet and are thus labeled proficient. Now imagine a new year brings a new coach who feels that three feet is no longer a competitive height in the world of track and field and so updates the proficiency benchmark to three feet, six inches. Because the bar has been raised (literally in this case), a fewer number of athletes may now qualify as proficient. However, to an outside observer it appears that the percentage of proficient track and field athletes has declined. In fact, the athletes have the same ability they had before, and it is unlikely that any athlete actually lost athletic ability. It is simply that what we take to constitute proficiency has been altered.

The common core assessments present educators with the same bitter but necessary pill: if we want U.S. students to graduate college and career ready in an increasingly global and competitive world, then more rigorous academic standards are required. That rigor demands that we raise the proficiency benchmark for all our students in the interest of long term gains. In the short term, however, transitioning to the new common core assessments may give rise to the perception that our educational progress has faltered.

It could have been different. Such a communication nightmare could have been avoided if over the life of *No Child Left Behind* (NCLB) we had taken a growth perspective and not a status perspective in describing student achievement. State departments of education could have focused on longitudinal growth over cross-sectional growth, and used vertical scales. As Smith wrote in 2004:

> Consequently, the real reason that the multiple measures requirement is on such a slippery slope is that our instruments do not have exchangeable scales. Without standard objective scales, like those employed in the hard sciences, educators will be left with less than satisfactory methods and very confusing, complicated schemes for reporting such data.

To punctuate the problem with a temperature analogy, imagine a world in which weather reporters did not report the actual temperature but just categorical labels (hot, warm, cool, cold). What Floridians consider "cold" would no doubt be very different from the actual temperature that a Minnesotan would report as being cold. However, even in this hypothetical world of temperature-free weather reporting, there is an obvious path away from the confusion: underlying these descriptive labels are vertical and exchangeable scales (Fahrenheit and Celsius). That is, we could explain that Floridians feel like it is cold when the temperature is 48°F and below, whereas Minnesotans consider it cold when it is 32°F and below. Even if weather forecasters in Minnesota used the Celsius scale instead of Fahrenheit, because the two scales are exchangeable, we could still impose order on these different perceptions.

By allowing each state to define its own level of proficiency, we created an academic Tower of Babel of reading achievement in this country. Unfortunately, the basic measurement principles that are commonplace and essential in the "hard sciences" have been for too long ignored in the "soft sciences." However, when it comes to the measurement of reading, that neglect is no longer justified. The Lexile Framework for Reading is a vertical, developmental scale that allows the measurement of students and text on a common scale.

Today, there are 20 states that report Lexile measures on their NCLB assessment. For the states already employing the Lexile metric, the message can be clear and concise as they transition to the more rigorous demands of the CCSS. To revisit the hurdle analogy, a state can report that last year's proficiency hurdle was a Lexile level of X, but that hurdle has been raised to Y. Additionally, the state can report that while the hurdle has been raised, student performance did not actually slide, and that the average Lexile measure for students is (hopefully) higher than the previous year. In other words, state officials can explain that their students are jumping as high (or higher) than last year, but we have simply raised the hurdle requirement. These sorts of clear analogies simplify the message and allow for better understanding.

Additionally, the fact that the Lexile scale can also be used in the measurement of text complexity provides students, parents and educators with a concrete, interpretive framework. This feature of the conjoint measurement of reading comprehension and text complexity is a critical component in realizing the full potential of the CCSS. As Standard 10 in the Common Core State Standards for English Language Arts/Literacy Appendix A indicates, the ability of students to read increasingly complex texts is one of the major shifts from previous standards (National Governor's Association, 2012). As ubiquitous as Lexile measures are in the measurement of student reading ability, on the text side, Lexile measures are even more pervasive. Today, every major text and trade book publisher have measured their books using the Lexile Framework. In addition,

millions of articles have also been measured and are freely available from library resource groups like EBSCO and ProQuest.

The total and complete implementation of CCSS will require incredible commitment, time and persistence. To facilitate this Herculean effort, our organization has provided frictionless and free resources to educators as they align and build new resources. This effort has resulted in over 100,000 educators, parents and students measuring over 700 million words through the fourth quarter of 2012.

The conjoint feature of measuring text and students on the Lexile scale, combined with the fact that the CCSS identified the Lexile level per grade for college and career readiness, provides students, parents and educators with clear empirical and external benchmarks. No longer will students, parents and educators have to guess and wait for some annual assessment to determine if they are on track for college and career readiness. Furthermore, like the adoption of the CCSS, the adoption and acceptance of Lexile measures has been a voluntary movement spawned by educators that recognized the utility and value of the measures for their classroom practice.

It may seem unimaginable to us now, but the unification of measurement has been a slow road. There was a time in our history when we did not have unification of the measurement of constructs such as temperature and time. In the early 1600s there were dozens of instrument makers measuring temperature on their own scale. However, once our scientific understanding of temperature was established, there was recognition that we did not need all these different instrument makers, but common scales. Additionally, as late as the 1850s, we did not have standardized time. A Canadian engineer, Sanford Fleming, frustrated with translating all the different railroad time tables in order to figure out when his nephew would arrive at the train station, convened a world summit to fix this problem by creating standard time zones.

Now the time has come in education that we learn the lessons from the measurement of temperature and time, and realize the importance of the unification of measurement. The education of our students will be best facilitated not by adding more tests with various scales, but by placing the tests that we administer on a common scale. Because the Lexile Framework for Reading is a vertical scale used universally by publishers and educators, the benefits of measuring longitudinal growth across an array of assessments is now possible. The conjoint feature of the measurement of text and individuals on the same scale permits a rich, interpretive framework for understanding test scores. The text demands for college and career readiness expressed in Lexile measures gives students, parents and educators a clear, transparent and empirical benchmark upon which to evaluate their progress. The late Steven Covey stated that one of the core principles of effective leaders is to “begin with the end in mind.” This core tenet, when

it comes to college and career readiness for reading, has now been identified through the CCSS and expressed by grade level. As we transition to these higher and more rigorous demands, the Lexile Framework provides educators and policy makers with another tool in their arsenal, not only for achieving these goals, but for communicating progress. Along with the analogy of the higher hurdle of the CCSS, we can reassure the public that student achievement in reading is not getting worse, but that what constitutes our new standards has become more rigorous. Through the Lexile scale for reading we can show that student performance is growing even as we transition to the more rigorous standards. Ultimately the success of the CCSS may rest on how well we communicate student test performance in this transitional moment, which at first glance, may look worse than it actually is.

REFERENCES

Camara, W. J., & Shaw, E. J. (2012). The Media and Educational Testing: In Pursuit of the Truth or in Pursuit of a Good Story? *Educational Measurement: Issues and Practice, 31*, 33-37.

Goldstein, D. (2012). The Schoolmaster. *The Atlantic*.

National Governor's Association, Council of Chief State School Officers. (2012). *Common Core State Standards Initiative: Standard 10: Range, Quality, & Complexity.* Retrieved from Common Core State Standards Initiative: http://www.corestandards.org/ELA-Literacy

Sireci, S. G., & Forte, E. (2012). Informing in the Information Age: How to Communicate Measurement Concepts to Education Policy Makers. *Educational Measurement: Issues and Practice, 31*, 27-32.

Smith, M. (2004). The Need for Objective Measurement Under the No Child Left Behind Act. Durham, North Carolina: MetaMetrics, Inc.

Smith, M., Turner, J. (2013). Looking Worse Before We Look Better: Student Performance in a Common Core World. Durham, North Carolina: MetaMetrics, Inc.

Why P-16/P-20 Educational Systems?

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D.

In response to the challenges of preparing students to compete in a global marketplace, policy makers have started to conceptualize education more broadly than kindergarten through high school. As such, a "21st century" view that includes preschool through postsecondary education (P–16/P–20), increasingly is becoming the norm.

Challenges of Implementing P–16/P–20 Systems

The shift to a P–16/P–20 perspective has uncovered the difficulties that many students encounter when making the transition from secondary to post-secondary settings. A number of studies and reports released over the past few years indicate that the transition from high school to the post-secondary world is anything but smooth. According to the study, "Out of Many, One: Toward Rigorous Common Core Standards From the Ground Up" (July 2008), by the bipartisan, nonprofit organization Achieve, Inc., "too many students across the country meet state standards, pass state tests...only to be placed into remedial courses once they enroll in college or find they are unqualified for training programs and skilled employment in the modern workplace." In 2003, the U.S. Department of Education reported that remediation rates of entering freshmen were 42 percent for public two-year institutions, 20 percent for public four-year institutions, and 12 percent for private four-year institutions. Data from ACT and The College Board also have pointed to the inabilities of many students to meet college- and career-readiness benchmarks.

Bridging the Gaps

In order to have a seamless, well-connected P–16/P–20 system, there must be a consensus about what is considered to be "proficient" in high schools, community colleges, four-year colleges and other post-graduation pursuits. How can we address the discontinuity that exists today?

First, secondary and post-secondary environments must have a consistent way (i.e., common metric) to evaluate what "proficient" means in both worlds. Imagine the confusion that would result in health care if a patient (student) at 18 years of age transitioned (graduated) from pediatrics (high school) to a general internist (college, community college) and there were no consistent and standard metrics that spanned the developmental lifespan of the patient (student). Common indices of health (e.g., weight, temperature) are measured on the same scale whether one is 10 or 50 years old.

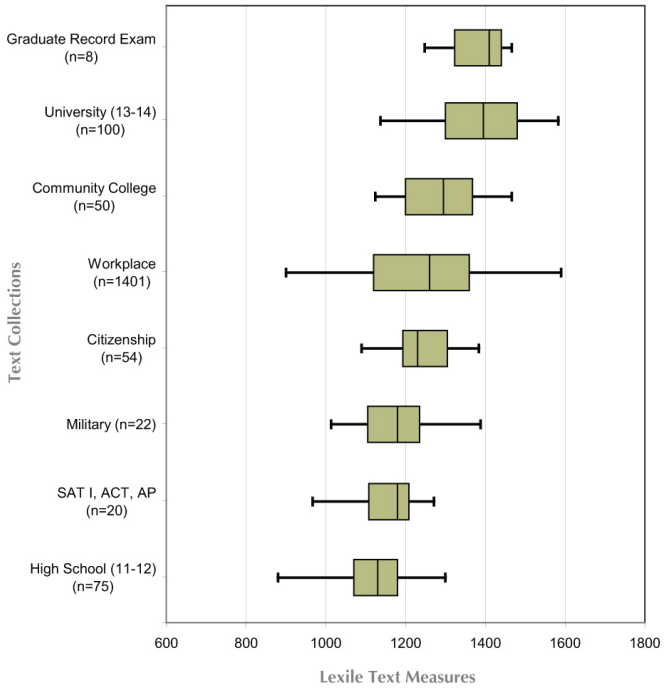
Consistency, clarity and continuity are just as critical in P–16/ P–20 education. For the measurement of reading, The Lexile Framework for Reading provides a common, developmental scale that allows one to examine the growth of reading proficiency over

the lifespan of an individual. Today, there are many instruments that measure and express student reading ability as a Lexile measure—from emergent reader (Dynamic Indicators of Basic Early Literacy Skills (DIBELS®) through adulthood (Test of Adult Basic Education (TABE)). For a list of instruments linked with the Lexile Framework, visit www.Lexile.com.

Measuring Reading Ability and Text Difficulty on a Common Scale

In addition to providing a consistent metric for reporting student reading ability across the lifespan of the individual, Lexile measures allow one to look at the reading demands of P–16/P–20 education in an objective, empirical fashion. Just as one can obtain a student's Lexile measure from a reading comprehension test that is linked with the Lexile scale, the difficulty of a particular text also can be measured and placed on the same Lexile scale. This allows educators to connect readers with appropriate text and forecast which students will likely need assistance with required readings. Today, more than 115,000 books across the P–16/P–20 landscape have been measured, as well as over 80 million articles and 60,000 Web sites. One does not have to guess about the reading demands of the texts within a particular environment or domain— they can be measured objectively.

Table 1: A Continuum of Text Difficulty for the Transition from High School to Postsecondary Experiences (Box Plot Percentiles: 5th, 25th, 75th, 95th)



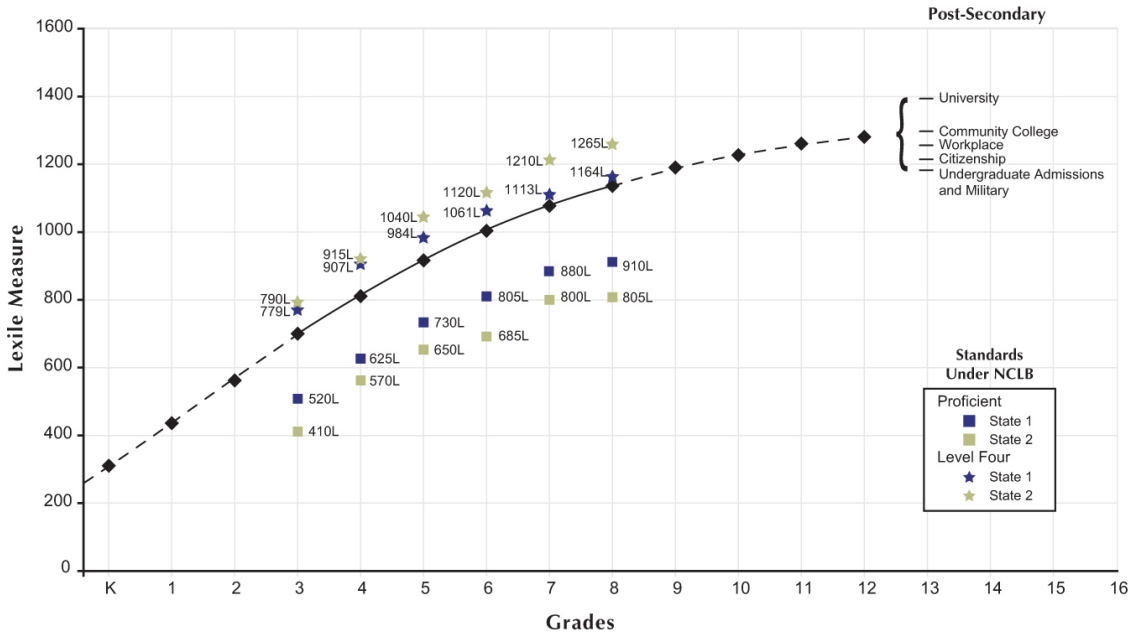
In 2004, Gary Williamson, Ph.D. conducted a study in which he analyzed the reading demands across a variety of post-secondary options, including the military, workplace and college. In terms of reading demands, the analysis illustrated in Table 1 clearly indicates a significant increase in the reading demands from high school to college. The postsecondary reading demand is substantially and significantly higher than the reading demand that the typical "proficient" high school student has encountered.

In addition, research conducted by the International Center for Leadership in Education shows that a significant gap exists between students' present reading levels and the reading requirements of the 16 Career Clusters defined by the U.S. Department of Education. Based on the results of a Lexile text analysis, the organization concluded that a large number of entry-level jobs have higher reading requirements than are required for high school graduation.

Better Policy Decisions and Better Results: An Objective Measure of College and Career Readiness

Because the Lexile Framework can be used to measure the text demands within various environments, as well as to measure the reading ability of students, policy makers now have better information to "align the journey with the destination." Table 2 shows the growth trajectory of approximately 70,000 North Carolina students who took the state's End-of-Grade Tests in grades 3–8 and extends that growth trajectory to their likely secondary school exit ability levels. When these projected ability levels are compared to the likely text demands that students will encounter in the postsecondary world, the misalignment is clear. This type of analysis can be done for each student, and allows educators and policy makers to make more informed observations about where a student is and where he or she is likely to exit. Stephen Covey, author of "The 7 Habits of Highly Effective Organizations," argues that a key principle of effective organizations is to "start with the end in mind." The Lexile Framework can help locate this end point and create a path to reach it so that educators and policy makers can better understand the growth demands over the entire P–16/P–20 span, and help prepare all students for success in college and their careers after high school graduation.

Table 2: Proficiency Standards Comparison with Median Postsecondary Text Measures



◆ Based on the estimated average growth for the 1999–2004 North Carolina Longitudinal Panel (Grades 3–8, N=67,908) $E(L) = 700.023 + 118.727\ (\text{Grade}-3) - 6.093\ (\text{Grade}-3)^2$ *Dashed lines indicate extrapolation or projection.*

REFERENCES

Achieve, Inc. (2006). Out of Many, One: Toward Rigorous Common Core Standards From the Ground Up. Washington, D.C.

Covey, Stephen R. *The 7 Habits of Highly Effective Organizations.* Simon & Schuster, 1998.

Daggett, Willard R. (2003). Achieving Reading Proficiency for All. A white paper from the International Center for Leadership in Education. Rexford, N.Y.

Parsad, Basmat and Lewis, Laurie. (2000). Remedial Education at Degree-Granting Postsecondary Institutions in Fall 2000. Statistical Analysis Report. National Center for Education Statistics. Washington, D.C.; Westat, Inc.; Rockville, Md.

Williamson, G. L. (2004). Student Readiness for Postsecondary Options. A white paper from MetaMetrics, Inc. Durham, N.C.

Williamson, G. L. (2006). Aligning the Journey With a Destination. A white paper from MetaMetrics, Inc. Durham, N.C.

SECTION 2: Looking to the Future

MetaMetrics White Paper | 1380L
Original Publication: October 10, 2012
Featured in the October 10th Issue of Education Week

Restoring Faith in Public Education

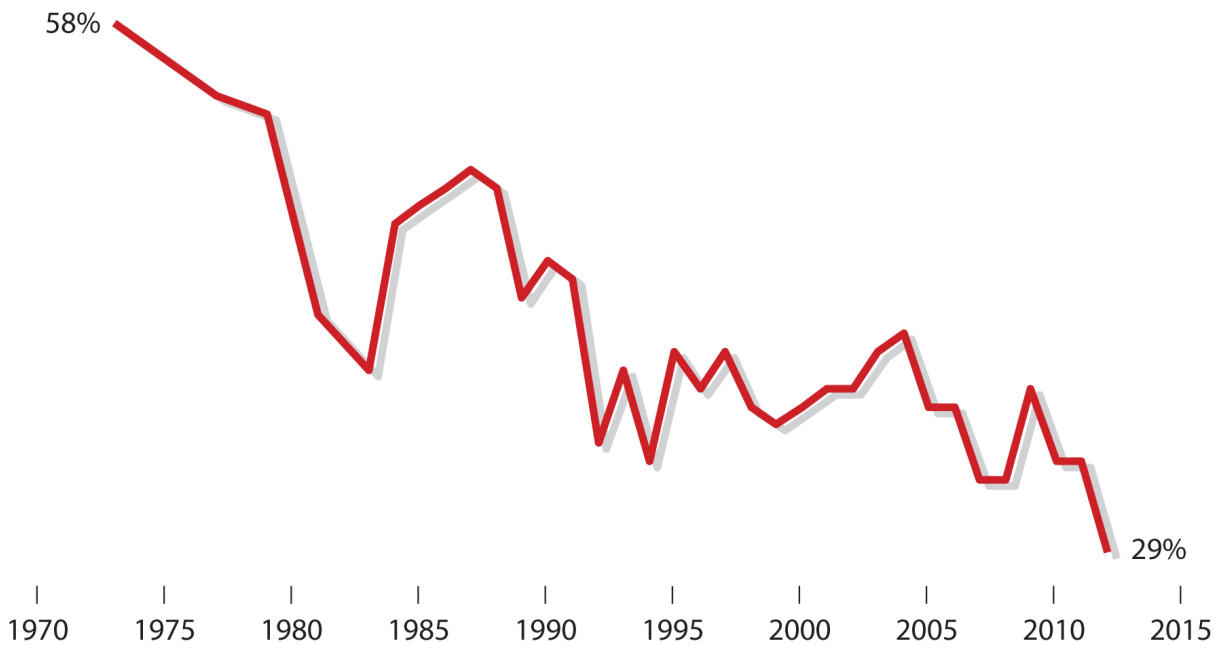
by MetaMetrics President and Co-founder Malbert Smith III, Ph.D., Director of Professional Development Jason Turner, and Research Engineer Steve Lattanzio

Restoring Faith in Public Education

Gallup's July 2012 'Confidence in Institutions' survey reveals a disheartening lack of confidence in U.S. public schools. While the majority of Americans continue to express confidence in institutions like the military and police, those same respondents expressed a much more dismal view of public education. Participants indicating 'a great deal' or 'quite a lot' of confidence in public K-12 education fell to an all-time low of around 29%— a 5% decrease from 2008 and a drop of 29 percentage points from 1973 when Gallup first began including public schools in its survey and public confidence was around 58% (Jones, 2012).

Unfortunately, faith in the public schools has been steadily eroding since 1973. By 1982, just 42% of respondents reported confidence in public schools; and while 1985 and 1988 saw a slight rebound in positive public perception, peaks and valleys notwithstanding, the trend has been clearly downward. As Figure 1 displays, this negative trend has been alarmingly consistent almost every year. For the first time ever recorded, less than 1/3 of the populace seems hopeful about the state of education (Jones, 2012).

Figure 1: Confidence in Public Schools

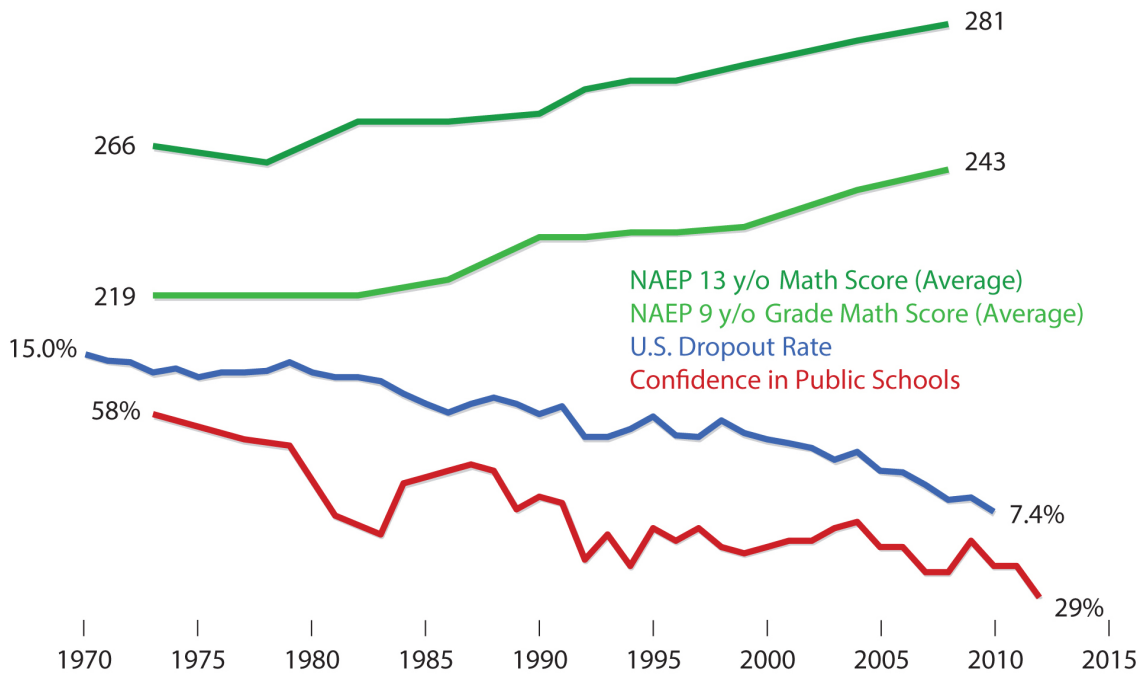


Public confidence: http://www.gallup.com/poll/155258/Confidence-Public-Schools-New-Low.aspx

Yet are things really this dismal? Is the public's perception of our educational system consistent with the reality of public education's performance? Is all of the student

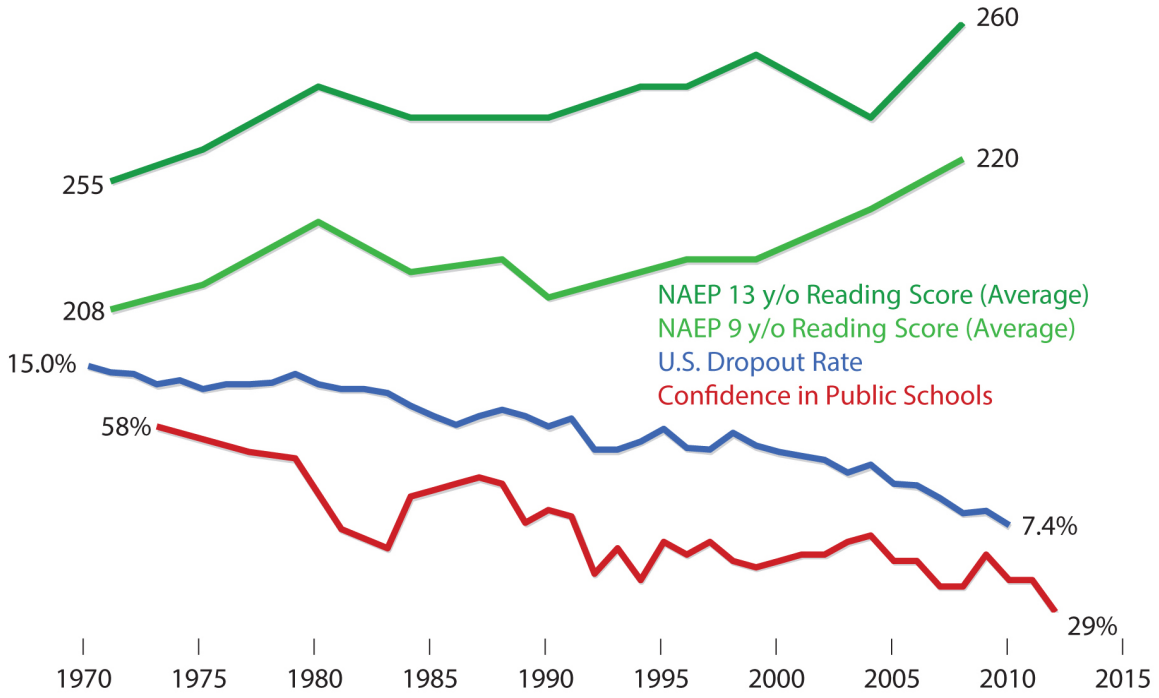
performance data negative? In fact, upon closer examination, the public's perception does not seem tightly tethered to the reality of the progress that has been made in public education over the last twenty years. In Figures 2 and 3, we mapped National Assessment of Educational Progress (NAEP) performance and high school dropout rates onto the public confidence graph. As Figures 2 and 3 illustrate, performance data on these dimensions is improving while public confidence is declining. In fact, NAEP scores for both 9 year old and 13 year old students have been steadily rising and trending upward since the 1970s. Though there were small dips throughout the 1980s, scores for 9 and 13 year old students rebounded in the 1990s and have been steadily rising ever since and now sit at an all-time high for both reading and mathematics. Compared to an average scale score of 219 in 1973 for 9 year old students, 2008's average scale score of 243 represents significant progress in math performance (see Figure 2) (National Center of Education Statistics, 2010).

Figure 2: Confidence in Public Schools vs. Student Performance



NAEP math scores: http://nces.ed.gov/programs/digest/d10/tables/dt10_140.asp
Dropout rate: http://nces.ed.gov/programs/digest/d11/tables/dt11_116.asp
Public confidence: http://www.gallup.com/poll/155258/Confidence-Public-Schools-New-Low.aspx

Figure 3: Confidence in Public Schools vs. Student Performance



NAEP reading scores: http://nces.ed.gov/programs/digest/d10/tables/dt10_124.asp
Dropout rate: http://nces.ed.gov/programs/digest/d11/tables/dt11_116.asp
Public confidence: http://www.gallup.com/poll/155258/Confidence-Public-Schools-New-Low.aspx

Graduation rates are worth a look too. The rate at which we graduate U.S. students—surely a key metric in assessing our educational system's overall success—has been rising. In 1999, the graduation rate was 66% and by 2009 the graduation rate had improved to 73.4% (Education Week, 2012). While the U.S. high school dropout rate remained stubbornly anchored in the double-digits for two decades, by 2005 it was reduced to less than 10%. As of 2010, the dropout rate decreased to 7.4% (National Center for Education Statistics, 2010). Considering that educators are being asked to educate more students than ever before—students with increasingly diverse backgrounds and levels of home support—the increase in high school graduation rates should be commended. While the U.S. admittedly still has more work to do to ensure that every student is graduating college and career ready, we have made remarkable progress over the last twenty years.

If public perception fails to mirror the reality of the commendable progress we have made in education over the last two decades, then what accounts for the erosion of confidence in our public schools? In light of clear positive performance metrics, is it an unarticulated distrust fueled by negative media coverage? Stories examining a wide

range of education topics—slashed budgets, students ill-prepared for the rigors of life after high schools, examples of ineptitude and bureaucratic waste—may reasonably shape negative public perceptions about the state of U.S. education. However, such explanations fail to account for the sheer depth and acceleration of the trend. Public confidence has been sliding down for almost forty years and seems to be getting worse despite some well-regarded empirical data demonstrating clear progress.

WHY THE NEGATIVE TREND?

One explanation for the diminishing confidence in public education is that Americans have become more cynical over the last thirty years and have expressed lessening confidence in most institutions. With the exception of the military, the police, and small business, which have notably remained above 50%, most public institutions have seen precipitous drops in public confidence. Banks, for example, once enjoyed confidence from a majority of the American public—60% in 1979. Today that number is 23%. Congress has always received low levels of approval. Even in 1973 only 42% of Americans expressed confidence in the job Congress was doing. Today that number sits at an abysmal 12%—by far the lowest level of confidence for any of the institutions included in Gallup's survey (Gallup News Service, 2012). This is not surprising. Elliot Gerson recently wrote, "According to a recent Rasmussen poll, only 17 percent of Americans believe our national government possesses the consent of the governed. These numbers may not seem shocking, because they've been low for so long" (Gerson, 2012).

Gerson's interpretation accords with Gallup's own summary of their most recent survey. As Lydia Saad writes, "The same poll found American's confidence in public schools, banks, and television news at their all-time lowest, perhaps reflecting a broader souring of American's confidence in societal institutions in 2012" (Saad, 2012). Our increasing cynicism toward some of our most fundamental public institutions may account for the lack of confidence Americans express in public education, *despite the fact that public education is improving as measured by certain essential metrics.*

An even more compelling reason for our loss of faith in public education—an explanation beyond the resigned cynicism of our populace—is the growing awareness of our place and role on the international stage. Over the last few decades, the U.S. has become a less insular, globally competitive, cosmopolitan society. While it is easier to commend our educational progress when considered only against the solipsistic backdrop of our own internal achievement, it is far more difficult to regard our progress with satisfaction when considered against our international peers. In fact, when seen through the prism of international educational progress, the U.S. appears to be making incremental gains in some areas and losing ground in others.

A review of the Trends in International Mathematics and Science Study (TIMSS) data reveals that the absolute score in mathematics for U.S. eighth grade students has risen only slightly (between 500 and 508) from 1995 through 2007. That fact becomes even more troubling when considered alongside our international ranking. In 1999, the U.S. ranked a disappointing 19th in mathematics out of 38 nations. In 2003, we ranked 15th and by 2007, we finally jumped to a promising 9th place among participating nations. That is good news. Still, the perception that the U.S. leads the world in education is hard to justify when considered against the wider backdrop of our international performance. While we have demonstrated clear progress in mathematics achievement, many other nations continue to consistently outperform our students. China, Singapore, and Hong Kong, for example, have ranked in the top five performing nations in mathematics from 1999 – 2007 (National Center for Education Statistics).

In the area of reading, the emerging trend is not encouraging. The Progress in International Reading Literacy Study (PIRLS) found that in 2001, U.S. fourth grade students had an average scale score of 542. By 2006 that score actually dropped a bit to 540 and our rank among participating nations dropped from 9th place in 2001 to 12th place in 2006 (see Figure 4) (National Center for Education Statistics). This reflects our international peers demonstrating clear progress and growing in achievement at a rate faster than the U.S.

Figure 4: U.S. Student Performance on the PIRLS and the TIMSS

	Year	Scale Score	Rank	Percentile
TIMSS	1995	500	28/41	33%
	1999	502	19/38	51%
	2003	504	15/45	68%
	2007	508	9/48	82%
PIRLS	2001	542	9/28	70%
	2006	540	12/28	59%

TIMSS math scores: http://nces.ed.gov/timss/table07_1.asp, http://nces.ed.gov/timss/TIMSS03Tables.asp?Quest=3&Figure=5, http://nces.ed.gov/timss/results99_1.asp, http://timss.bc.edu/timss1995i/HiLightB.html
PIRLS reading scores: http://nces.ed.gov/pubs2008/2008017.pdf

Growing awareness of our global ranking in education has taken a toll on public confidence and the view that we enjoy a dominant position on the world stage is no longer sustainable. As Gerson writes, "In college attendance, our previous preeminence has long faded; we are now 9th in percentage of younger workers with two-year or four-year degrees, and 12th in college graduation rate" (Gerson, 2012). These trends point to significant cause for concern: many U.S. students graduate unprepared for the challenges they will likely face in college and careers. This unpreparedness not only portends significant academic challenges, but increasingly dire consequences at both the individual and macro-economic levels. At the individual level, students may find themselves unable to compete academically and miss out on employment opportunities in some of the world's fastest growing career sectors.

At the macro level, below average academic performance suggests a troubling outlook for our country's competitiveness in the international arena. In response to the 2009 NCES report, Education Secretary Arne Duncan said, "We are lagging the rest of the world, and we are lagging it in pretty substantial ways. I think we have become complacent. We've sort of lost our way" (Holland, 2009). Duncan is right. A growing number of STEM (science, technology, engineering, and mathematics) doctoral students matriculate from outside the U.S. However they return to their native countries, leaving the U.S. with a shortage of qualified graduates in the STEM fields. The National Science Board's 'Science and Engineering Indicators: 2010' report stated that only 15.6 percent of bachelor's degrees were awarded in STEM fields (Business Higher Education Forum).

Given that by a number of important benchmarks we are—to paraphrase Secretary Duncan—lagging the rest of the world in some significant ways, it is no surprise that many Americans view the effectiveness of K-12 public education with increasing skepticism. While public education has shown green shoots of educational progress, these budding indicators fail to garner the attention of the press as compared to the disappointing news of our slippage in international rankings. It is difficult to recall a national news story or national press release celebrating any of the signs of positive progress in K-12 education. Recent developments, however, provide a reason for hope and support a more optimistic outlook.

REVERSING THE NEGATIVE TREND

We are cautiously optimistic that the public's confidence in public education will begin to rise as we take a number of significant steps. A major step in that direction has already been taken with the adoption of the Common Core State Standards. The Common Core State Standards were fueled by the recognition that our country needs to

adhere to a set of clearly articulated standards of sufficient rigor to ensure that all students graduate college and career ready. The Standards establish a clear pathway for students and set clear markers for what it means to be college and career ready. This is a good first step.

An essential function of legitimate educational systems around the world is readying students for advanced studies and careers and for life after secondary school. There is a tight correlation between a country's education level and its economic productivity, which is why retaining our nation's competitiveness among our international peers requires serious educational reform and a more rigorous and standardized curriculum, one that establishes readiness for the postsecondary world as the goalpost. An OECD study predicted that an increase of 25 points on the PISA (Programme for International Student Assessment) over the next 20 years would result in an economic gain of $41 trillion for the U.S. economy (Armario, 2010).

That sort of commitment to educational reform—like the sort advocated by the authors of the Common Core State Standards—will require a heady mix of social capital and political will. The United States' place on the global stage is at stake. To remain economically and politically relevant, it is essential that we remain globally competitive in our educational outcomes. Our high school students must graduate ready for the demands of an increasingly competitive global workforce. The shift away from proficiency—where each state set their own standards, standards divorced from the rest of the nation and untethered to the demands of the global economy—toward a common set of shared curriculum standards represents a move away from an insular and self-referential standard of academic success toward a more rigorous and global set of standards. The new standards are designed for the global stage and are crafted to allow us to regain and retain our place of global competitiveness.

The second critical step in restoring confidence in education is the implementation of the Common Core. The success of the Common Core hinges upon how well we as educators transition from adoption of the new standards to actual implementation. This step is critical and small incremental changes will not be sufficient. Implementation will require rigorous and dedicated focus throughout all of education. The importance of this step has been recently described in 'Transitioning from Adoption to Implementation of the Common Core State Standards' (Smith, 2012).

A third step in the process of renewing public confidence must be taken by the leaders of our political parties. As Matt Miller recently argued, both parties have failed to provide significant time, leadership, and political will toward education (Miller, 2012). Miller reported that in the Republican debates, less than 1% of time has been devoted to education thus far. Democrats have proven equally guilty, devoting just as little time as

their colleagues on the other side of the aisle. As Miller pointed out, President Obama devoted only two minutes to education in the recent State of the Union address—a speech lasting more than an hour (Miller, 2012).

Our leaders on both sides of the political aisle need to do more substantive work in education than merely complain about our lack of progress. In fact, our leaders would do well to point to some of our successes along the way. It is worth remembering that while we may trail our international peers in STEM related fields, many countries continue to send their students to our shores for our cultural significance and to study American ingenuity and creativity. However almost all players along the educational food chain—from the classroom teacher, the principal, the local superintendent, and ultimately the state chief—feels assailed on all sides. The accumulation of negative news reports and the labeling of teachers and schools as 'failures' do little to provide a conducive environment for productive change.

A fourth step in the process of restoring confidence in public education rests with our educators. An almost universal trait of educators is optimism. Educators tend toward their profession because of a deep-seated belief that all children can learn, that children can have a better tomorrow, and because education makes a difference in individual lives and to society as a whole. It is critical that our performance matches our beliefs. Educators must be realistic about the need to improve our educational system while simultaneously helping the public celebrate the successes and improvements we have made and will continue to make. The old adage that one can choose to see the glass as half empty or half full is undeniably applicable here. Unfortunately, the public overwhelmingly has come to see the education glass as half empty. While we should continue to have candid conversations about the areas we need to improve, let us also not hesitate to point to our many successes as these green shoots take root.

REFERENCES

Armario, C. (2010, December 7). Wake-up call: US students trail global leaders. Retrieved from MSNBC: http://www.msnbc.msn.com/id/40544897/ns/us_news-life/t/wake-up-call-us-students-trail-global-leaders/

Business Higher Education Forum. (n.d.). Confronting the STEM Challenge: A New Modeling Tool for US Education Policymakers. Washington D.C.: Business Higher Education Forum.

Education Week. (2012). National Graduation Rate Keeps Climbing; 1.1 Million Students Still Fail to Earn Diplomas. Retrieved from EdWeek.org: http://www.edweek.org/mediadiplomascount2012_presspacket_final.pdf

Gallup News Service. (2012). Confidence in U.S. Public Schools at New Low. Gallup Poll Social Series. Gallup New Service.

Gerson, E. (2012, July 10). To Make America Great Again, We Need to Leave the Country. The Atlantic.

Holland, S. (2009, August 25). U.S. Students Behind in Math, Science, Analysis Says. Retrieved from CNN: http://articles.cnn.com/2009-08-25/us/students.science.math_1_math-and-science-fourth-and-eighth-graders-math-scores?_s=PM:US

Jones, J. M. (2012, June 20). Gallup Politics. Retrieved from Gallup.com: http://www.gallup.com/poll/155258/Confidence-Public-Schools-New-Low.aspx

Miller, M. (2012, April). The Third Party Education Agenda We Need. Retrieved from mattmilleronline: http://www.mattmilleronline.com/pdfs/third-party-education-speech-2012.pdf

National Center for Education Statistics, National Assessment of Educational Progress (NAEP), NAEP 2004 Trends in Academic Progress; and 2008 NAEP Long-Term Trend Mathematics Assessment, retrieved from NCES.ed.gov: http://nces.ed.gov/programs/digest/d10/tables/dt10_140.asp

National Center for Education Statistics. (2010). National Center of Education Statistics. Retrieved from NCES.ed.gov: http://nces.ed.gov/fastfacts/display.asp?id=16

National Center of Education Statistics. (2010). National Center for Education Statistics. Retrieved from NCES.ed.gov: http://nces.ed.gov/programs/digest/d10/tables/dt10_124.asp

National Center for Education Statistics. (2010). National Center of Education Statistics. Retrieved from NCES.ed.gov: http://nces.ed.gov/pubs2008/2008017.pdf

National Center for Education Statistics. (n.d.). Progress in International Reading Literacy Study. Retrieved August 2012, from National Center for Education Statistics: http://nces.ed.gov/surveys/pirls/

National Center for Education Statistics. (2009). The Condition of Education: A Close Look. National Center for Education Statistics.

National Center for Education Statistics. (n.d.). Trends in International Mathematics and Science Study (TIMSS). Retrieved August 2012, from National Center for Education Statistics: http://nces.ed.gov/timss/

Saad, L. (2012, July 12). US Confidence in Organized Religion at Low Point. Retrieved from Gallup Politics: http://www.gallup.com/poll/155690/Confidence-Organized-Religion-Low-Point.aspx

Smith, M. (2012, July 19). Transitioning from Adoption to Implementation of the Common Core State Standards. Retrieved from Lexile.com/about-lexile/Policy-Briefs/.

U.S. Department of Commerce, Census Bureau, Current Population Survey (CPS), October 1967 through October 2010. Retrieved from NCES.ed.gov: http://nces.ed.gov/programs/digest/d11/tables/dt11_116.asp

Miles to Go Before We Sleep: The Successes of U.S. Education and the Promises Left to Keep

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D., Director of Professional Development Jason Turner, and Research Engineer Steve Lattanzio

"The woods are lovely, dark and deep.
But I have promises to keep,
And miles to go before I sleep,
And miles to go before I sleep."

-Robert Frost, "Stopping by Woods on a Snowy Evening"

Miles to Go Before We Sleep

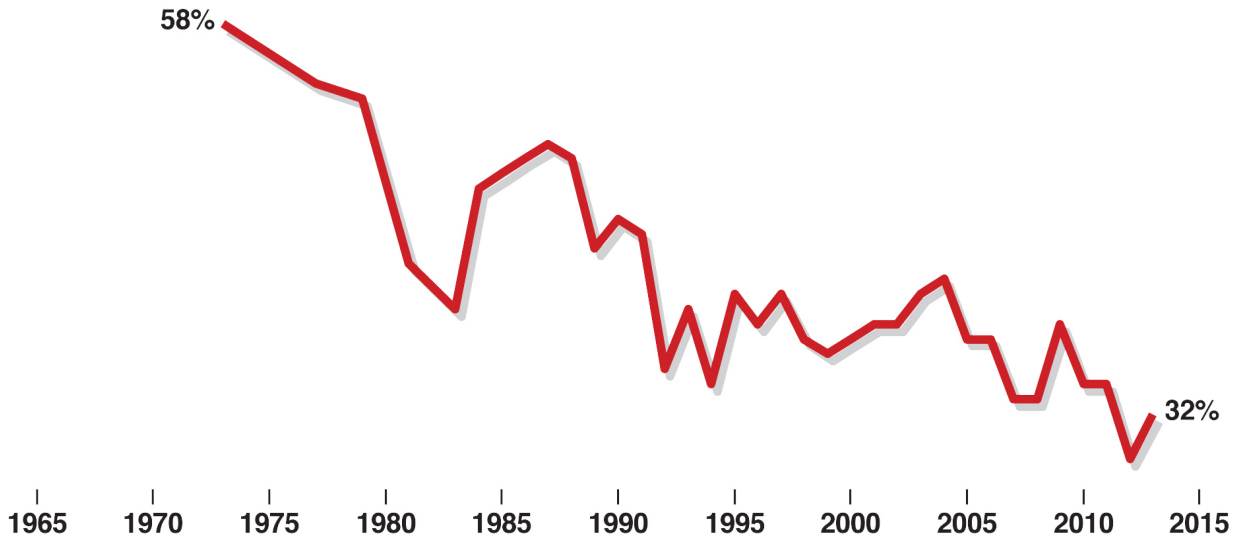
Since 1973, Gallup has conducted an annual public confidence survey in which Americans rate their confidence in sixteen various public institutions. Last year's results generated the headline, "Confidence in U.S. Public Schools at New Low" (Jones, 2012). Puzzled and concerned by this trend, we examined empirical performance measures of U.S. public schools to see if public perceptions were, in fact, tethered to reality. In our paper, 'Restoring Faith in Public Education' (Smith, Turner & Lattanzio, 2012), we plotted National Assessment of Educational Progress (NAEP) Long Term scores in reading and mathematics, Trends in International Math and Science Study (TIMSS) scores, Progress in International Reading Literacy Study (PIRLS) scores, and high school drop-out rates against the plummeting public school confidence trend line. Our analysis indicated that such a dismal perception was not warranted when considered against these empirical benchmarks.

On June 13, 2013, Gallup reported this year's survey results and fortunately, education was not the headline story. The big story generated from this year's survey was that public confidence in Congress had reached an all-time low. Just 10 percent of respondents reported having confidence in Congress (Mendes & Wilke, 2013).

Confidence in public schools, on the other hand, experienced a slight uptick with 32 percent of respondents reporting confidence in public education (Mendes & Wilke, 2013). That's up 3 percentage points over last year's poll results. While that may seem like an encouraging sign, it's worth noting that the poll's margin of error is +/- 3

percent[3]. This means, for all intents and purposes, public confidence in education remains essentially unchanged.

Figure 1: Public School Confidence



Public School Confidence: http://www.gallup.com/poll/163052/americans-confidence-congress-falls-lowest-record.aspx

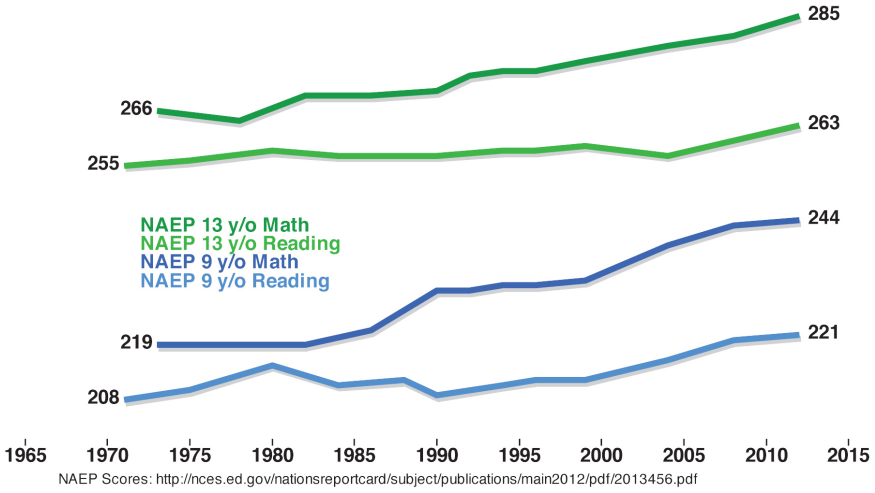
As Figure 1 illustrates, our confidence in public schools has been progressively declining since the 1970s. Other than a few brief rebounding peaks in the mid-80s, the American public has adopted a primarily pessimistic view on the state of education in the U.S. As we argued in last year's paper, this declining confidence in education does not appear to be justified (Smith, Turner & Lattanzio, 2012). In addition to another year of Gallup data, we also have additional data on student performance. In July of 2013, the National Center for Education Statistics (NCES) released NAEP Long Term results for 2012. On the international assessment side, we have 2011 test data for PIRLS and

[3] Concerning polling results for confidence in various American institutions, Gallup states, "For results based on the total sample of national adults, one can say with 95 percent confidence that the maximum margin of sampling error is ±3 percentage points." This does not mean that every value as a percent in the poll has a 95 percent confidence of ±3 percentage points, but rather, that no value in the poll has a 95 percent confidence interval greater than ±3 percentage points. Given the total sample of 1,529 national adults in 2013, the 32 percent of respondents who said they had a "great deal/quite a lot" of confidence in public schools has a 95 percent confidence interval of ±2.3 percentage points. Given the total sample of 1,004 national adults in 2012, the 29 percent of respondents who said they had a "great deal/quite a lot" of confidence in public schools has a 95 percent confidence interval of ±2.8 percentage points. When taken together, there is approximately a 95 percent probability that the true percentage of Americans with a "great deal/quite a lot" of confidence in public schools is higher in 2013 than it was in 2012, which borders on statistical significance. The fact that the probability is 95 percent and the difference in value between the two years is 3 percent is coincidental.

TIMSS. We have also added an additional variable worth considering—trend data on the percentage of free and reduced lunches being served in public schools from 1969 to 2012.

So with another year of Gallup data, another year of NAEP Long Term results, PIRLS data, TIMSS results, drop-our rates, and hard data on the increase of free and reduced lunches being served, is perception tied to reality? The short answer is the same as last year – we think not. After all, U.S. students have continued to demonstrate positive growth in reading and mathematics as measured by the NAEP[4]. As Figure 2 illustrates, 9 and 13-year-old U.S. students have markedly improved in both reading and mathematics since 1970 (National Center for Education Statistics, 2012). In fact, in reading, the average score of a U.S. 9-year-old student has risen from 208 in 1970 to 221 in 2012. In mathematics, the growth is even more dramatic with the average score of a 9-year-old student rising from 219 in 1970 to 244 in 2012. For thirteen year old students, the trend has been similarly positive. In reading, thirteen year old students have risen from 255 to 263; and in mathematics they have grown from an average of 266 in 1970 to 285 in 2012 (National Center for Education Statistics, 2012). This marked increase in mathematics achievement demonstrates a remarkable growth trend among American students and belies the pessimistic view adopted by the public on the state of public education.

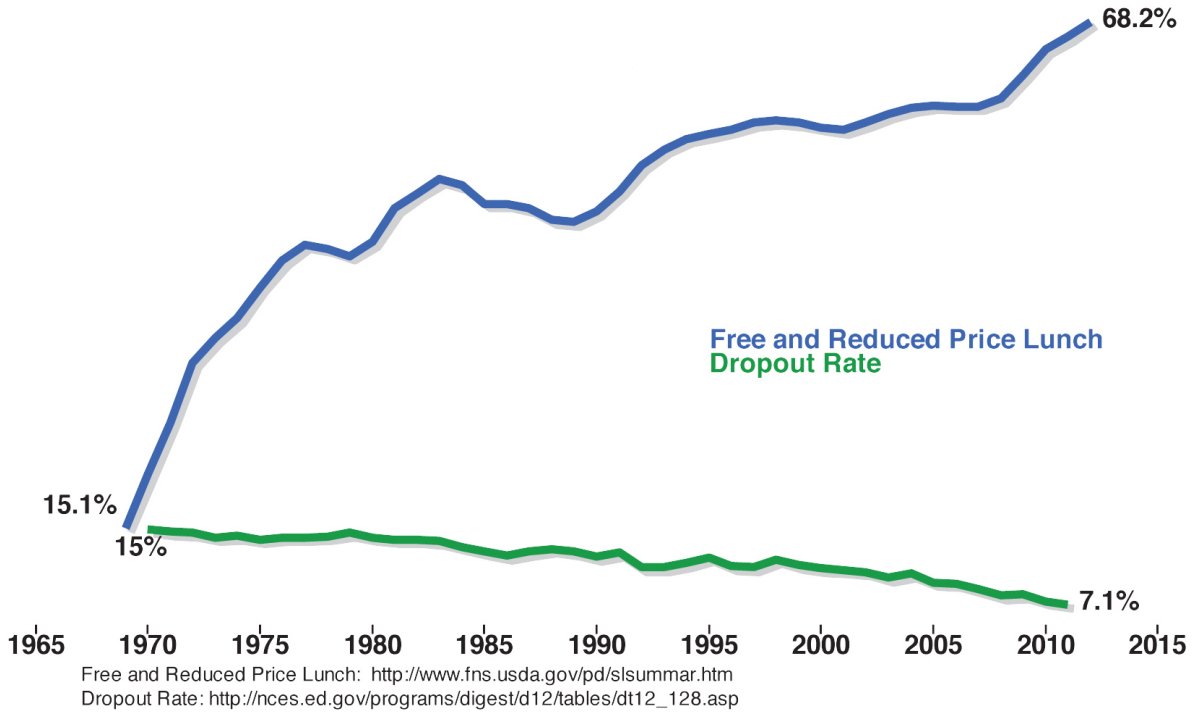
Figure 2: NAEP Math and Reading Scores



[4] The average NAEP scores for all students are used instead of the average NAEP scores for just public school students because data on public vs. private or Catholic school data does not extend past 1978 for mathematics or 1980 for reading. The average scores for public schools tell the same story as the NAEP scores that are shown for the available years. Students that take the NAEP are overwhelmingly public school students and over the years the percentage of total students in public schools has actually increased, which should alleviate any concerns of this positive trend being driven by non-public schools.

While it's worth noting that reading and mathematics scores for 17-year-old students have remained virtually unchanged since the early '70s (e.g., 285 in reading in 1973 and 287 in reading in 2012) (National Center for Education Statistics, 2012), some have argued that there's nothing particularly worrisome or illustrative about that trend. Seniors in high school are well aware that the assessment has no effect on their academic ambitions, and it makes little difference whether they do well or poorly. In fact, as Diane Ravitch has written, "The National Assessment Governing Board (NAGB), which oversees NAEP, has known for years that 12th graders don't try to do well on the tests" (Ravitch, 2010). And we believe that this same concern applies equally to 17-year-old students who participate in the NAEP Long Term study. While the concern about motivation and good faith efforts may cast doubt upon the validity of seventeen year old performance, the fact that the trend has been basically flat is actually remarkable when one considers that 17-year-old students today represent a far more diverse population than in the 1970s.

Figure 3: Free/Reduced Price Lunch and Dropout Rate



All of this commendable progress has been made despite the fact that today's educators are working with larger percentages of low-income students and much higher

percentages of English Language Learners (ELL). As of 2009, around 10 percent of U.S. students were ELL and 80 percent of this group were native Spanish speakers (Center on Education Policy, 2012). Student populations of low-income students have been rising as well. Figure 3 illustrates that from 1969 through 2012, there has been a steady upward trend in the percentage of students receiving free and reduced lunch[5]. In 1969, only 15 percent of U.S. students (of those who received school lunches) were free and reduced price recipients. That number, however, has been increasing ever since, and by last year, that number was over 68 percent (USDA: Food and Nutrition Service, 2013).

Table 4: U.S. Student Performance on the PIRLS and the TIMSS

	Year	Scale Score	Rank	Percentile
TIMSS (8th grade)	1995	500	28/41	33%
	1999	502	19/38	51%
	2003	504	15/45	68%
	2007	508	9/48	82%
	2011	509	9/42	80%
PIRLS (4th grade)	2001	542	9/35	76%
	2006	540	18/45	61%
	2011	556	6/45	88%

TIMSS: http://nces.ed.gov/timss/; http://timss.bc.edu/timss1995i/HiLightB.html (1995)
PIRLS: http://nces.ed.gov/surveys/pirls/index.asp

In the face of such challenges, the U.S. has continued to make gains. Even more compelling is U.S. student performance when compared to other developed nations, as measured by the PIRLS and TIMSS. As Table 4 illustrates, the United States has continued to improve in terms of both the average student scale score and the rank of

[5] Free and reduced price lunches served as a percentage of total school lunches is used as a proxy for percent of students eligible for free and reduced price lunches because there is data that extends back to 1969, vs. the statistics on percent eligible for free and reduced price lunch which extends only back to 1998 (see http://nces.ed.gov/ccd/elsi/tableGenerator.aspx). However, the trend for percent students eligible for free and reduced price lunches between 1998 and 2010 matches almost exactly with the percentage of free and reduced price lunches served. Both increase by 8 percent between 1998 and 2010 and have a correlation of 0.95. Thus, we feel that the metric is justifiable as a proxy for student poverty in public schools.

the U.S. relative to other nations[6]. On the TIMSS, student performance has risen from a score of 500 in 1995 to 509 in 2011 (National Center for Education Statistics, 2012). And the increase in rank has been even more dramatic – from 33 percent in 1995 to 80 percent in 2011 (National Center for Education Statistics, 2012). As Reardon, Valentino and Shores have acknowledged regarding student performance on the PIRLS:

> "Once again, there is little evidence of an imbalance. In 2006 the United States scored above average in both reading for literary purpose and for informational purpose, ranking twelfth in both categories" (Reardon, Valentino, & Shores, 2012).

All of these results point to commendable progress in public education and there is much to laud about the work our educators are doing. Public cynicism and decreasing confidence appear not entirely justified given that the last thirty years have seen improvement in U.S. student performance. Possible reasons for the deepening cynicism of the American public are outside the scope of this paper, though we have argued for several possible explanations in our previous paper, 'Restoring Faith in Public Education' (Smith, Turner, & Lattanzio, 2012).

That the academic performance of U.S. students is generally improving is good news, and it should not be given short shrift, however; there is some cause for concern. Despite these notable and admirable successes, digging a little deeper in the data reveals there is still much work to be done to shore up the dream of educational opportunity for all U.S. students. Two gaps that continue to plague educators – and perhaps contribute to the public's declining confidence in schools – are the achievement gap and the college and career readiness gap. As Eric Robelen argues, though there has been general improvement in U.S. student performance, the academic black-white achievement gap remains an intractable thorn in the side of our educational system (Robelen, 2013). Though the persistent gap has been well documented and the target of a multitude of educational policies, a magic bullet solution has remained elusive, and the gap persists despite our well-intentioned efforts.

As Eric Robelen recently wrote:

> A new report from "the nation's report card: emphasizes progress in closing achievement gaps for black and Hispanic students between the early and mid-1970s and today....

---

[6] It should be noted that the composition of countries participating in TIMSS and PIRLS changes from year to year and the scale score is relative to the countries participating; thus the scale scores and percentiles need to be viewed cautiously. However, the absolute ranking of the United States among the countries can be viewed more confidently as a metric for U.S. performance as most of the countries that do not participate in all of the years are towards the bottom end of the scale while the top of the scale has more consistent participation.

> Much of that narrowing of the achievement gap was actually accomplished by the mid to late 1980s, the data indicate. It's ebbed and flowed a bit since then, but in most cases, the gaps are no smaller today than they were two decades ago. In fact, they're sometimes larger, though not by amounts deemed statistically significant.
>
> For example, the black-white achievement gap for 13-year-olds in reading reached its narrowest point in 1988, at 18 points, compared with 23 points in 2012.
>
> In math, the black-white achievement gap for 9-year-olds was 25 points in 1986, the exact same figure as for 2012 (Robelen, 2013).

To be fair, there has been *some* progress in ameliorating the black-white achievement gap, particularly when taking a wider historical view and considering the size of the gap in the early 1970s. Comprehensive educational policies – like No Child Left Behind (NCLB) and more recent efforts like Race To The Top – have attempted to explicitly address the gap, and have had some mild successes. As Reardon, Valentino, and Shores have argued, there have been some positive steps in narrowing the gap, however modest:

> The gap widened modestly in the early 1990s before beginning to narrow again in the late 1990s; that narrowing continued slowly through 2008. This pattern is evident in Scholastic Achievement Test score trends as well as in other large studies with nationally representative samples of students. The most recent NAEP-LTT data (from 2008) indicate that the black-white gap is now roughly 0.6 of a standard deviation, about half of what it was forty years ago, although almost all of the progress in closing the gap was made in the 1970s and 1980s (Reardon, Valentino & Shores, 2012).

Given that Black and Hispanic students enter high school with literacy skills approximately three years behind their white peers, it's easy to see how such a gap could persist (Reardon, Valentino, & Shores, 2012). And given the size of the gap, only the most intensive remediation efforts are likely to have a significant impact. Recent research, however, indicates that the achievement gap may, in fact, be more a function of income than race or ethnicity. Low-income 8th graders are approximately five years behind their more affluent peers, and there appears to be some evidence that the socioeconomic gap is widening:

> ECLS-K data indicate that socioeconomic disparities in reading achievement are much larger than racial or ethnic gaps. ...These socioeconomic achievement gaps appear to have widened substantially in recent decades. ...For children born in the 1950s, the reading gap between students from high- and low-income families was

> smaller than the black-white gap; the income gap is now much larger than the black-white gap (Reardon, Valentino, & Shores, 2012).

That's not surprising considering the opportunity gap between the socioeconomic levels. Children from high-income families often benefit from a wide array of additional educational efforts, including tutors and supplemental educational tools and resources (Reardon, Valentino, & Shores, 2012). Additionally, high-income families are more apt to spend a greater amount on enrichment activities (e.g., summer camps, travel, language immersion programs, extracurricular academic activities, etc).

> In the early 1970s families in the top income quintile invested 4.2 times more a year in child enrichment expenditures than did parents in the lowest income quintile; by 2005 parents in the highest income quintile spent 6.8 times more a year on child enrichment activities than did their counterparts in the lowest income quintile (Reardon, Valentino, & Shores, 2012).

Matthew O'Brien argues similarly that the U.S. is becoming more stratified and that the well-to-do are pouring a tremendous amount of resources into the education and additional enrichment activities of their children (O'Brien, 2013). O'Brien points to a recent study by the Pew Economic Mobility Project that found, strikingly, that high socioeconomic students without a college degree are 2.5 times more likely to end up affluent than low socioeconomic students with a college degree.

> It's what outgoing Council of Economic Advisers chief Alan Krueger has dubbed the "Great Gatsby Curve" – the more inequality there is, the less mobility there is. As Tim Noah put it, it's harder to climb our social ladder when the rungs are farther apart (O'Brien, 2013).

According to NCES, almost two-thirds of Black and Latino students attend schools in which more than half of the student population is from low-income families (Center on Education Policy, 2012). In addition, almost 20 percent of U.S. students are from families with incomes below the federal poverty line (Center on Education Policy, 2012). The burgeoning opportunity gap may go a long way in accounting for the persistent achievement gap, an ostensible black-white gap, but more accurately recognized as the gap between those with abundant resources and those without.

The second gap that has received increasing national attention is our failure to graduate every student prepared for the rigors of college and career. This 'college and career gap', and our collective resolve to address it, led to the creation of the Common Core State Standards. With the almost unanimous adoption and implementation of these standards, there has been unparalleled national, regional and local media attention on this historic moment in education. As a small example, Exxon Mobil launched a

concerted media campaign entitled 'Let's Solve This', directed at tackling major educational deficiencies (Exxon Mobil, 2013). 'Let's Solve This' uses media such as national TV commercials and social media to highlight, target and address these educational issues.

A consequence of all the media attention on the Common Core State Standards is the highlighting of our failures in public education. Principally, our failures revolve around these two attainment gaps. In none of the media campaigns have the positive results discussed above or our significant progress as a nation been chronicled. Balanced perspective is certainly needed if we are going to simultaneously raise our standards and public confidence.

Balancing the dual priorities of equity and excellence represents a significant challenge. This challenge is best expressed in Secretary of Education Arne Duncan's assertion that the Holy Grail of education is to graduate all students from high school, and to secondly ensure that all students graduate college and career ready (Education Week, 2009). Never in the history of education in our country have we graduated 100 percent of our students. While steadily improving our overall graduation rate, our drop-out rate remains stubbornly high; we still lose approximately 1.3 million students per year (Alliance for Excellent Education, 2013). Even worse, the data on the number of high school graduates requiring remedial educational support once they enroll in a community college or university demonstrates a significant gap. According to the Alliance for Excellent Education, 42 percent of freshmen at community colleges – and 20 percent of freshmen at four-year institutions – enroll in at least one remedial course (Alliance for Excellent Education, 2006).

Seen through this lens, the U.S. has a long way to go in terms of public education. The long road ahead, and the notable work that must be done, may justifiably give rise to the skepticism and doubt expressed by the American public in Gallup's recent poll. But when President Kennedy announced in 1961 our intent to put a man on the moon, it likely sounded as far-fetched and unlikely as Secretary Duncan's vision that we graduate every student college and career ready. In the same way that President Kennedy's goal and commitment became a reality in 1969, we believe that Duncan's views can be realized in the years to come.

The distance between the goal of graduating every student college and career ready and the status of today's U.S. student population is obviously significant. A burgeoning ELL student population and a significant number of low-income students mean that many students begin school far behind their more affluent peers. Unfortunately, educational research over the last hundred years has demonstrated that there are no easy solutions and no silver bullets to solve some of education's most intractable problems. Realizing

the Holy Grail will require sustained effort, tireless commitment, and dedicated execution – as well as a public committed to facing the myriad – and often uncomfortable – issues that complicate our educational landscape; it will require confidence in our ability to reach our long term goal of achieving educational opportunity for all of our students. As we pointed out in our paper, 'Restoring Faith in Public Education', our challenge will be to educate the public in such a way as to renew their confidence by recognizing the gains we are making and being realistic about the challenges we face (Smith, Turner & Lattanzio, 2012). Have we made significant strides in improving the quality of education for K-12 students? Looking back at the overall data trends, the answer is clearly yes. Looking forward, however, we still have significant steps left in this journey, and remedying the achievement gap and the college and career readiness gap may be our last two mile markers.

REFERENCES

National Center for Education Statistics. (2012). Trends in Academic Progress. Washington, D.C.: Institute of Education Sciences. Retrieved from: http://nces.ed.gov/nationsreportcard/subject/publications/main2012/pdf/2013456.pdf

Alliance for Excellent Education. (2006). Paying Double: Inadequate High Schools and Community College Remediation. Washington, D.C.: Alliance for Excellent Education.

Alliance for Excellent Education. (2013). Graduation Rates. Retrieved from Alliance for Excellent Education: http://www.all4ed.org/about_the_crisis/students/grad_rates

Center on Education Policy. (2012). A Public Education Primer: Basic (and Sometimes Surprising) Facts About the U.S. Educational System. Washington, D.C.: Center on Education Policy.

Education Week. (2009, December 02). An Interview with Arne Duncan. Retrieved from Education Week: http://www.edweek.org/ew/articles/2009/12/02/14duncan-transcript.h29.html

Exxon Mobil. (2013). Let's Solve This. Retrieved from let's solve this: http://letssolvethis.com/

Jones, J. M. (2012, June 20). Gallup Politics. Retrieved from Gallup.com: http://www.gallup.com/poll/155258/confidence-public-schools-new-low.aspx

Mendes, E., & Wilke, J. (2013, June 13). Gallup Politics. Retrieved from Gallup.com: http://www.gallup.com/poll/163052/americans-confidence-congress-falls-lowest-record.aspx

National Center for Education Statistics. (2012). TIMSS 2011 Results. Washington, D.C.: Institute of Education Sciences.

National Center of Education Statistics. (n.d.). Digest of Education Statistics:. Retrieved from nces. ed.gov: http://nces.ed.gov/programs/digest/d12/tables/dt12_128.asp

O'Brien, M. (2013, June 18). RIP, American Dream? Why It's So Hard for the Poor to Get Ahead Today. Retrieved from The Atlantic.com: http://www.theatlantic.com/business/archive/2013/06/rip-american-dream-why-its-so-hard-for-the-poor-to-get-ahead-today/276943/

Ravitch, D. (2010, November 23). Bridging Differences. Retrieved from Education Week: http://blogs.edweek.org/edweek/Bridging-Differences/2010/11/12th_grade_naep_scores_are_mea.html

Reardon, S. F., Valentino, R. A., & Shores, K. A. (2012, Fall). Patterns of Literacy among U.S. Students. Retrieved from futureofchildren.org: http://futureofchildren.org/futureofchildren/publications/docs/22_02_02.pdf

Robelen, E. (2013, June 28). NAEP Report: A Closer Look at Trends in the Achievement Gap. Retrieved from Education Week: http://blogs.edweek.org/edweek/curriculum/2013/06/naep_report_a_closer_look_at_the_achievement_gap.html

Smith, M., Turner, J., & Lattanzio, S. (2012, November 05). Restoring Faith in Public Education. Retrieved from Lexile.com: https://d1jt5u2s0h3gkt.cloudfront.net/m/cms_page_media/135/Public%20Confidence_3.pdf

USDA: Food and Nutrition Service. (2013). Retrieved from USDA.gov: http://www.fns.usda.gov/pd/slsummar.htm

MetaMetrics White Paper | 1350L
June 15, 2012

Four Global Trends in Education (And Why They Matter)

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D., Executive Director of Global Services Todd Sandvik, and Research Engineer Steve Lattanzio

Preface

During the past year, we have traveled the world to meet with leaders in educational assessment, technology, publishing, and research. As they have described the challenges facing education within their countries and organizations—and the strategies for confronting them—four common trends have emerged. Generally speaking, they represent advancements in thinking that can be traced to an increasingly global orientation and growing digital capabilities. Each trend represents real opportunities to improve learning and better meet the needs of students, parents, and educators. Our work with the Lexile Framework for Reading offers examples of benefits being achieved today in support of these trends.

#1- University and Career Readiness

The first concern of most education systems around the world is the level of effectiveness in readying students for advanced studies and jobs. The tight correlation between a country's education level and its economic productivity holds the attention of policy makers everywhere. It is widely held that individuals and communities with higher education levels have more and better opportunities in the long run. However, for a country to stay competitive among our globally connected economies, it must stay globally competitive in its educational outcomes. This means students must graduate high school ready for the demands of increasingly competitive universities and workplaces. In other words, they must be university (or "college") and career ready. As a penultimate goal of K-12 education, this issue is at the forefront of current policy and research. Arne Duncan, the U.S. Secretary of Education, has referred to this achievement standard as the "Holy Grail" of education reform (Klein, 2009). This suggests not only the importance, but also its elusive quality; thus, useful definitions are critical.

> **"If we can dramatically increase high school graduation rates, if we can dramatically increase the number of graduates who are college and career ready, that's what this is about. Everything's a means to that end. That's the Holy Grail here. Are our students being prepared to be successful?"**
>
> ***Arne Duncan, U.S. Secretary of Education***
> ***Education Week, December 9, 2009***

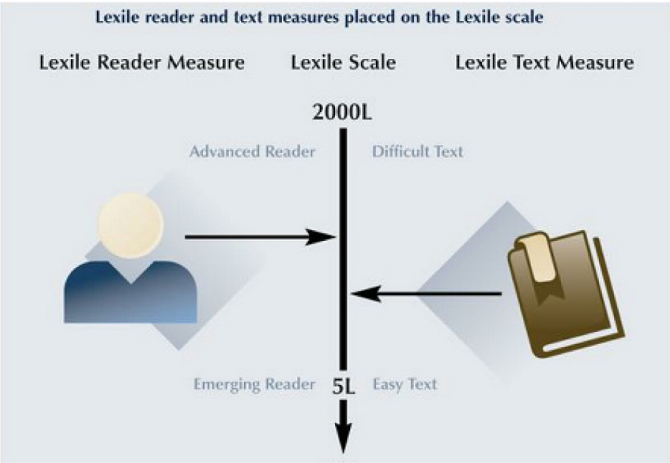
In the case of reading in English, there is general consensus with the notion that students should graduate from high school able to read the increasingly complex texts that await them in universities and jobs. Through our research in text complexity, we are demonstrating how reading demands can be quantified to define university and career readiness. The Lexile Framework enables the scientific measurement of text on a vertical scale that describes the full range of text complexities. Lexile analysis offers a

means to evaluate representative groups of texts from specified environments, such as freshman-level university courses or corporate HR collections. To date, we have measured over a 130,000 English books and more than 100 million articles, conducting along the way numerous studies that examined the characteristic reading demands within a wide range of settings. As a result, we were able to objectively quantify a meaningful target for university and career readiness in reading. In Lexile terms, this was found to be approximately 1200L to 1400L, which corresponds to a *TOEFL iBT®* Reading score of 18 or above. Unfortunately, we know that too few students can read at that level when they graduate from high school.

 # 2- Longitudinal Perspective

There is growing recognition that we need to take a longitudinal perspective on education if we are to effectively address the challenges of its fundamental goal of creating university and career ready citizens. Historically, many countries (including the U.S.) have segmented their educational systems into different groups with discrete and independent governance. That is, higher education has been separate from K-12 education, which in turn has been detached from early (preschool) education. Not surprisingly, the groups have not been well aligned. The result for those students who are unprepared if and when they advance is costly remediation to address gaps in their skills. Those who are prepared get and stay ahead.

In response to this issue, many countries are recognizing the importance of taking an integrated perspective that accommodates P-20 education (preschool to higher education) and beyond—something that can be humorously described as a pediatric-to-geriatric view on learning. This trend toward longitudinal perspectives in education reinforces the need for vertical scales that span the developmental continuum. Of course, this also must be reflected in our measurement systems.



Closer alignment of standards across the P-20 continuum is a key aspect of a longitudinal perspective, but a vertical scale is also essential for the meaningful analysis of growth to evaluate progress. Using a vertical scale—especially one that connects to practical definitions of readiness at all levels—provides a means of interpreting status and growth from a common perspective. Just imagine the confusion that would ensue in

measuring growth norms for height if we utilized non-vertical, non-exchangeable scales. Yet, too often in education, that is exactly what we have done. There are many good tests in use, but results are not easily comparable.

MetaMetrics produces scientific, vertical scales for educational measurement that promote longitudinal perspectives. But rather than limiting our scale to a particular test, we link our scale to the wide array of measurement instruments in use. In other words, we enable results from different assessments to report a common metric on a vertical scale. To date, we have linked close to a hundred assessments across the spectrum of high stakes, formative, and summative programs to our Lexile scale.

The trend toward longitudinal perspectives through a vertical scale is especially important for the measurement of reading since the end goal—university and career readiness—has been quantified in Lexile measures. To borrow from Steven Covey's work, an organization is better served when they start with a clear destination point. When we "begin with the end in mind," as Covey suggests, we can apportion the growth requirements in reading ability across the lifespan of the learner (Covey, 1989.). In fact, the authors of the U.S. Common Core State Standards using Lexile measures recently did this work.

Table 1: Text Complexity Grade Bands and Associated Lexile Ranges

Grade	Lexile Range
1	Up to 300L
2	140L to 500L
3	330L to 700L
4	445L to 810L
5	565L to 910L
6	665L to 1000L
7	735L to 1065L
8	805L to 1100L
9	855L to 1165L
10	905L to 1195L
11 and 12	940L to 1210L

(Source: Revised Grade Bands from the Common Core State Standards)

3- Digital Content

The shift from print to digital content in education—as in other segments of the publishing industry—presents unprecedented challenges and opportunities. The days of backpacks crammed full of "one size fits all" textbooks are fading fast in the emergent digital age of education. The production and distribution of physical textbooks is a costly undertaking with substantial investment risk. Economies of scale and least common denominators of curriculum drive development. Smaller education markets have long had to settle for materials developed with larger, higher value customers in mind. Few if any textbooks have adequately addressed the spectrum of needs and interests within a given classroom. The digital transition, however, is enabling customization at scale.

For publishing companies, the shift to digital requires new models for selling and delivering their high-quality content. It seems virtually inevitable (excuse the pun) that access will continue to shift toward online education portals, tablets, and other mobile devices—mediums that drastically expand distribution options and lower barriers to localized adaptation.

For students and educators, the shift to digital is generating new ways of viewing curriculum and assessment. However, major challenges emerge for navigating efficiently the abundance of online resources while retaining the deep curricular connections that education needs. While this digital trend holds great promise to truly democratize the delivery of educational opportunity across the globe, efficient tools and methodologies for finding 1) the right content at 2) the right level at 3) the right time must coincide with the adoption of digital resources. Without useful indexing and search mechanisms, the learner will be awash in a flood of disconnected and potentially fallacious digital content. The main opportunities inherent to digital would be largely lost if focused only on the medium.

As an example, Lexile measures can be integrated with digital library services to allow teachers and students the means to search for articles within a specified range of text complexity. Using automated processes, a Lexile measure can be produced and stored with meta-data. For example, the company EBSCO has measured *tens of millions* of articles from a broad range of publications, and incorporated Lexile-based search capabilities. Such digital enhancement of educational content is itself a transformative opportunity. But it is also an essential ingredient of the final common trend in worldwide education— individualized learning—which promises to harness and marshal the ocean of digital content in new and revolutionary ways.

#4- Individualized Learning

Perhaps the most remarkable innovations being realized during this transformative moment in education relate to the individualization of learning. Technology is breaking down barriers to taking an entirely student-centric approach to learning, one that enables "just-in-time" educational delivery and extends the boundaries of educational time and space well beyond the classroom. Versatile online platforms are emerging that can store and use vast amounts of student data to inform and adapt learning paths in real time. Computer-based assessments are being blended seamlessly into differentiated instructional activities that provide immediate feedback. This data is being used to power sophisticated, real-time reports for teachers and make the management of diverse needs in a classroom far more practical and efficient. Teachers are getting unprecedented views into student activity outside the classroom, such as actual time spent on homework assignments, along with highly efficient ways to communicate around those activities. The incorporation of vertical scales is ensuring that progress and goals are always front and center from a common, longitudinal perspective.

Our Engaging English® reading service, which is powered by Lexile technology, was designed with these benefits in mind. Students have daily access to thousands of online reading options that match their ability and specified interests. As they read, computer-based items provide immediate feedback on their reading experience and track improvement toward goals on the Lexile scale. When a student's Lexile measure increases, the system adjusts in real time the recommended text complexity of subsequent readings to ensure that every reading experience presents the optimal amount of challenge for a given day. Progress reports help students see their rate of improvement over time, and clearly understand how their measure compares to reading goals in Lexile terms, such as university and career readiness or target scores for a linked assessment (such as the TOEFL iBT® test). Additional monitoring tools help readers see the amount and diversity of their reading over time and reinforce the importance of daily practice. Game-like features help keep daily practice fun and interesting.

The Engaging English platform demonstrates how computer-adapted delivery of instructional content can be combined with computer-adaptive measurement of the learner on a common, vertical scale for a meaningful longitudinal perspective. When every instructional interaction can be mined for assessment data, the traditional paradigms of testing will become increasingly obsolete. No longer will we be limited to archaic testing windows at the end of the school year, during which time students are administered a fixed set of items from a costly exam. Rather, students will be continuously assessed and teachers will stay focused on their progress year round.

#5- Conclusion

These four trends place education at the dawn of a new day with the legitimate potential to revolutionize how we teach and assess. For many generations, the delivery of education has remained largely unchanged: children are grouped by age and advance together through a series of classrooms along a prescribed path of print-based exams and curricular materials. Teachers have had few supports for differentiating instruction inside and outside of school. Today however, preschool children share the promise of a very different educational experience: individual learning needs will be foremost at all times; progress toward university and career readiness will be clear and present throughout their educational life span; continuous assessment will be used to inform their daily instruction and practice without intrusion; and enhanced digital resources will engage more fully their passion for learning.

To realize this potential, it is critical to establish foundations upon which to build and grow. The use of common, vertical scales that empower adaptive assessment, differentiated instruction, and persistent longitudinal views with meaningful standards are all essential ingredients. When it comes to reading, the Lexile Framework provides the requisite foundation upon which this future can be realized.

REFERENCES

Covey, S. (1989). The 7 Habits of Highly Effective People. New York: Simon & Schuster.

Klein, A. (2009). An interview with Arne Duncan. Education Week, 29(14), Retrieved from http://www.edweek.org/ew/articles/2009/12/02/14duncan-transcript.h29.html

SECTION 3:
The Lexile Framework and Text Complexity

MetaMetrics White Paper | 1490L
February 10, 2012

Not So Common:

Comparing Lexile Measures With the Standards' Other Text Complexity Tools

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D.

Executive Summary

Since the Common Core State Standards were published last year, much national attention has focused on the importance of text complexity in evaluating college- and career-readiness. Common Core authors David Coleman and Sue Pimentel have stated that understanding and measuring text complexity is a major shift in the new English Language Arts Standards and that these criteria are key to determining if students are adequately prepared for the academic and professional reading demands they will likely face after high school (The Hunt Institute, 2011). Subsequent reports,"Publishers' Criteria for the Common Core State Standards in English Language Arts and Literacy" (for grades K-2 and Grades 3–12) and the more recent, "Measures of Text Difficulty: Testing Their Predictive Values for Grade Levels and Student Performance," have echoed these same text complexity themes.

MetaMetrics focuses on the importance of matching individual readers with targeted texts that provide the right level of challenge to support continued reading growth. Long before the Common Core movement, The Lexile Framework for Reading played an important role in articulating the reading demands typically encountered in first grade through college and careers. In fact, MetaMetrics' research on K-12 reading demands and ultimately those of the postsecondary world are annotated in the text complexity "staircase" in the Standards' English Language Arts Appendix A. This staircase approach to text complexity is designed to help guide students' reading comprehension development through their school years.

The subsequent reports noted earlier were intended to provide policy makers, educators and publishers with additional information and guidance on the value of measuring text complexity. While the "Publishers' Criteria" reports primarily reiterate much of the research contained in the Common Core, the "Measures of Text Difficulty" report offers a detailed analysis of the text complexity landscape in terms of the tools commonly found in the marketplace. The report compares six text complexity tools—Carnegie Mellon University's and the University of Pittsburgh's REAder-specific Practice(REAP), Renaissance Learning's ATOS, Questar Assessment's Degrees of Reading Power® (DRP®), Pearson's Reading Maturity Metric, ETS's SourceRater and MetaMetrics' Lexile measure—using various criterion outcomes. A description of a seventh tool, Coh-Metrix's Text Easability Assessor, is also mentioned. In summary, the report states that "there is no agreed upon gold standard" for evaluating text complexity. Its comparisons of the text complexity tools demonstrate that while they share some commonalities, there are also distinct differences (Nelson, Perfetti, Liben, & Liben, 2011). Building upon the report's findings, this document provides a contextual framework for how these similarities and differences could be interpreted and used by the educational and publishing communities when selecting a text complexity tool.

As state departments of education, and the districts and schools within those respective states, transition from adopting the new Common Core State Standards to the more difficult task of implementing them,the challenge now becomes how to translate these higher standards into tangible, practical and cost-effective curricula. Implementing the Common Core will require districts and schools—and the educational publishers who supply curricular materials—to develop new instructional strategies and complementary resources that are not only aligned with the national college- and career- readiness standards, but also utilize and incorporate validated and budget-friendly tools that are universally accessible to all stakeholders.

The Relevance of Text Complexity

The Standards for English Language Arts focus on the importance of text complexity. As stated in Standard 10, students must be able to "read and comprehend complex literary and informational texts independently and proficiently" (Common Core State Standards for English Language Arts, College and Career Readiness Anchor Standards for Reading, NGA Center and CCSSO, 2010, p.10). The Common Core notes the following reasons for incorporating these more rigorous standards:

1. The text complexity of K-12 textbooks has become increasingly easier over the last 50 years (Chall, Conrad,& Harris, 1977; Hayes, Wolfer, & Wolfe, 1996).

2. The text demands of college and careers have remained consistent or increased over the same time period (Common Core State Standards for English Language Arts, Appendix A, NGA Center and CCSSO, 2010, p. 2).

3. As a result, there is a significant gap between students' reading abilities and the text demands of their postsecondary pursuits. The Common Core states, "Being able to read complex text independently and proficiently is essential for high achievement in college and the work-place and important in numerous life tasks" (Common Core State Standards for English Language Arts, Appendix A, NGA Center and CCSSO, 2010, p. 4).

The Common Core recommends a three-part model for evaluating the complexity of a text that takes into account its qualitative dimensions, quantitative measure, and reader and task considerations. It describes text complexity as "the inherent difficulty of reading and comprehending a text combined with consideration of reader and task variables...a three-part assessment of text [complexity] that pairs qualitative and quantitative measures with reader-task considerations" (NGA Center and CCSSO, 2010,

p. 43). In short, text complexity is a transaction between text, reader and task. The Lexile Framework for Reading is based on this transaction and is significant because it allows for matching individual readers with specific texts on the same developmental scale.

MetaMetrics' research on the importance of the reader-with-text match and the typical reading demands of college and careers contributed to the Common Core as a whole and, more specifically, to the Lexile-based bands in Table 1.

Table 1: Text Complexity Grade Bands and Associated Lexile Ranges

Text Complexity Grade Bands	Lexile Ranges Aligned to College- and Career-Readiness Expectations
K-1	N/A
2-3	420L–820L
4-5	740L–1010L
6-8	925L–1185L
9-10	1050L–1335L
11-CCR	1185L–1385L

Note: There is an infinity of ways of connecting a single starting point in grade 1 to a single ending point in grade 12. Each of these ways may find a supportive constituency and a constituency in opposition.

Text Complexity and The Lexile Framework for Reading

Since its inception in 1984, MetaMetrics has focused on the development and refinement of a scientific scale that measures text complexity and, more importantly, places individuals on that same scale to evaluate reading ability. Today, The Lexile Framework for Reading is used by nearly half of U.S. state departments of education and tens of millions of individuals worldwide (primarily because *ETS®'s TOEFL iBT®* test, *TOEFL® JuniorTM* test and *TOEIC®* test report a Lexile measure). Consistent with the Common Core's definition of text complexity as the transaction between reader, text and task, the originating psychometric method used to develop the Lexile Framework was based on the relationship between individuals' actual reading comprehension (for a given task) while reading specific texts. The Lexile Framework is not a readability formula. Readability formulas are based only on the features and/or characteristics of a

text. As stated in the Common Core, "...[A]n important difference between the Lexile system and traditional readability formulas is that traditional formulas only assign a score to texts, whereas the Lexile Framework can place both readers and texts on the same scale" (NGA Center and CCSSO, 2010, p. 7).

Table 2: Six Text Complexity Tools' Correlations with Text Complexity Measures from Texts with Previously Determined Complexity Levels (summarized by MetaMetrics)

Reference Set	REAP	ATOS	DRP	Lexile Measure	Reading Maturity	SourceRater
EdSphere™						
All	0.629	0.924*		0.946*	0.879	Did Not
Subset	0.679	0.921*	0.893	0.935*	0.0871	Participate

*Significantly different from all other Spearman rho correlations based on non-overlap of 95% confidence internals for each estimate.
Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2011). Measures of Text Complexity: Testing their Predictive Value for Grade Levels and Student Performance. Student Achievement Partners.

MetaMetrics' vision of a common scale to measure both text complexity and reading ability was shared by policy makers and researchers at the National Institutes of Health, which supported the organization's work through five grants over a ten-year period. As the education community transitions to the Common Core, stakeholders will need to be cognizant of both text complexity and reading ability, and how they can use this information to accurately match individual readers with targeted texts. Accelerating reader growth and scaffolding support to the more rigorous college- and career-readiness levels can be accomplished most effectively when a common scale connects readers with appropriate texts.

Comparing the Standards' Text Complexity Tools

In addition to the Common Core, two more supporting documents were published last year on the importance of text complexity. The "Publishers' Criteria for the Common Core State Standards in English Language Arts and Literacy, Grades K-2" and "Publishers' Criteria for the Common Core State Standards in English Language Arts and Literacy, Grades 3-12" reinforce the value of measuring text complexity for educational publishers. A third report, "Measures of Text Difficulty: Testing Their Predictive Value for Grade Levels and Student Performance," released in December

2011 describes the various text complexity tools available to the education community: Carnegie Mellon University's and University of Pittsburgh's REAder-specific Practice (REAP), Renaissance Learning's ATOS, Questar Assessment's Degrees of Reading Power (DRP), the Pearson Reading Maturity Metric, ETS's SourceRater, MetaMetrics' Lexile measure, and Coh-Metrix's Text Easability Assessor.

Noting that "there is no clear gold standard measure for text difficulty against which to compare the various metrics," the report authors evaluated each of the tools (except Coh-Metrix's Text Easability Assessor) using three basic types of criterion measures (Nelson, Perfetti, Liben, & Liben, 2011). Table 2 provides an analysis of the report's findings on one of the criterion measures. The "Reference Set" column includes the five reference sets that were used as criterion measures. Each of the six text complexity tools in the top-row headings was correlated (using Pearson Product Moment Correlations) with predetermined text complexity levels in the reference sets. None of the correlations were corrected for the measurement error or any other artifactual sources of variation.

While there are five different criterion reference sets, the sets fall into three categories:

1. CCSS (Common Core State Standards) exemplars comprised of passages in which the text complexity has been determined based on human judgment.

2. State test passages, SAT-9, and GMRT are based on empirical student test performance within a grade level.

3. EdSphere™ data is based entirely on empirical data of student responses to text passages.

The first two categories of criterion measures suffer from serious limitations which compromise their utility. The first category is comprised of a disproportionate number of passages that had a significantly different measure than the text complexity measure for the complete book. This bias calls into question its validity and utility. The second set of criterion measures while at least based on empirical student performance is compromised due to the truncation of scores into grade levels. The categorization of the criterion into grade-level-groupings restricts the correlation coefficients of text complexity metric that is a vertical scale. Even with these limitations all the tools were moderately correlated with the criterion measures.

The third criterion measure is based entirely on empirical student performance without the restriction of collapsing into grade levels. For the EdSphere criterion, all of the text complexity tools were moderately correlated to highly correlated.

Table 3: Text Complexity tools

	REAP	ATOS	DRP	Lexile Measure	Reading Maturity	SourceRater
Year Developed	2011	1998*	1980**	1996	1996	1996
Scale Type	Grade Level	Grade Level	Vertical	Vertical***	Grade Level	Grade Level
Based on Student-Test-Track transaction	Yes	Yes	Yes	Yes	No	No
Free Text Analyzer Available	Yes	Yes	Yes	Yes	Yes	No
Spanish Scale and Online Analyzer	No	Yes	No	Yes	No	Yes

*"The ATOS™ Readability Formula for Books and How it Compares to Other Formulas". (2000). School Renaissance Institute. Madison, WI.
http://www.trademarkia.com/drp-77072604.html * http://www.lexile.com/about-lexile/grade-equivalent

---

Table 4: Number of Texts Measured to Date

	REAP	ATOS	DRP	Lexile Measure	Reading Maturity	SourceRater
Books Measured	--	124,861	38,000	168,000*	--	--
Textbooks Measured	--	--	7,000	5,000+	--	--
Articles Measured	--	0	--	100,000,000**	--	--
Webpages Measured	--	0	0	60,000**	--	--

* Updated August 2013 **https://d1jt5u2s0h3gkt.cloudfront.net/m/uploads/positionpapers/TheReading-WritingConnection.pdf
--As of November 2011, this information was not publicly available.

Table 5: Number of Assessments Reporting the Measure

	REAP	ATOS	DRP	Lexile Measure	Reading Maturity	SourceRater
State Assessments	--	0	1*	21**	0	0
Norm-Referenced Assessments	--	0	--	10	--	--
Benchmark Assessments	--	0	--	9	--	--
Student Measures Each Year	--	30,000+***	--	35,000,000	--	--

*http://en.wikipedia.org/wiki/Connecticut_Mastery_Test
**http://www.lexile.com/using-lexile/lexile-measures-and-the-ccssi ***http://www.renlearn.com/ar/overview/atos/
--As of November 2011, this information was not publicly available.

It is important to note that the correlations for the text complexity tools (except for ETS's SourceRater which was not part of this reference set) were higher on this criterion than for the other four references sets. While all of the correlations were higher, the Lexile and ATOS scales had the highest correlations. MetaMetrics believes that this criterion measure is the ultimate litmus test for text complexity tools: How well does the tool estimate text complexity from actual student comprehension? A second and related point is the importance of placing students in terms of their reading ability on the same scale.

Selecting and Using a Text Complexity Tool

Besides the correlation coefficients, there are many other features and dimensions that should be considered when selecting a text complexity tool. MetaMetrics believes that the most important feature is the ability to place both a reader and a text on the same developmental scale, which is the essence of the Lexile Framework. Simply ordering text from easy to difficult on a scale is of limited value without measuring and ordering an individual's reading ability on that same scale.

Tables 3, 4 and 5 on the previous page highlight some of the key features and dimensions of the Standards' text complexity tools. Table 3 focuses on the foundational characteristics; Table 4 focuses on the type and quantity of text that has been measured

by each tool; and Table 5 focuses on the measurement of students. As these tables illustrate, there area number of distinct advantages for using Lexile measures.

First, the Lexile Framework is the only text complexity measurement system that is originally psychometrically derived from individuals reading texts—the reader-text-task transaction. MetaMetrics theorized this transaction and discovered that select semantic and syntactic text features could serve as proxies for the mental processes individuals use when reading texts. These text proxies account for an extremely high amount of variance (94 percent) in readers' comprehension.

MetaMetrics has conducted studies to explore additional proxies that could explain the remaining six percent of variance(Hanlon, Swartz, Stenner, Burdick, & Burdick, 2010; Kershaw, Barth, Francis, Swartz, & Stenner, 2011; Stenner, Burdick, Burdick, Hanlon, & Swartz, 2010; Stenner, Swartz, Burdick, Burdick, & Hanlon, 2010; Swartz, Burdick, Hanlon, Stenner, Kyngdon, Burdick, & Smith, 2011). The organization has also provided its data sets to third-party researchers and organizations to see if they could improve the Lexile measure. While this research continues, it is important to note that, to date, none of the studies has led to the discovery of additional proxies (in a statistically significant way). While additional proxies could be added, MetaMetrics has found that they do not provide any more explanatory power to the Lexile equation. And as any researcher will attest, simply adding more variables to an equation does not necessarily produce a more accurate result.

Second, the Lexile Framework was built using a vertical, developmental and equal-interval scale, permitting finer-grained analyses than grade-related scales and avoiding the pitfalls associated with grade-related scales. More than three decades ago, the International Reading Association resolved that grade-equivalent reports should be abandoned because they were being misconstrued and misused (IRA, 1981).

MetaMetrics' and Questar's tools are built upon vertical, equal-interval scales. Renaissance Learning, ETS and Pearson Educational Measurement have opted to create grade-based or grade-equivalent scales. On the surface, grade equivalents appear to be an intuitive way to report students' test scores. However, this seemingly simplistic method glosses over some significant limitations that often promote misleading and inaccurate interpretations of the data. These misconceptions include: 1) norms are standards of what should be; 2) grade equivalents indicate the appropriate grade placement for a student; 3) all students should be expected to grow one grade-equivalent unit per year; 4) the units are equal throughout the score range; 5) grade equivalents for different tests are comparable; and 6) the scores that are based on extrapolations to grades well above or below the test level are meaningful (Miller, Linn & Gronlund, 2009).

Third, the Lexile Framework places reader and text on the same developmental (Lexile) scale, which enables for more accurate reader-with-text matches. As numerous researchers and educational practitioners attest, this feature provides classroom teachers with an actionable resource to support differentiated instruction (Chall & Dale, 1995; Hall & Moats,1999; Hiebert, 2009; Hiebert & Mesmer, 2011; Mesmer, 2007).

Fourth, a topic that is often overlooked but has profound implications is the importance of "editing" a text for analysis, as different tools have different rules for how text should be prepared for measurement. MetaMetrics makes its rules transparent by publishing them on the Lexile website. Without precise rules for how text should be prepared for analysis there will be volatility in the measures. MetaMetrics also provides open access to its Lexile Analyzer to generate Lexile measures for various length text files, from short samples to entire books. More than 228,000 individuals have registered for this free resource.

Fifth, pragmatic factors should be considered when selecting a text complexity tool. Factors such as ubiquity of use, access to complementary resources, and the number of books, assessments and products that are built on these text complexity tools are paramount. Lexile measures, for example, are available for more than 150,000 books from more than 200 publishers in MetaMetrics' free "Find a Book" search, and in the search tools from domestic and international booksellers like Barnes & Noble and Korea's Interpark Books. In addition, more than 5,000 textbooks from more than 20 publishers have been measured. More and more, publishers are submitting their materials for Lexile measurement to ensure alignment with the Common Core. Tens of millions of articles also have Lexile measures and are available from content aggregators such as EBSCO, ProQuest, Gale Cengage and Newsbank. And more than 200,000 web pages have been measured. All of these resources with Lexile measures—many of which are used in schools and libraries—allow students to read targeted narrative and informational text on classroom relevant topics or personal interests.

A final consideration involves the assessment of the reader. As noted previously, it is imperative that readers and texts are placed on a common scale. Currently, there are only three systems (ATOS, DRP and the Lexile Framework) which measure the reader and the text. However, the availability of Renaissance Learning's ATOS and Questar Assessment's DRP metrics are limited to those organizations' proprietary tests. Schools and districts must purchase the STAR test from Renaissance Learning or the DRP test from Questar. MetaMetrics, in contrast, does not publish or sell its own test. The organization maintains open, nonproprietary relationships with major educational test publishers, having linked the Lexile Framework with more than 60 popular reading assessments and programs. ETS and Pearson alone offer Lexile measures from a total of eleven reading assessments. This approach provides districts and schools with more

options and greater flexibility in determining which instrument to use. (Please see Appendix A for a complete list of assessments and reading programs that report Lexile measures.)

Conclusion

Only three of the text complexity tools included in the "Measures of Text Difficulty" report have a direct and empirical link to readers and texts: ATOS, DRP and the Lexile measure. The other three tools, SourceRater, REAP and the Reading Maturity measure, evaluate text only and, to date, are not reported on any standardized tests.

When it comes to selecting a tool to measure reading ability and text complexity, there are two important criteria to consider. The first is ubiquity, or how many resources like books, articles and websites have been measured. As mentioned earlier, Lexile measures are available for more than 150,000 books, tens of millions of articles, and hundreds of thousands of web pages.

The second criteria is the "accessibility" of the tools for measuring text complexity. The success of the Common Core will be determined by how the Standards are translated into classroom practice. MetaMetrics strives to provide the tools and complementary resources to make this a reality. For example, the Lexile Analyzer is freely available on the Lexile website for educators and others to determine the Lexile measure of a text. In addition, publishers can license the software to build commercial products or to report certified Lexile measures in public domains like websites, databases and catalogs. To date, MetaMetrics has provided more than 228,000 users with access to the free Lexile Analyzer. Over the past two years alone, users have analyzed nearly one million text files. And the number of words analyzed for commercial use exceeds 1.5 billion.

As Stephen Covey points out in his book, "The Seven Habits of Highly Successful People," it is best to begin with the end in mind. The authors of the "Measures of Text Difficulty" report offer a text complexity staircase for college- and career–readiness in Appendix C. By mapping required Lexile growth backwards across grade levels and providing access to resources to measure text complexity and monitor student growth, policy makers and educators will be better able to implement and sustain the Common Core as they prepare all students for the reading demands of the postsecondary world.

REFERENCES

AERA/APA/NCME Joint Committee. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Airasian, P. W. (1994). Classroom assessment (2nd ed.). New York: McGraw-Hill, Inc.

Chall, J. S., Conrad, S., & Harris, S. (1977). An analysis of textbooks in relation to declining SAT scores. Princeton, NJ: College Entrance Examination Board.

Chall, J. S & Dale, E. (1995). Readability Revisited: *The New Dale-Chall Readability Formula*. Cambridge, Massachusetts: Brookline Books.

Hall, S. L. & Moats, L. C. (1999). *Straight Talk About Reading: How Parents Can Make a Difference During the Early Years*. Chicago, Illinois: Contemporary Books.

Hanlon, S. T., Swartz, C. S., Stenner, A. J., Burdick, H., & Burdick, D. S. (2010). OasisTM Literacy Research Platform. Durham, North Carolina: http://www.alearningoasis.com. Retrieved November 15, 2010. Available from MetaMetrics.

Hayes, D.P., Wolfer, L.T., & Wolfe, M.F. (1996). Source book simplification and its relation to the decline in SAT-Verbal scores. *American Educational Research Journal*, 33, 489-508.

Hiebert, E. H., & Mesmer, H. A. (2011). *Text complexity in the primary grades*. Paper presented at the meeting of the Pacific Coast Research Conference, San Diego, CA.

Kershaw, S., Barth, A., Francis, D., Swartz, C. W., & Stenner, A. J.(2011, February). *A factor analysis of text cohesion.* Paper presented at the annual meeting of the Pacific Coast Research Conference, San Diego, CA.

Mesmer, H. (2007). Tools for Matching Readers to Text: *Research-Based Practices*. New York, New York: The Guilford Press.

Miller, M., Linn, R., & Gronlund, N. (2009). Measurement and assessment in teaching (10th ed.). New Jersey: Pearson Education.

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

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2011). Measures of Text Difficulty: Testing their Predictive Value for Grade Levels and Student Performance. Student Achievement Partners.

Stenner, A. J., Burdick, H., Burdick, D. S., Hanlon, S. T., & Swartz,C. W. (2010, January). Evaluating predictors of text complexity for extended passages. D. Coleman (Chair). *Text Complexity*. Symposium conducted at the meeting of Student Achievement Partners, Chicago, IL.

Stenner, A. J., Swartz, C. W., Burdick, H., Burdick, D. S., & Hanlon,S. T. (2010, September). Relationship between theoretical and empirical estimates of text complexity for two different reading item types. D. Coleman (Chair), *Text Complexity*. Symposium conducted at the meeting of Student Achievement Partners, Brooklyn, NY.

Stiggins, R. J. (2009). Student-centered classroom assessment (Second Edition). Upper Saddle River, NJ: Merrill, an imprint of Prentice Hall.

Swartz, C. W., Burdick, D. S., Hanlon, S. T., Stenner, A. J., Kyngdon, A., Burdick, H., & Smith, M (2011, April). *Toward a theory relating text complexity, reader ability, and reading comprehension*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

The English Language Arts Standards: Key Changes and their Evidence. (September 1st, 2011). The Hunt Institute. Durham, North Carolina.

Williamson, G. L. (2008). A Text Readability Continuum for Postsecondary Readiness. *Journal of Advanced Academics*, vol. 19(4), 602-632.

Appendix: Assessments and Reading Programs that Report Lexile Measure

State Assessments

- Arizona's Instrument to Measure Standards
- California English-Language Arts Standards Test
- Delaware Comprehensive Assessment System*
- Georgia Criterion-Referenced Competency Tests and End-of-Course Tests (Ninth Grade Literature and Composition, and American Literature and Composition)
- Hawaii State Assessment
- Illinois Standards Achievement Test
- Kansas State Assessments of Reading
- Kentucky Performance Rating for Educational Progress
- Minnesota Comprehensive Assessments
- New Mexico Standards-Based Assessment
- North Carolina End-of-Grade Tests and English I End-of-Course Test
- Oklahoma Core Curriculum Tests
- Oregon Assessment of Knowledge and Skills
- Proficiency Assessments for Wyoming Students
- South Carolina Palmetto Assessment of State Standards
- South Dakota State Test of Educational Progress
- Tennessee Comprehensive Assessment Program Achievement Test
- Texas Assessment of Knowledge and Skills**
- Virginia Standards of Learning Tests
- West Virginia WESTEST 2

Norm-Referenced Assessments

- CTB/McGraw-Hill: TerraNova (CAT/6 and CTBS/5) and Tests of Adult Basic Education (TABE)
- ERB: Comprehensive Testing Program, 4th Edition
- Pearson: Stanford 9 & 10, MAT 8 and Aprenda 3
- Riverside Publishing: The Iowa Tests (ITBS and ITED) and Gates-MacGinitie Reading Tests, Fourth Edition

Interim/Benchmark Assessments

- American Education Corporation: A+ Learning Link
- Curriculum Associates: i-Ready® Diagnostic & Instruction
- Dynamic Measurement Group: DIBELS
- Measured Progress: Progress Toward Standards (PTS3)

- Measurement, Inc.: Item Bank
- NWEA: Measures of Academic Progress® (MAP®)
- Pearson: Stanford Learning First, Stanford Diagnostic
- Reading Test, Fourth Edition, and AIMSweb
- Scantron: Performance Series
- Scholastic: Scholastic Reading InventoryE/S
- School Speciality: Making Connections Intervention

Reading Programs and Interventions

- Achieve3000: KidBiz3000 and TeenBiz3000
- Cambium Learning Group (Sopris/Voyager): Language! and Passport Reading Journeys
- Capstone Digital: myON reader
- EDmin: Total Reader
- EPS: MCI (Making Connections® Intervention)
- Hampton-Brown: Edge and Inside
- Houghton Mifflin: Earobics Reach
- McGraw-Hill: SRA FLEX LiteracyTM
- Pearson/Longman/Prentice Hall: MyReadingLab
- Scholastic: Read 180 and Reading Counts!

International Assessments

- ETS: TOEFL iBT® test, TOEFL® JuniorTM test and TOEIC® test
- GL Assessment: Progress in English (PIE) Assessment
- Natmal: E-LQ Assessment

• Spanish Assessments

- Achieve3000: KidBiz3000 and TeenBiz3000
- New Mexico Standards-Based Assessment
- Pearson: Aprenda 3
- Scholastic: Scholastic Reading Inventory

*Linking study in 2012
**Grades 10 and 11 only

MetaMetrics White Paper | 1290L
October 25, 2013

Quick and Easy Access to Measures of Text Complexity Using the Lexile Framework for Reading

By Matt Copeland, Lexile Curriculum Specialist, MetaMetrics Inc. and David Liben, Senior Content Specialist, Student Achievement Partners (SAP)

Quick and Easy Access to Measures of Text Complexity Using the Lexile Framework for Reading

As schools across the country continue their implementation of standards for college and career readiness, educators find themselves re-examining and reconsidering the complexity of the texts they ask students to read. And rightly so. Research[7] "makes clear that the complexity level of the texts students read are significantly below what is required to achieve college and career readiness" (Coleman & Pimentel, 2012).

Likewise, parents understand the need to provide their students with the variety, quality, and volume of texts that will help them progress toward their personal college and career goals. Parents understand the importance of reading as a gateway skill that helps students build knowledge and understanding of the world around them that will be drawn upon throughout their lifetimes. Students must be able to comprehend and embrace a range of texts that cross genres, cultures, and eras and that also models the kind of thinking and writing they should aspire to in their own work as well.

To this end, new educational standards for college and career readiness "hinge on students encountering appropriately complex texts at each grade level to develop the mature language skills and the conceptual knowledge they need for success in school and life" (Coleman & Pimentel, 2012).

One of the questions most frequently asked is how educators and parents can find valid and reliable measures of text complexity to address this need. Luckily, there are a number of quick and easy methods an individual can use to find a Lexile measure of a text. With this information, educators and parents can better match their students to texts with the appropriate level of complexity.

The Lexile Framework for Reading is an approach to reading and text measurement. There are two Lexile measures: the Lexile text measure and the Lexile reader measure. A text receives a Lexile measure by running it through the Lexile analyzer which utilizes a linguistic algorithm that examines the semantic and syntactic features of the text. The lower the Lexile text measure the easier it is to read. For example, *Frog and Toad Together* (Lobel) is a 330L text, *Charlotte's Web* (White) is a 680L text, *The Pearl* (Steinbeck) is a 1010L text, and *The House of the Spirits* (Allende) is a 1280L text. Lexile text measures are rounded to the nearest 10L. 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, her reader measure is 880L. Higher Lexile measures represent a higher level of reading ability. A Lexile reader measure can range from below 200L for early

[7] [7] See Chall, Conard, Harris (1977); Hayes, Wolfer, and Wolfe (1996); Williamson (2006); ACT, Inc. (2006).

readers to above 1600L for advanced readers. More information about the Lexile Framework for Reading can be found by visiting the following URL: http://www.lexile. com/about-lexile/lexile-overview/

The Lexile Framework, which comprises the Lexile measures and Lexile scale, is not an instructional program any more than a thermometer is a medical treatment. But just as a thermometer is useful in informing medical care, the Lexile Framework is useful in informing a student's reading development.

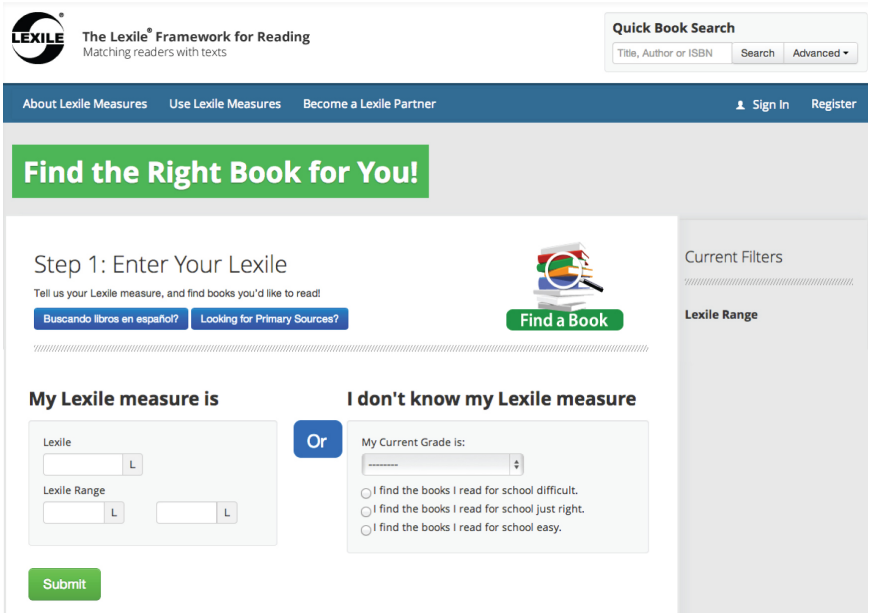
Among the convenient methods of finding a Lexile measure of a text and evaluating its level of complexity are the following:

- The "Find a Book" feature at Lexile.com
- Websites of prominent booksellers Barnes & Noble.com and Amazon.com
- Online educational and research databases
- The Lexile Analyzer at Lexile.com

"Find a Book" at Lexile.com

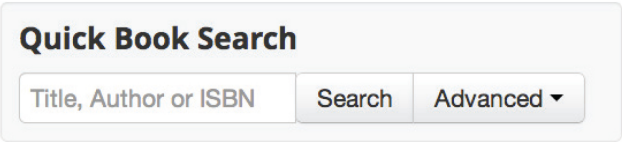
Visitors to the Lexile.com website have access to the "Find a Book" feature which will allow them to quickly and easily identify a Lexile measure of a book. Figure 1 shows the user interface for "Find a Book."

Figure 1: The "Find a Book" user interface



For visitors who know the title, author, or ISBN of the text they are searching for, they can use the quick book search feature located in the upper right-hand corner of the screen (Figure 2). This quick search allows the user to locate the Lexile measure of a text.

Figure 2: Lexile.com's quick book search feature



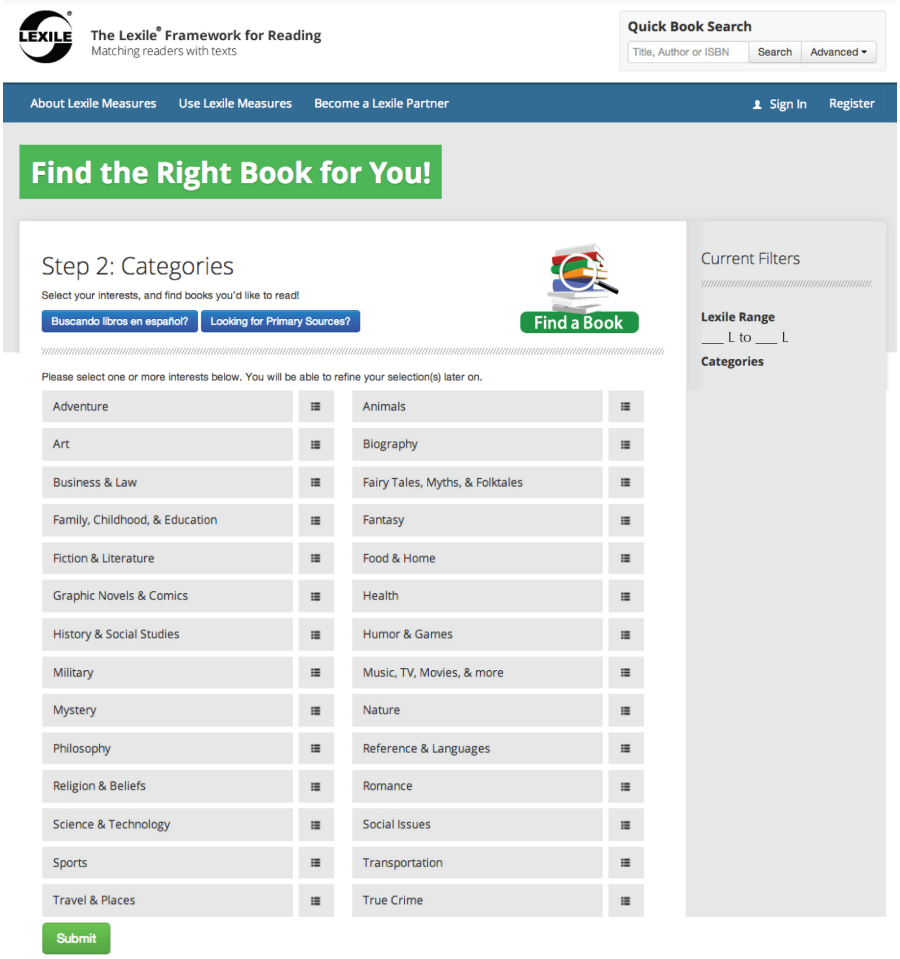
For example, if a visitor wanted to find a Lexile measure for Joseph Bruchac's young adult biography of Jim Thorpe, she could simply type the title, author, or ISBN into the quick book search feature. From the results list, she could then click on the link to the book's page and quickly find the Lexile measure (circled in red in Figure 3).

Figure 3: Book Detail Page for Jim Thorpe, Original All-American with the Lexile measure circled in red

The book detail page also includes quick links to be able to purchase the book from an online bookseller, if desired, or locate the nearest library that has the book available for check out.

Additionally, visitors who may not know the precise text they are looking for can also utilize the "Find a Book" feature to search for titles within a particular Lexile range. By entering a Lexile reader measure or a Lexile range, visitors can search for new books in one or more of 350+ categories and sub-categories (as shown in Figure 4 on the following page).

Figure 4: Interest categories on "Find a Book"



These results, too, report a Lexile measure for each title. Searches can even be saved and returned to at a later date for easy access.

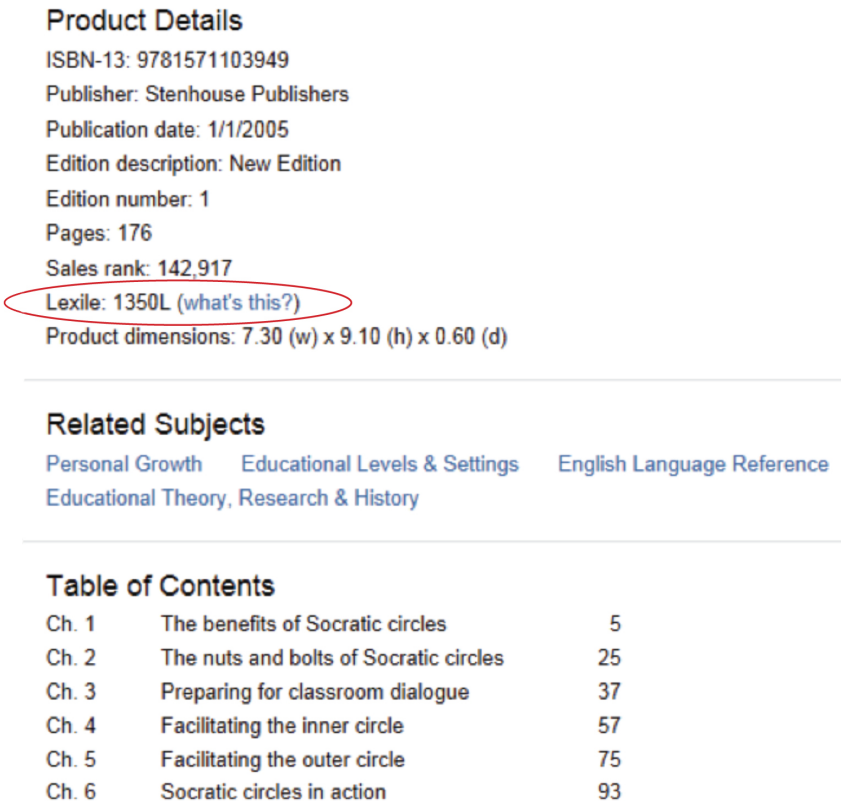
Occasionally, there are titles that are not included in the "Find a Book" feature or shorter works that are not book length. For texts such as these, please consult the Lexile Analyzer section below.

Websites of Prominent Online Booksellers

Major online booksellers like Amazon.com and Barnes & Noble.com provide Lexile measures for over 150,000 of their books. Barnes & Noble offers a Lexile wizard that allows users to build a custom book list based on their Lexile measure and interests. For more information on Lexile measures and Barnes & Noble, visit www.lexile.com/using-lexile/barnes-noble/.

The Barnes & Noble website often lists a Lexile measure for titles in the "Product Details" section of their webpages. Figure 5 shows an example of this idea with the Lexile measure circled in red.

Figure 5: An example of a Barnes & Noble webpage with the Lexile measure circled in red



Product Details

ISBN-13: 9781571103949
Publisher: Stenhouse Publishers
Publication date: 1/1/2005
Edition description: New Edition
Edition number: 1
Pages: 176
Sales rank: 142,917
Lexile: 1350L (what's this?)
Product dimensions: 7.30 (w) x 9.10 (h) x 0.60 (d)

Related Subjects

Personal Growth Educational Levels & Settings English Language Reference
Educational Theory, Research & History

Table of Contents

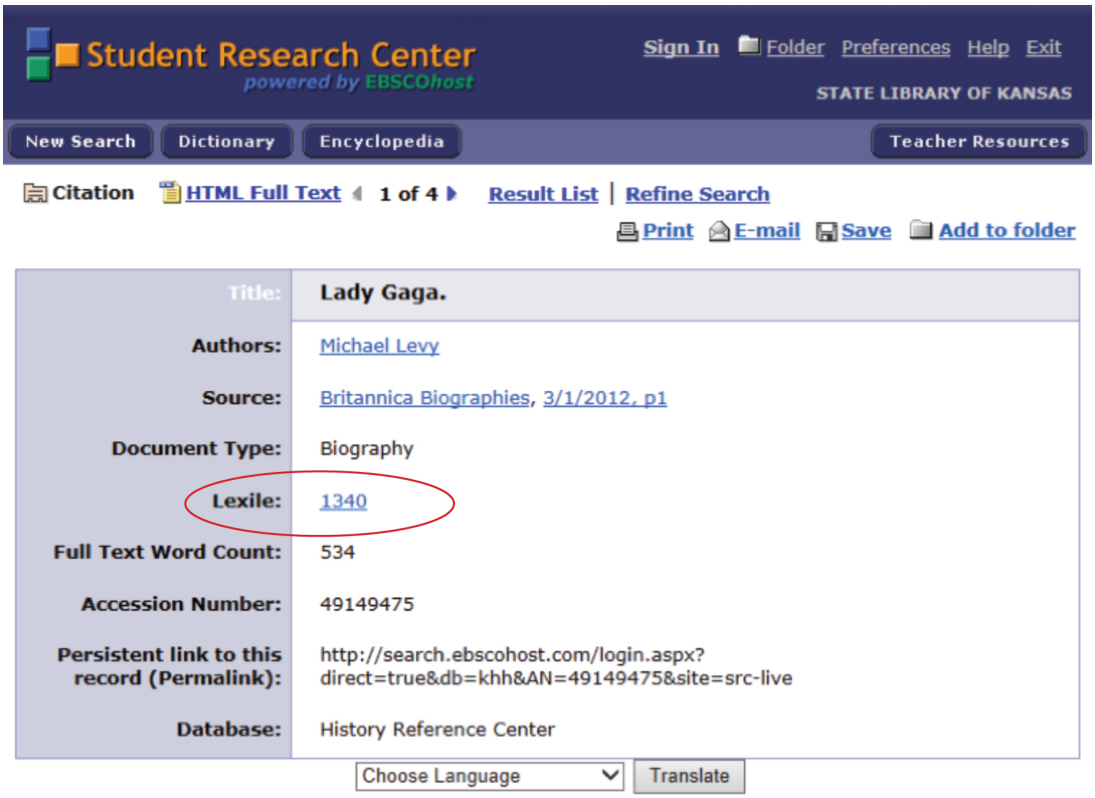
Ch. 1	The benefits of Socratic circles	5
Ch. 2	The nuts and bolts of Socratic circles	25
Ch. 3	Preparing for classroom dialogue	37
Ch. 4	Facilitating the inner circle	57
Ch. 5	Facilitating the outer circle	75
Ch. 6	Socratic circles in action	93

Online Educational Research Databases

Finding Lexile measures for short, article length pieces is also fast and easy thanks to the many educational and research databases such as EBSCO®, Gale®, World Book®, and ProQuest®. Many states offer these databases free of charge to constituents living in their state. For a complete list of available databases listed by state, please visit here.

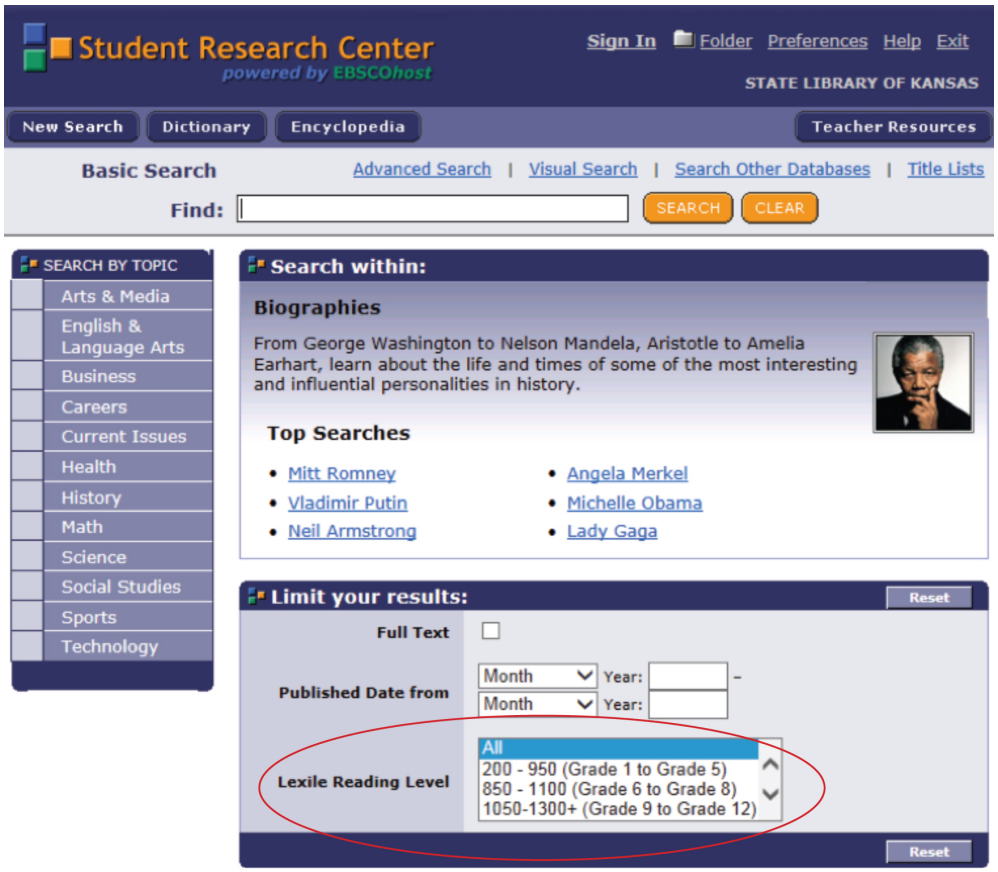
The vast majority of the online databases are quite user friendly and operate in a similar fashion. As an example, Figure 6 shows a screen shot of the user interface for Student Research Center powered by EBSCO available from the state library of Kansas. Much like the "Find a Book" feature at Lexile.com, many of these online databases allow users to limit their results to an identified Lexile range.

Figure 6: The user interface for the Student Research Center powered by EBSCO with ability to limit search results to certain Lexile ranges circled in red



And once users have identified an article, many of these online databases also report a specific Lexile measure for that particular text. For example, Figure 7 shows the bibliographic information for a text available online. Here, too, the Lexile measure is circled in red.

Figure 7: Bibliographic information for an article from an online database with the Lexile measure circled in red



The Lexile Analyzer

When educators or parents have an electronic copy of a text for which they cannot locate a Lexile measure, the Lexile Analyzer can provide an estimate measure. Access to the Lexile Analyzer is free for registered users. Initially, a person will have a 1000-word limit; however, she can also request access to the Professional Lexile Analyzer to measure longer pieces of text.

Before being analyzed, a text file requires preparation. There will likely be elements within a book or passage that should be removed, such as chapter headings and footnotes. Complete instructions for preparing the text file and accessing the Lexile Analyzer can be found at the following URL: www.lexile.com/analyzer/.

The Lexile Framework for Reading

The Lexile Framework for Reading, developed by educational research organization MetaMetrics, is an indispensable part of any child's literacy development. Lexile

measures take the guesswork out of connecting a child with appropriately challenging reading materials. More information regarding the Lexile Framework for Reading, including a six-minute animated overview video, is also available at the following URL: http://www.lexile.com/about-lexile/lexile-video/.

Regardless of the methods used to access a Lexile measure, educators and parents alike can make use of the Lexile Framework to ensure students are receiving regular practice with both the kinds and volume of complex text they need to grow as readers as they progress toward college and career readiness.

REFERENCES

ACT, Inc. (2006). *Reading between the lines: What ACT reveals about college readiness in reading*. Iowa City, IA: Author.

Allende, I. (2005). *The house of the spirits* (Vol. 281). Random House Digital, Inc.

Chall, J. S., Conard, S., & Harris, S. (1977). *An analysis of textbooks in relation to declining SAT scores*. Princeton, NJ: College Entrance Examination Board.

Coleman, D., & Pimentel, S. (2012). Revised publishers' criteria for the Common Core State Standards in English language arts and literacy, grades 3–12. Retrieved from the Common Core Standards Initiative at www.corestandards.org/assets/Publishers_Criteria_for_3-12.pdf

Hayes, D. P., Wolfer, L. T., Wolfe, M. F. (1996). Sourcebook simplification and its relation to the decline in SAT-Verbal scores. *American Education Research Journal*, 33, 489-508.

Lobel, A. (1971). *Frog and toad together*. Harper.

Steinbeck, J. (1994). *The pearl*. Penguin.

White, E. B., & DiCamillo, K. (1952). *Charlotte's web*. HarperCollins.

Williamson, G. L. (2006). *Aligning the journey with a destination: A model for K-16 reading standards. Durham, NC: MetaMetrics, Inc.*

Featured in Knowledge Quest. Vol 42, No. 3.
January/February 2014 | 1460L

Beyond the Classroom

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D., Senior Vice President Government Relations Anne Schiano and Marketing Manager Elizabeth Lattanzio

New Movement Shaping Education Today

In 1957 the launch of Sputnik sparked a transformative movement in public education in the United States, a movement that dramatically changed educational policy and practice. Now, over fifty-six years later, we are at another transformative moment in education with the almost universal adoption (forty-five states, the District of Columbia, and four territories) of the Common Core State Standards (CCSS). As we move from adoption to implementation of these standards across the country, the climate for educational reform has led to expectations of change that are unprecedented in scope. The challenge before educators today has never been greater. The end goal of the CCSS is best expressed by Education Secretary Arne Duncan's description of the Holy Grail of K–12 education: to graduate all of our students and make sure that they are "college-and career-ready." Let's be clear; this is the most ambitious goal we have ever had in public education. We have never in the history of our country, nor in the history of any other country, graduated every student. In addition to significantly improving the graduation rate, we are also challenged to raise the bar for all students by requiring that they graduate prepared to meet the challenges of postsecondary endeavors—that they are college- and career-ready. As educators and policy makers embark on implementing these new standards, they're seeking ways to effectively maximize the use of existing resources and strengthen partnerships in both the public and private sectors. There is no doubt that school and public libraries and librarians across this country play an essential role in reaching this "Holy Grail."

Librarians as a Driving Force

A substantial body of research clearly shows strong school library programs staffed with certified librarians have a significant impact on student achievement without regard to socioeconomics, teachers' experience levels, or other common correlations to student performance. Research has documented that schools where library resources have been used in support of instruction in literacy, information literacy, and technology skills have witnessed increased levels of motivation in students, as well as higher scores on student achievement measures and higher graduation rates. A study in Illinois involving schools with flexible scheduling and where students have increased access to the school library found that on the Illinois Standards Achievement Test fifth-grade students performed 10 percent better in reading and 11 percent better in writing compared to those with less access (Lance, Rodney, and Hamilton-Pennell 2005). The "School Library Impact Studies" conducted by Library Research Service concluded that libraries have a significant impact on student test scores, even for elementary-level students, and that school libraries play an important role in helping to close the achievement gap.

Equally important is the established role of public libraries in supporting all learners, especially underrepresented populations, including English language learners, socioeconomically disadvantaged students, and students with disabilities. In addition to being a valued community-based resource for recreational reading, these institutions provide students access to resources and technology that may not be available at home, as well as opportunities for after-school and summer learning programs. Experts agree that 21st-century information—and digital-literacy skills are essential for all learners. Public librarians have embraced this need by providing instruction and programs to help young people acquire these skills.

SIX CRITICAL SHIFTS

1 STAIRCASE OF TEXT COMPLEXITY
To read and comprehend text at the level of complexity required for success in college courses and the workplace, students need to be able to read and comprehend increasingly complex text throughout their K–12 experience.

2 BALANCING INFORMATIONAL AND LITERARY TEXT
Students need to read more nonfiction text. The CCSS proposes that about half the reading in elementary school and 75 percent in high school, across the entire day, should be nonfiction.

3 ACADEMIC VOCABULARY
Students must constantly build the transferable vocabulary they need to access grade-level complex texts across content areas.

4 WRITING FROM SOURCES
Students need to be able to write from multiple sources about a single topic, relying on evidence from texts to present careful analyses, well-defined claims, and clear information, as opposed to personal experience and opinion.

5 TEXT-BASED ANSWERS
Students must engage in rich and rigorous evidence-based conversations and writing about text.

6 KNOWLEDGE IN THE DISCIPLINES
Through reading, students need to build knowledge in history, social studies, science, and technical subjects.

While all six of these shifts are rich in their implications for educators, for the purposes of this discussion we want to focus primarily upon the issue of text complexity and the use of the Lexile® Framework for Reading.

With the implementation of the CCSS, libraries should be one of the most valued and trusted resources for teachers, parents, and students. Why are school and public libraries so well positioned to take on this role? A look at the six critical shifts from previous standards to the Common Core State Standards for English Language Arts and Literacy (EngageNY n.d.) brings this connection to light. Other than classroom teachers, no other professionals are so well suited to address these core issues as librarians are.

Text Complexity and the Lexile Framework for Reading

The CCSS recommends a three-part model for evaluating the complexity of a text (Common Core State Standards Initiative 2010). The heart and soul of text complexity is best conceptualized in the graphic in Figure 1. It takes into account the qualitative dimensions (levels of meaning or purpose, structure, language conventionality, and clarity, knowledge demands), quantitative

measures (word length or frequency, sentence length, and text cohesion), and reader and task consideration (motivation, interest, prior knowledge, and experiences). Text complexity is defined in the Common Core State Standards as "the inherent difficulty of reading and comprehending a text combined with consideration of reader and task variables; in the standards, a three-part assessment of text difficulty that pairs qualitative and quantitative measures with reader-task consideration" (Common Core State Standards Initiative 2010, 43).

In applying these considerations for instruction, educators need to keep in mind two very important sentences from Common Core documentation: "The use of qualitative and quantitative measures to assess text complexity is balanced in the Standards' model by the expectation that educators will employ professional judgment to match texts to particular students and tasks. Numerous considerations go into such matching" (Common Core State Standards Initiative 2010, 7).

Figure 1: Three components of assessing text complexity (Common Core State Standards Initiative 2010, 4).

The CCSS Initiative stresses the importance of text complexity if we are to successfully prepare students for reading demands after high school. As stated, "One of the key requirements of the Common Core State Standards for Reading is that all students must be able to comprehend texts of steadily increasing complexity as they progress through school" (Common Core State Standards Initiative 2010, 2). There are, however, two alarming trends in terms of text complexity that must be noted across the P–20 continuum (preschool through postgrad studies). First, over the last fifty years the text complexity of K–12 texts has trended downward (Chall, Conrad, and Harris 1977; Hayes, Wolfer, and Wolf 1996; Williamson 2008). Second, the text complexity of reading required by college courses, careers, and citizenship has held steady or increased over this same time period (Hayes, Wolfer, and Wolfe 1996). This disparity, along with the

finding that reading comprehension breaks down as students read more complex text on college admissions tests such as the ACT and SAT, led to the development of the "staircase of text complexity" by grade level. The Common Core cites Lexile measures as key indicators of text complexity and provides recommended Lexile grade bands for reading development to ensure students are on track for college and career text demands.

Table 1

GRADE BAND	LEXILE BAND
K–1	N/A
2–3	420L – 820L
4–5	740L – 1010L
6–8	925L – 1185L
9–10	1050L – 1335L
11–CCR	1185L – 1385L

In short, the Lexile Framework for Reading is based on this model and is significant because it, unlike any other technology, allows for matching individual readers with specific texts on the same developmental scale. MetaMetrics, developer of the Lexile Framework, has performed research on the importance of the reader-with-text match and the typical reading demands of college and careers. This research contributed to the Common Core as a whole and, more specifically, to the Lexile-based bands in Table 1 above (Common Core State Standards Initiative 2010).

By specifying the Lexile reading demands across the K–12 continuum, a few points become obvious. First and foremost, we need to take a more longitudinal perspective as we prepare all students for the reading demands post high school. Secondly, every grade, every subject, and every education professional is important in growing the literacy skills of our students. Too often we have viewed only a subset of our educators (K–3) as responsible for the reading growth of students. A third point is that we now have a quantitative measure to evaluate whether a student is reading on grade level, a measure that is consistent across districts, states, and our nation.

Since the Lexile Framework is used extensively throughout the nation by teachers and librarians who are currently implementing the CCSS, we thought it would be helpful to provide background on the development, purpose, and utility of Lexile measures.

Background on the Development of the Lexile Framework

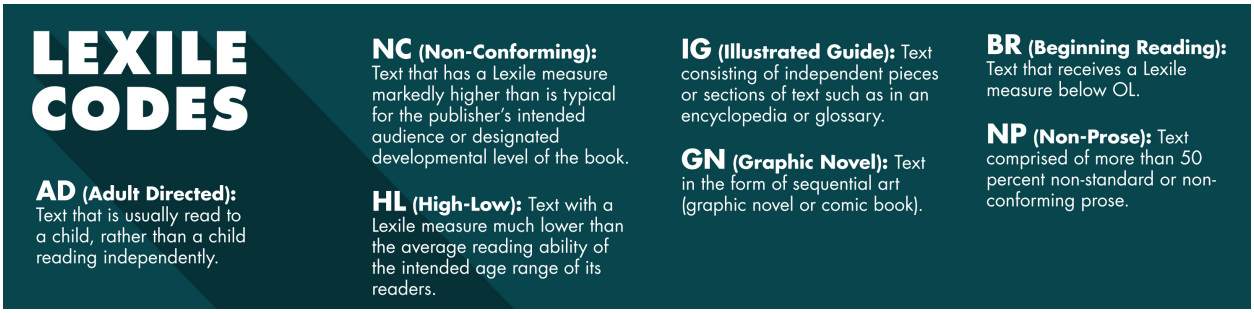
Founded in 1984, MetaMetrics cofounders Jackson Stenner, PhD, and Malbert Smith III, PhD, set out to build an empirical scale that would help match readers to text (placing reader ability and text complexity on the same scale) so that educators could respond appropriately to the tremendous heterogeneity of reading ability within a given class, grade level, or age group. In essence, Stenner and Smith wanted to build a stable and reliable technology tool to support what school and public librarians do daily: find the right book for each student. A second and related goal was to make test scores more meaningful and actionable. The typical scores that were reported from the major reading assessments in use at the time were norm-referenced scores that simply ranked student performance. There is not much a teacher, parent, or librarian can do with a percentile rank, stanine, or normal curve equivalent.

A final consideration in the development of the Lexile Framework was motivated by the need for what the philosophers of science call "unification of measurement." Simply stated, all of the different reading assessments with their various scales had created a psychometric Tower of Babel. Like other constructs, such as temperature, if we could unify the measurement of reading, we could reduce the need for so many different assessments and do more with the assessments that were administered.

The initial research on reading and psychometric theory that culminated in the development of the Lexile Framework was funded over a decade through a series of grants from the National Institute of Child Health and Human Development, part of the National Institutes of Health. While Stenner and Smith were the principal investigators on these grants, scholars from Duke University, University of North Carolina at Chapel Hill, and University of Chicago also played a significant role.

Consistent with the conceptual triangle of text complexity described earlier, the Lexile Framework was created within the transactional nature of students' relationship with text. Unlike some quantitative text-complexity tools that are just "text-centric," the Lexile Framework was created through a conjoint measurement model of both reader and text. In the creation of the Lexile Framework, the importance of the qualitative portion of the triangle was also recognized. These qualitative features are important and include such characteristics as developmental appropriateness, intended audience, purpose, and even factors such as the book's jacket art. When matching readers to books, it is important to pay attention to all of these features.

In trying to address some of these qualitative features, MetaMetrics also provides a series of codes. These codes, while not exhaustive, are intended to capture some of this information that is outside the quantitative measurement of Lexile measures. Below is a list of these codes along with a brief explanation.



Lexile Measures for Both Texts and Readers

The Lexile Framework for Reading is an approach to reading and text measurement. Therefore, two Lexile measures exist: the Lexile text measure and the Lexile reader measure. A book, article, or piece of text receives a Lexile measure by running it through the MetaMetrics Lexile Analyzer, which uses a linguistic algorithm that examines the semantic and syntactic features of the text. The lower the Lexile text measure the easier the text is to read. For example, "Frog and Toad Together" (Arnold Lobel) is a 330L text; "Charlotte's Web" (E. B. White) is a 680L text; "The Pearl" (John Steinbeck) is a 1010L text, and "The House of the Spirits" (Isabel Allende) is a 1280L text. Lexile text measures are rounded to the nearest 10L.

What makes the Lexile Framework for Reading such a powerful technology is the fact that students receive a Lexile reader measure, which is placed on the same developmental scale as the text complexity of reading materials, allowing teachers, parents, and librarians to match a reader with resources at the student's ability level. A Lexile reader measure can range from below 0L for beginning readers to above 1600L for advanced readers. 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, her reader measure is 880L. Higher Lexile measures represent a higher level of reading ability.

Use of Lexile Measures

In addition to the utility of any tool, another important component is the ubiquity of that tool. Lexile measures are used at the school level in all fifty states. More than thirty-five million Lexile measures are reported worldwide each year. In over twenty states, students receive Lexile measures from the state's end-of-grade reading tests. More than

sixty popular reading assessments and programs from major test publishers—including CTB/McGraw-Hill, ETS (formerly known as Educational Testing Service), NWEA (Northwest Evaluation Association), Measured Progress, Pearson, Scholastic, Scantron, and Cambium-Voyager— report Lexile reader measures. In addition to test publishers, all the major text and trade book publishers have elected to use Lexile measures as a way to describe text complexity.

More than one hundred million books, articles, and websites have been measured and received Lexile measures. The nation's largest periodical database services have provided Lexile measures for their newspaper and magazine articles, as well as encyclopedia and reference content. These popular content aggregator services—including EBSCO, World Book, Gale, ProQuest, and Grolier—are easily accessed in libraries through online state databases, and all provide a Lexile search for their content. The country's leading library automation services providers like Follett and Alexandria integrate Lexile measures into school library catalogs.

With the onset of the development and adoption of the CCSS, these numbers are growing dramatically. Lexile measures have become so widely used that there has been a transformation of the term "Lexile" from an adjective to a verb. When inquiring about texts, potential users often want to know if the text has been "Lexiled." The fact that such a large collection of resources has been Lexiled and the measures made available free of charge to teachers and parents enhances the ability of school and public librarians to support differentiated instruction in all content areas at all levels. The abundance of Lexile measures enables librarians to assume more of a leadership role in strengthening their partnership with classroom teachers to enhance instruction.

Kathy Mansfield, library media and textbook consultant from the Kentucky Department of Education, stresses the importance of using the Lexile search features like those in the KY Virtual Library "to help teachers differentiate instruction for their students and to help students find just the right materials that they can best comprehend" (2013). As a school librarian, Mansfield promotes the use of Lexile measures in helping teachers identify valuable resources and materials; the method she recommends for this advocacy is through collaboration that has meaningful goals, such as curriculum development and lesson design.

Kim Shearer, 2012 Kentucky Teacher of the Year and ELA teacher and school librarian at Boone County High School in Kentucky, emphasizes the importance of teachers using "stretch text" in classrooms "to expose students to texts which are more challenging than what they're used to" (2013). Shearer values and actively promotes collaboration between classroom teachers and school librarians in bringing Lexile-measured resources

—books, articles, encyclopedia entries—into classrooms to support student achievement in all content areas.

Common Misconceptions of the Lexile Framework

Many factors affect the relationship between a reader and a book, including its content, the age and interests of the reader, and the design of the actual book. The Lexile measure is an important tool in the book-selection process. However, no tool can replace the professional judgment of a teacher, parent, or librarian in helping students select books for educational and recreational reading. Lexile measures are an indicator of whether the book will be accessible to a reader, not whether it is developmentally appropriate for a reader. Lexile. com provides recommended age-appropriateness ranges for books when this information is available. Lexile.com also offers an age-appropriateness filter for the popular "Find a Book" book search tool. Age information is provided by publishers and Bowker, the world's leading provider of bibliographic information management solutions. In addition to these features, Lexile codes, as described earlier, provide additional information about a book; this information relates to its developmental appropriateness, reading difficulty, and common or intended usage.

The Lexile Framework is not a panacea but a tool that helps match readers to text by measuring reading ability and text complexity on a common scale, the Lexile scale. As all educators, including school librarians, know, there is a tremendous heterogeneity in the reading ability of students within a given class, grade, or age level. The Lexile Framework helps in finding the "sweet spot" in terms of finding text at the appropriate challenge level. What Lev Vygotsky called the "zone of proximal development" (McLeod 2012) is engineered into the matching of reading ability and text complexity through the Lexile Framework.

When a reader's Lexile measure and the Lexile measure of a book match, a targeted reading experience occurs. For example, if a reader has a Lexile measure of 1000L, he or she will be forecasted to comprehend approximately 75 percent of a book with the same Lexile measure (1000L). With the targeted reading experience, a reader will comprehend enough to understand the text but will be exposed to new vocabulary and face some reading challenge. This 75 percent comprehension rate is based on independent reading; if the reader receives assistance, the comprehension rate will increase. Important scaffolds like audio support, visual aids, and preteaching vocabulary can all contribute to a higher comprehension rate or help students read above their Lexile range. Of course, the Lexile Framework should never be used to restrict or pigeonhole a reader.

Librarians' Essential Role in Reading Progress

The ambitious goal of graduating every student ready for college or a career will be reached only if we marshal all of our collective resources in a concerted effort to promote literacy. Huge arrows in the quiver of educational resources are the school and public libraries and their importance in extending learning time. Unfortunately, our school calendar of 180 days a year with 3 months of summer off was built upon an agrarian society that no longer exists.

A large body of research documents the pernicious effects that this three-month break has in terms of reading growth for our students, in particular our low-income students. "Summer loss," "summer slide," and "summer fade" all refer to the reality that students who qualify for free or reduced-price lunch lose ground during the summer at a higher rate than other students. In fact, researchers now argue that up to two-thirds of the achievement gap can be explained by twelve summers of turning off the educational spigot (Alexander, Entwisle, and Olson 2007a, 2007b).

To the credit of leadership within the library community, your profession has long recognized this problem and attempted to address it through participation in the Collaborative Summer Library Program. At MetaMetrics, we have attempted to promote library use over the summer as well. We have teamed with the Council of Chief State School Officers to encourage state departments of education to engage in programs to increase targeted reading over the summer.

As mentioned earlier, over twenty states report Lexile measures from their state-wide reading assessments. These states are now able to take advantage of the link between reading ability and the millions of resources that have a Lexile text measure. When these states send score reports to parents, they are able to directly link test scores through Lexile measures to actionable resources. This capability has enabled over twenty state departments of education to use the state test reports to promote the use of libraries through summer reading programs.

Through our partnerships with state departments of education, MetaMetrics has worked directly with four state libraries. In Colorado, Illinois, Kansas, Kentucky, and South Dakota the state librarians urged school and public librarians, educators, and families to use the Lexile-based "Find a Book" search tool. "Find a Book" enables individuals to build custom reading lists based on a Lexile range and personal interests and then to check the availability of books at the local library. The development of "Find a Book" was motivated by research performed by Harvard 's Dr. James Kim and others, research that has demonstrated the importance of paying attention to both interest and Lexile match when encouraging summer reading (Kim 2005).

The idea behind the Lexile Framework for Reading is simple; if we know how well a student can read and how hard a specific book is to comprehend, we can predict how well that student is likely to understand the book. The Lexile Framework for Reading provides a common developmental scale and measure to match readers with resources and activities that are targeted to readers' ability levels. Lexile measures help educators, school librarians, and families select books, articles, and other materials that provide the right level of challenge for a reader's skills and goals, and to monitor growth in reading ability. For more information on Lexile measures, visit www.Lexile.com.

REFERENCES

Alexander, Karl L., Doris R. Entwisle, and Linda Steffel Olson. 2007a. "LastingConsequences of the Summer Learning Gap." *American Sociological Review* 72 (2):r67-80.

--. 2007b. "Summer Learning and Its Implications: Insights from the Beginning School Study." *New Directions for Youth Development* 114, 11-32.

Chall, Jeanne S., SueS. Conrad, and Susan H. Harris. 1977. *An Analysis of Textbooks in Relation to Declining SAT Scores*. Princeton, NJ: College Entrance Examination Board.

Common Core State Standards Initiative. 2010. *Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects, Appendix A: Research Supporting Key Elements of the Standards; Glossary of Terms*. <www.corestandards.org/assets/Appendix_A.pdf> (accessed October 12, 2013).

EngageNY. n .d. "Pedagogical Shifts Demanded by the Common Core State Standards." <www.engageny.org/sites/default/files/resource/attachments/common-core-shifts.pdf> (accessed October 30, 2013).

Hayes, Donald P., Loreen T. Wolfer, and Michael F. Wolfe. 1996. "Sourcebook Simplification and Its Relation to the Decline in SAT-Verbal Scores."*American Educational Research Journal* 33 (2): 489-508.

Kim,J. S. 2005. "Project READS (Reading Enhances Achievement during Summer): Results from a Randomized Field Trial of a Voluntary Summer Reading Intervention." Paper presented at Princeton University, Education Research Section, November 7.

Lance, Keith Curry, Marcia J. Rodney, and Christine Hamilton-Pennell. 2005. Powerful Libraries Make Powerful Learners: The Illinois Study. Canton, IL: Illinois School Library Media Association .

Mansfield, Kathy. 2013 . "Just Right Text: A Librarian Discusses Assisting Students and Teachers Using the Lexile Framework." <www.youtube.com/watch?v=LrkCipEh5gl> (accessed October 31, 2013).

McLeod, Saul. 2012. "Zone of Proximal Development." <www.youtube.com/watch?v=zEwlDIIW4vg> (accessed October 31, 2013) .

Nelson, Jessica, et al. 2012. *Measure of Text Difficulty: Testing Their PredictiveValue for Grade Levels and Student Performance*. <www.ccsso.org/Documents/2012/Measures%20oiText%20Difficultyfinal.2012.pdf> (accessed October 31, 2013).

Shearer, Kim. 2013. "Master Teacher Discusses the Lexile Framework for Reading,Resources for the Common Core." <www.youtube.com/watch?v=zEwlDIIW4vg> (accessed October 31, 2013).

Smith, Malbert, III. 2012. "Not So Common: Comparing Lexile Measures with the Standards' Other Text Complexity Tools." <http://cdn.lexile.com/m/cms_page_media/135/Not%20So%20Common_1.pdf> (accessed October 31, 2013).

Williamson, Gary L. 2008. "A Text Read ability Continuum for PostsecondaryReadiness. *Journal of Advanced Academics* 19 (4): 602-32.

SECTION 4:
Assessment and Educational Reform

A white paper from The Lexile Framework for Reading | 1360L
April 1, 2004

The Need for Objective Measurement Under the No Child Left Behind Act

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D.

Overview

With the passage of the No Child Left Behind Act, Congress reauthorized the Elementary and Secondary Education Act (ESEA)—the principal federal law affecting education from kindergarten through high school. In amending ESEA (commonly referred to as No Child Left Behind, or NCLB), the new law represents a sweeping overhaul of federal efforts to support elementary and secondary education in the United States.

Some of the major provisions of NCLB include:

- Accountability for results
- Expanded local control and flexibility
- Requiring every child to be on-grade level/proficient by the end of the 2013-2014 school year
- Expanded parental options
- Ensuring every child can read
- Adequate yearly progress (AYP) standards
- Promoting English proficiency

One of the major weaknesses of reading education today is the lack of meaningful measurement systems. The key in the "hard" sciences is unification of measurement. In the case of the measurement of temperature in the 1600s, there were literally dozens of instrument makers with their own scale. However, once a theory of temperature had been developed and accepted, measurement unification was possible. Today, it is inconsequential whether a temperature is taken with a thermometer purchased at CVS or K-Mart — the scale is independent of the manufacturer of the instrument.

As stated in the 2003 report from the Northwest Evaluation Association (NWEA):

> In some states lack of consistency between grade levels poses a serious problem that can be avoided if standards are calibrated. Standards that are not calibrated give students, parents, and educators an inaccurate perception about the child's standing relative to the expected level of performance. Students are reported as proficient in one grade who may not remain proficient in later grades even if they show normal growth.
>
> Assume Xavier, for example, is a 3rd grader living with his family in Yuma, Arizona. Xavier scores at about the 46th percentile in mathematics on his state assessment, which is the minimum score for a rating of proficient, or meets standard. Xavier's

> teacher and parents believe he is performing at a level that is satisfactory relative to grade level standards.
>
> Now let's move forward to 8th grade. Xavier takes the state level assessment again and achieves at the same level of performance, a 46th percentile score, relative to other students. This does not put Xavier anywhere close to the level required to meet the standard. Xavier's parents are alarmed that he is no longer meeting grade level standards and his teachers may come in for criticism because Xavier's performance "slipped." But Xavier's performance didn't slip. Instead he was the victim of a poorly calibrated standard that was too low at 3rd grade to reflect the performance that would be needed by 8th grade (Kingsbury, et. al.).

The Lexile Framework for Reading is a scientific approach to measuring reading ability and text difficulty. All of the major test publishers have linked their norm-referenced tests so that they can report out Lexile measures to students and parents. Each year, millions of students receive a Lexile measure from one of the "instrument" makers. Tens of thousands of trade and textbooks have Lexile measures, and tens of millions of articles have Lexile measures through library database-service providers. MetaMetrics believes that its unifying efforts for the measurement of reading can address the issues in NCLB.

Measurement and Accountability

Many of the assessment issues and concerns that have typically been of interest only within the psychometric community have now become more visible with the high-stakes assessments of NCLB. For example, the fairly standard professional and industry guidance in not relying exclusively on a single test score has become memorialized in standard 13 in Standards for Educational and Psychological Testing (National Council on Measurement in Education, 1999). Simply stated, "in educational settings a decision or characterization that will have a major impact on a student should not be made on a single test score." Recognizing this standard and principal, Section 111 of NCLB requires that the mandated assessments in grades three through eight should employ "multiple, up-to-date measures of student academic achievement" (Koretz, 2003).

A number of researchers have attempted to provide educational practitioners with advice on how to handle the multiple-measure requirements within their accountability models. As these researchers currently point out, the need for multiple measures arises out of the recognition that measurement instruments are not infallible, and thus, should be interpreted within a range of uncertainty that our reliability estimates indicate. While we should view any score from an assessment through the prism of reliability theory,

Baker (2003) rightly points out that multiple measures should be viewed from a validity perspective as well. As Baker states, "despite multiple measures' paternity in reliability arguments, the mother of multiple measures is validity and should exert full sway on the design and continuing evaluation of assessment and accountability systems" (Baker, Linn, Herman and Koretz, 2002).

The Problem of Multiple Measures

In essence, multiple measures within NCLB have provoked both reliability and validity discussions that should force the psychometric community to re-examine the way we measure basic constructs, such as reading ability. As states grapple with how and what they will include as multiple measures, perhaps a thought experiment might be helpful.

Consider for a moment how an assessment system with multiple measures might be designed if NCLB was focused on physical (*health*) outcomes as opposed to cognitive constructs. For example, if NCLB had been enacted to eradicate obesity in the K-12 population, our outcome measures would focus on weight (*pounds, grams*). Since weight and height are positively correlated, we would have to control for height. Perhaps, at the end of this approach, there would be three cut points and four groups: ectomorph, mesomorph, endomorph and below endomorph.

In the reading world, one can think of these three cuts as advanced, proficient and basic readers. Unlike the reading example, however, the cut points for physical assessment would consist of uniform definitions and exchangeable scales upon which all could agree. The physical assessment measures would still be left with the reliability concerns and thus, still in need of multiple measures. Nonetheless, because of the uniform metrics (pounds and grams, inches and meters) the assessment framework would be uncomplicated and the comparisons across states would be straightforward.

Unfortunately, many of the cognitive constructs that are of most interest to educators and policy makers are on nonexchangeable, proprietary scales. Thus, we cannot move easily from one measure to another like we can with constructs in the hard sciences. For example, because of measurement unification in temperature, height and weight, conversions from Celsius to Fahrenheit, inches to meters and pounds to grams are actively used. These scales are exchangeable and data collected from different instruments can be placed on a common scale.

The serendipitous benefit of the high-stakes consequences of NCLB is that it will expose one of the most profound limitations of measurement in the social sciences: the lack of unification of metrics (universal and standard scales). Without universal, exchangeable scales in the social sciences, our assessment systems across states may employ the same

labels (advanced, proficient, basic and below basic), but may vary dramatically in the achievement implied by these labels. An NWEA study proposes that, "States have set proficiency levels using different definitions of 'proficiency.' These standards are now being pressed into service as proficiency indicators under the No Child Left Behind. It is not surprising that the proficiency levels differ, but the degree to which they differ and the potential for misinterpretation are surprising." A third-grade student labeled "proficient" in State A may differ dramatically from a third-grade student in State B as demonstrated in the following table (for a more detailed description of this conundrum, see Education Week, May 2003).

Table 1: Cut scores representing "proficient" or "meets standards" level of performance on 14 state assessments

Reading

Grade 3			Grade 4			Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10		
State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile
SC	205	67	**WY**	214	73	**SC**	220	73	**SC**	221	63	**SC**	227	70	**WY**	232	74	**MT**	224	43	**OR**	236	77
CA	200	51	**SC**	213	70	**CA**	214	54	**CA**	216	46	**WA**	226	67	**SC**	230	68	**IA**	224	43	**WA**	227	51
MN	193	35	**WA**	207	53	**AZ**	210	45	**MT**	211	35	**CA**	221	50	**OR**	227	58	**ID**	221	37	**ID**	224	44
OR	193	35	**CA**	205	46	**OR**	209	42	**ID**	211	35	**MT**	218	43	**CA**	226	54	**CO**	204	9	**MT**	224	44
ID	193	35	**ID**	200	34	**IL**	207	37	**IN**	210	32	**IA**	216	37	**AZ**	224	49				**IA**	223	42
MT	193	35	**MT**	196	26	**MT**	206	35	**IA**	209	30	**ID**	215	35	**IN**	219	35				**CO**	209	15
IL	193	35	**IA**	196	26	**ID**	206	35	**TX**	208	28	**TX**	210	24	**MT**	219	35				**CA**	208	14
IN	192	32	**CO**	191	18	**IA**	205	32	**CO**	197	11	**CO**	206	18	**IA**	219	35						
IA	191	31				**MN**	204	30							**ID**	218	32						
AZ	190	29				**TX**	204	30							**IL**	218	32						
TX	179	13				**CO**	197	18							**MN**	218	32						
CO	179	13													**CO**	206	12						

Mathametics

Grade 3			Grade 4			Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10		
State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile
SC	208	75	**WY**	221	83	**SC**	227	76	**SC**	235	78	**SC**	242	78	**WY**	257	89	**MT**	242	47	**WA**	257	73
CA	204	60	**WA**	218	76	**CA**	225	70	**CA**	230	67	**WA**	242	78	**SC**	251	80	**IA**	241	44	**MT**	247	40
IN	201	50	**SC**	217	74	**AZ**	220	59	**IN**	221	47	**CA**	238	70	**AZ**	248	75	**ID**	240	42	**IA**	247	40
OR	199	46	**CA**	212	59	**OR**	215	46	**ID**	219	42	**ID**	225	44	**CA**	240	59	**CO**	235	32	**OR**	245	33
AZ	199	46	**ID**	205	39	**ID**	213	41	**IA**	218	40	**MT**	224	42	**OR**	235	50				**ID**	242	25
MN	198	42	**IA**	205	39	**MT**	212	38	**MT**	218	40	**IA**	222	38	**ID**	233	46				**CO**	233	14
MT	197	39	**MT**	205	39	**IA**	212	38	**CO**	207	19	**TX**	221	35	**MN**	231	42				**CA**	232	13
IA	197	39				**MN**	210	33				**CO**	216	26	**IN**	231	42						
ID	196	36				**IL**	210	33							**IL**	230	40						
IL	193	29				**TX**	209	31							**MT**	228	36						
					CO	201	15							**IA**	228	36							
														CO	225	31						

• Indiana tests students in the fall. Their cut scores were adjusted to reflect equivalent spring performance.
• Colorado uses the partially proficient level of performance for NCLB reporting. To maintain consistency, NWEA reports the level each state uses for NCLB reporting here.
• The Texas estimate is based on the level for proficient performance that will be implemented in 2005.
(Source: NWEA report, 2003)

Consequently, the real reason that the multiple measures requirement is on such a slippery slope is that our instruments do not have exchangeable scales. Without standard objective scales, like those employed in the hard sciences, educators will be left with less-than-satisfactory methods and very confusing, complicated schemes for reporting such data.

Looking back to the late 1800s, one can find a direct analogue to the reporting dilemmas and confusion confronting the state assessment system. Before the introduction of the railroad system, our country literally had thousands of time zones, as each community would set their clocks to noon when the sun reached its zenith. This meant that every community was on their "local" system. Two neighboring communities might differ significantly, and traveling from one town to another meant that a person had to reset his watch upon arrival. With the introduction of the railroad system it was no longer feasible or practical to have all these "localized" time zones, and movement was begun by a Canadian engineer, Sir Sanford Fleming, to unify (standardize) the measurement of time (Blaise, 2000).

Just as the unification of time was borne out of a Canadian engineer's frustration in trying to figure out what time to pick up his nephew at the railroad station, the complexity and confusion of our state accountability systems may serve as the impetus to agree to standard, universal scales for the constructs of reading and math. A search of Buros Mental Measurement Yearbook (BMMY) for an instrument to measure reading ability or mathematics, yields hundreds of choices that are each on a unique, proprietary scale that is nonexchangeable across instruments.

There are countless constructs — such as temperature, time and weight — that looked like "reading ability" in terms of the number of ways to measure them in their early days of inception. Unification of these constructs was driven by two forces: First, as the underlying scientific theories were developed, there was implicit recognition that the underlying "scales" were important, not the plethora of instruments per se; and secondly, applications forced unification.

The Lexile Framework in Context

Today, at least when it comes to the measurement of reading, our theoretical understanding of the construct is sufficient for unification of scales, and perhaps the application that will accelerate unification is the legislation of NCLB.

A promising candidate in the unification of measurement of reading is The Lexile Framework for Reading. The Lexile Framework is an approach that makes it possible to place readers and text on the same scale (the Lexile scale). The Lexile Framework systemizes two common measurement assumptions:

- Text can be ordered as to difficulty (see Chall, 1996, for a thorough review of readability and the Lexile Framework)
- Readers can be ordered as to reading ability

By placing readers and text on the same scale (the Lexile scale), the difference between a reader's Lexile measure and a text's Lexile measure can be used to forecast the comprehension that the reader will enjoy with the text. One of the realities in U.S. K-12 education that tends to be neglected is student mobility. Every year, a great number of students move from one state to another. To the degree that all states are using tests that have been linked to the Lexile Framework, the students' test scores can travel with them.

With this continuity of the measurement system, schools will not lose important test data. Using the Lexile Framework, each state can establish its own proficiency level benchmarks, but by using the same scale, improvement can be viewed in very concrete terms. States could define their AYP for reading in terms of the Lexile scale and use the Lexile measures from the test results to document the growth.

Currently, every major test publishing company has linked their norm-referenced reading tests to the Lexile scale. Some examples include:

Harcourt Assessment

- Stanford Achievement Tests, Ninth and Tenth Editions (SAT-9 and SAT-10)
  Metropolitan Achievement Test, Eighth Edition (MAT-8)
  Stanford Diagnostic Reading Test, Fourth Edition (SDRT-4)

CTB/McGraw-Hill

- TerraNova Assessment Series (CTBS/5 and CAT/6)

The Riverside Publishing Company

- Gates MacGinitie Reading Tests, Fourth Edition (GMRT-4)
- The Iowa Tests (ITBS and ITED)

Northwest Evaluation Association (NWEA)

- NWEA Achievement Level Tests (print and electronic versions)
- Measures of Academic Progress (MAP)

In addition, Scholastic Reading Inventory, or SRI (Scholastic Inc., 1999), is a standardized assessment designed to measure how well students read literature and expository texts of varying difficulties. SRI began as a targeted-level pencil-and-paper test, but is now available in a computer-adaptive test format.

All of these instruments are able to report out Lexile measures for every student. Because the Lexile scale is a common, supplemental scale that has been linked to the underlying scale of each instrument, we have moved closer to the concept of objectivity in measurement, and hence a unification of a construct. Just as when we measure temperature using a thermometer, we assume that the measure we obtain is not dependent on which thermometer we used. Likewise, in the measurement of reading ability, we assume that the measurement is not dependent on which assessment we used (e.g., the SAT-10, TerraNova, GMRT, etc.). This attribute, termed "general objectivity," is what has historically distinguished measurement in the physical sciences from that in the behavioral sciences.

Many states use tests that are already linked to the Lexile Framework. Any time a student takes one of these tests, he or she can receive a Lexile measure. Having state assessment results also reported on the Lexile scale also enables parents, teachers and students to act on the information. With a Lexile measure, parents can actively support and encourage reading by helping their children select appropriately targeted books (tens of thousands of titles are available at www.Lexile.com). Lexile measures help teachers to differentiate instruction and select textbooks, classroom materials and periodicals that have been measured on the Lexile scale. Since tens of thousands of trade books, thousands of textbooks, and tens of millions of articles have been measured, the annual state test data can now be linked to the classroom text resources.

Students can also benefit from knowing their Lexile measure. Depending on the age of the student, the Lexile Framework can help older students select appropriately targeted research materials for projects. For younger children, a Lexile measure helps to ensure a positive reading experience. Targeted readers report confidence, control of the text, and comprehension and enjoyment of the reading material.

Another benefit is in describing "proficiency levels" with real-life text. What does proficiency at the fourth-grade level mean compared to eighth grade? The Lexile Framework provides a way for teachers, school districts and states to describe proficiency levels in terms of actual text that can be read and comprehended. Using the same label ("proficient") across all grades fails to communicate efficiently with parents. For example, fourth-grade proficiency could be described in terms of the types of text that a reader can comprehend. Concretely, this information could be presented with well-known titles at the different grade levels.

A final benefit is that the Lexile Framework permits school administrators to build longitudinal growth profiles on each student. Since the Lexile scale is a common supplemental metric that cuts across multiple instruments, these growth profiles can be built from multiple data points over many years. For example, if a district is using the

SAT-9 as their norm-referenced test, a statewide test that has been linked to the Lexile Framework (e.g., North Carolina) and any interim assessments that reports Lexile scores (e.g., SRI), then there could be three data points for every year on the student's profile.

Conclusion

As our various sanctioning and professional bodies in the behavioral sciences have rightfully pointed out, we should not make high-stakes decisions from a single administration of a test. This standard has resulted in the necessity of multiple measures. Unless, however, there is general objectivity of measurement of the underlying constructs (i.e., reading, mathematics and science), we are still left with subjectivity and sliding state standards where "proficiency" in one state means something entirely different in another state.

REFERENCES

Baker, Eva L., "Multiple Measures: Toward Tiered Systems," Educational Measurement: Issues and Practice, Vol. 22, No. 2, 2003.

Baker, E. L., R. L. Linn, J. L. Herman, D. Koretz, "From the Directors: Standards for Educational Accountability Systems," The CRESST LINE, Winter 2002.

Blaise, Clark "Time Lord: Sir Sanford Fleming and the Creation of Standard Time," Pantheon Books: New York, Summer 2000.

Kingsbury, G. Gage, Allan Olson, John Cronin, Carl Hauser, Ron Houser, "The State of State Standards: Research Investigating Proficiency Levels in Fourteen States," Northwest Evaluation Association, 2003.

Koretz, Daniel, "Using Multiple Measures to Address Perverse Incentives and Score Inflation," Educational Measurement: Issues and Practice, Vol. 22, No. 2, Summer 2003.

Smith, Malbert III, Ph.D., "Multiple Measures within NCLB: The Need for Objective Measurement," MetaMetrics, Inc., July 2003.

"2001 'No Child Left Behind' Act," The Lexile Times, Vol. 2, Issue 1, August 2002.

MetaMetrics Position Paper | 1250L
January 10, 2010

Best Practices for Next-Generation Assessments

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D.

On Dec. 10–11, 2009, the National Academies' Board on Testing and Assessment (BOTA) held a two-day conference on "Best Practices for State Assessment Systems." The conference's overarching theme was the need to move beyond current assessment systems in favor of more innovative approaches and technologies. Interest in this topic is being fueled by a number of trends, including the Common Core Standards Initiative; RTTT funding; concerns about international benchmarking and assessments; college and career readiness; the belief that NCLB assessments have narrowed the curriculum; and U.S. Education Secretary Duncan's statements that *we can do better than the bubble sheet*. The collective force of these trends and the new funds committed to "innovative assessments" has created a *zeitgeist* for change.

The conference presenters did an outstanding job of painting a picture of the many issues that will need to be addressed, including the costs (one-time versus on-going); psychometric issues; the need for different item types; the tension between curriculum and assessment staff; and the role assessment systems should play in education. A recurring theme was to be mindful of past lessons to avoid the fate that some of our previous innovations have experienced.

Some 30 years ago, legendary assessment guru Oscar K. Buros reflected on the past 50 years of testing (Buros, 1977). His concern about the lack of progress made in the testing field was punctuated in the following: "If you would examine these books and the best of the achievement and intelligence tests then available, you might be surprised that so little progress has been made in the past fifty years—in fact, in some areas we are not doing as well. Except for the tremendous advances in electronic scoring, analysis and reporting of test results, we don't have a great deal to show for fifty years of work. Essentially, achievement tests are being constructed today in the same way they were fifty years ago—the major changes being the use of more sophisticated statistical procedures for doing what we did then—mistakes and all" [p. 10].

OK, no pun intended, what major advances in testing have we witnessed since Buros' critique over thirty years ago? Clearly, the testing field has advanced in many ways—with computer-adaptive testing, IRT models, latent variable theory and robust simulation models—as our computing power has exploded. Yet, many of the basic assumptions about how and when testing should be done, and the inferences we make from them, have changed very little. While testing is done much more frequently than 30, or even 80, years ago, the ultimate question that needs to be answered is "are we getting better information from the tests that we administer?" Unfortunately, I'm not confident that many of us would answer this question in the affirmative.

As we look back at these lessons learned, we need to look even further into the annals of history than Buros did. Consider, for example, some of the past measurement issues

in the hard sciences. A major theme across the sciences is the acceptance and unification of common scales and metrics. In the late 1600s, there were literally dozens of instrument makers (think test publishers) claiming to measure temperature. Each had an eponymous scale that was correlated with the others (similar to the current state of educational measurement constructs). Regardless of whether these instruments used water or mercury, the real breakthrough in measuring temperature was the realization that our scientific understanding of temperature could guide the instrumentation and acceptance of a common scale. This resulted in the acceptance of Daniel Fahrenheit's scale in 1714 and Anders Celsius' scale in 1744. And if the British and French could have sorted out their differences, today we would use a single scale.

When it comes to the measurement of reading, writing and mathematics, we have the same potential for unification as we had with temperature. The underlying measurement issues in education have nothing to do with cosmetic differences in item types, but, rather, in the towering Babel of too many assessments measuring the same construct on different and non-exchangeable scales. Therein lies the confusion on what proficiency means on NAEP versus a specific state test, PIRLS or other testing instrument. Just imagine the confusion in the healthcare industry if we had as many different metrics and scales for measuring temperature as we have for reading.

Unfortunately, today there seems to be more attention and debate on the item types within a test than on the construct being measured—the metric or the utility of the scale. The focus on item type probably is the result of falsely equating the item type with the construct of interest. Often people will look at the item type as illustrative of how the construct should be taught; practice on items like these to improve your reading ability. Item types can vary tremendously across many dimensions—from task complexity to costs to psychometric issues—but they still are just proxies for the measurement of the attribute or construct. We can have many different item types that measure the construct of "reading ability," without privileging any one of these as the most relevant.

The first breakthrough in a new era of meaningful assessments rests upon the idea that reading, writing and mathematics can be measured on vertical (developmental) scales. As E.L. Thorndike stated, "Whatever exists at all, exists in some amount" (Osterlind, 1997). Constructs and measures can be explained using common, vertical scales which facilitate communication and clarity. A primary goal of education is to foster growth, and it is time that we routinely measure individual student growth.

The second breakthrough is predicated upon the premise that just like we can order students from low to high across the vertical scales of reading, writing and mathematics, we also can order instructional content along the same vertical scales.

Placing students and instructional tasks (e.g., readers and books) on the same scale enables us to bridge the curriculum divide. Test and resource publishers can link their products to these underlying scales in such a way that educators can connect assessment with day-to-day instruction in the classroom.

The de minimus psychometric standards of validity and reliability are prerequisites for any assessment. Innovative assessments can adhere to these same psychometric principles while making the following features possible:

1. Assessment and instruction can be usefully blurred, proving that it is possible to "mine the exhaust" of the instructional experience for assessment data as the student engages in instructional tasks. Assessment and instruction are two sides of the academic coin.

2. Computer-adaptive engines can be applied to instructional content, just as they are applied to the test item bank. Both the creation and delivery of content and test items are targeted to the individual.

3. Assessment engines can connect day-to-day progress with year-to-year summative tests by reporting on common developmental scales. Having multiple measurements on a common scale across time and various assessment instruments permits a more reliable and stable estimate of the learner's true ability. We can have more confidence in the inferences that we make about a student's current status and growth trajectory when we rely on multiple measures across the year, as opposed to a single administration of a high-stakes assessment.

4. Test items can be created "on the fly" as students interact with instructional content. Machine-generated test items can be created and discarded as needed throughout the experience of the student. The storehouse of value is in the underlying scale, not in a secure set of test items.

5. Scoring and reporting can be immediate for students, teachers, parents and policymakers. The learning experience and the assessment data mined from the experience is not constrained by calendar, time or location. Delivery is accessible 24/7 via the Web.

6. Perspectives and monitoring can be longitudinal across the developmental lifespan of the student for each construct. As LEA's move from K–12 to P–20 systems of accountability, the importance of optimizing growth for each individual student requires the monitoring and documentation of longitudinal

data. Within these utilities, growth over the lifespan of the learner can be measured and expressed with unparalleled precision (Williamson, 2006).

7. The focus is "student-centric," as opposed to "teacher-centric." A student-centric approach breathes life and reality into the ideal of individual educational plans (IEPs) by paying attention to the critical components of skill acquisition: targeted practice, real-time corrective feedback, intensive practice, distributed practice and self-directed practice.

Innovative assessments built on these metrological principles unlock the realities that are manifested in the hard sciences. We can look at longitudinal data across the developmental span of the learner. By building formative and summative assessments on a common scale, educators, families and policymakers can connect day-to-day learning activities with year-to-year activities. The storehouse of value is no longer locked up in a specific item type but in the value of the scale. Items can be constructed through theory and on-the-fly through artificial intelligence.

An area that needs to be examined in more detail is the mode of assessment. Dirk Mattsen offered the example of assessment delivered through mobile devices and interactive games. Truly innovative assessments will blur the distinction between assessment and instruction. Today, by using existing technologies and frameworks, it is possible to mine the exhaust of an instructional experience and extract assessment data. Individuals who "play" a video game are being assessed in real-time, although they are generally oblivious to this assessment. As they get better, they progress to higher levels within the game. There is no reason why the same instructional and psychometric considerations can not be applied to the teaching and measurement of reading, writing and mathematics.

In conclusion, Stephen Lazar began his presentation with a quote from Ecclesiastes: "What has been will be again, what has been done will be done again; there is nothing new under the sun" (Ecclesiastes 1:9 New International Version). The statement certainly is true when we examine the educational and measurement issues confronting us today. Nearly 90 years ago during the dawn of educational measurement, there was tremendous optimism about the role of assessment. Wilson and Hoke wrote: "The college instructor blames the high school teacher, the high school teacher complains of the grade teacher, each grade teacher above first grade finds fault with the poor work of the teacher in the grade below, and the first grade teacher in turn is chagrined at the shortcomings of the home training. Must this go on indefinitely? Whose opinion shall prevail? Is it not possible to get away from personal opinion to an agreed-upon consensus of opinion? May we not replace the constantly conflicting subjective standards with definitely defined objective standards?" (Wilson & Hoke, 1921).

This time, let's hope we get it right!

REFERENCES

Buros, O.K. (1977). Fifty years in testing: Some reminiscences, criticisms, and suggestions. *Educational Researcher*, 6(7), 9-15.

Holy Bible: New International Version. Authentic Media, 2008.

Osterlind, Steven J. (1997). *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance and Other Formats (Evaluation in Education and Human Services)*. New York: Springer.

Williamson, G. L. (2008). A Text Readability Continuum for Postsecondary Readiness. *Journal of Advanced Academics, vol. 19* (4), 602-632.

Wilson, G.M., & Hoke, K.J. (1921). *How to Measure.* New York: The Macmillan Company.

Next-Generation Assessments

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D.

Today, it is "in vogue" to write, talk and think about the measurement of 21st century skills. Generally, these discussions focus on what should be measured (e.g., critical thinking, digital literacy, cultural awareness), but not necessarily how these constructs should be measured.

More than 30 years ago, legendary assessment guru Oscar K. Buros reflected on the last 50 years of testing (Buros, 1977). His concern about the lack of progress made in the testing field was punctuated in the following statement: "If you would examine these books and the best of the achievement and intelligence tests then available, you might be surprised that so little progress has been made in the past fifty years—in fact, in some areas we are not doing as well. Except for the tremendous advances in electronic scoring, analysis and reporting of test results, we don't have a great deal to show for fifty years of work. Essentially, achievement tests are being constructed today in the same way they were fifty years ago—the major changes being the use of more sophisticated statistical procedures for doing what we did then—mistakes and all" [p. 10].

OK, no pun intended, but what major advances in testing have we witnessed since Buros' critique over thirty years ago? Clearly, the testing field has advanced in many ways—with computer-adaptive testing, IRT models, latent variable theory and robust simulation models—as our computing power has exploded. Yet, many of the basic assumptions about how and when testing should be done, and the inferences we make from them, have changed very little. While testing is done much more frequently than 30, or even 80, years ago, the ultimate question that needs to be answered is "are we getting better information from the tests that we administer?" Unfortunately, I'm not confident that many of us would answer this question in the affirmative.

21st Century Assessments

What should 21st century assessments look like and how should they be utilized? First and foremost, they should assess the three R's: reading, writing and arithmetic. While it has become common practice to disparage the acquisition and measurement of these basic skills, this practice often is misguided. Arguably, the three R's are the foundation of all 21st century skills. And the ability to access and read books is the most efficient and cost-effective pathway to cultural awareness. I'm struck by the continuing influence of our greatest thinkers on current issues. For example, many of our debates today regarding intellectual property remain grounded in the thought-provoking essays of Thomas Macaulay and Thomas Jefferson (J. Boyle, 2009). In most cases, these individuals were never "taught" cultural awareness or critical thinking. Instead, they acquired these skills through learning the three R's.

While these fundamental and critically important skills are the foundation for everything educators attempt to teach in school, in most cases, educators essentially are teaching and measuring these constructs in much the same way as previous generations. Tests still are characterized as falling somewhere on the formative-summative continuum and their results rarely inform instruction in a meaningful way.

The first breakthrough in a new era of meaningful assessments rests upon the idea that reading, writing and mathematics can be measured on vertical/developmental scales. The construction of common vertical scales for these disciplines facilitates communication and clarity. One purpose of education is to foster growth, and it is time that we measure individual student growth. The second breakthrough is predicated upon the premise that just like we can order students from low to high across the vertical scales of reading, writing and mathematics, we also can order instructional content along the same vertical scales.

In the cases of reading, writing and mathematics, these breakthroughs have resulted in the creation of The Lexile Framework for Reading, The Lexile Framework for Writing and The Quantile Framework for Mathematics. Using these frameworks, the educational community has seen many benefits. Test and text publishers can and have linked their products to these underlying scales in such a way that educators now can connect assessment with day-to-day instruction in the classroom.

These breakthroughs also rest upon the assumption that reading, writing and mathematics are skills that can be taught and there are critical instructional components that facilitate their development. Research suggests that a novice develops into an expert through an intricate process that includes the following components (Ericson, 2006; Glaser, 1996; Kellogg, 2006; Shea & Paull, 1996; Wagner & Stanovich, 1996):

- *targeted practice* in which one is engaged in developmentally appropriate activities;
- *real-time corrective feedback* that is based on one's performance;
- *intensive practice* on a daily basis that provides results that monitor current ability;
- *distributed practice* that provides appropriate activities over a long period of time (i.e., 5–10 years),
  which allows for monitoring growth towards expert performance; and
- *self-directed practice* for those times when a coach, mentor or teacher is not available.

Seven Principles

In order to bring more meaning to measurement and to integrate assessment into actionable information for educators, students, and parents, the following seven principles should guide the development of 21st century assessments.

1. **Assessment and instruction should be blurred,** utilizing technology that makes it possible to "mine the exhaust" of the instructional experience for assessment data as the student engages in instructional tasks. Assessment and instruction are two sides of the academic coin.

2. **Computer-adaptive engines must be applied to instructional content, just as they are applied to the test item bank.** Both the creation and delivery of content and test items should be targeted to the individual.

3. **Assessment engines should connect day-to-day progress with year-to-year summative tests by reporting on common developmental scales.** Having multiple measurements on a common scale over time and various assessment instruments permits a more reliable and stable estimate of the learner's true ability. We have more confidence in the inferences that we make about a student's current status and growth trajectory when we rely on multiple measures across the year, as opposed to the once a year administration of a high-stakes assessment.

4. **Test items should be created "on the fly" as students interact with the instructional content.** Test items literally are fungible and appear and disappear as needed throughout the experience of the student. The storehouse of value is in the underlying scale that is being measured, not in a secured set of test items.

5. **Scoring, feedback and reporting must be immediate for students, teachers, parents and policymakers.** The learning experience and the assessment data mined from the experience should not be constrained by calendar, time or location. Delivery should be accessible 24/7 via the Web.

6. **Perspectives and monitoring ought to be longitudinal across the developmental lifespan of the student for each content area.** As we move from K–12 to P–20 systems of accountability, the importance of optimizing growth for each individual student requires the monitoring and documentation of longitudinal data. Within these utilities, growth over the lifespan of the learner can be measured with unparalleled precision (Williamson, 2006).

7. **The focus should be "student-centric," as opposed to "teacher-centric."** A student-centric approach breathes life and reality into the ideal of individual educational plans (IEPs) by paying attention to the critical components of skill acquisition: targeted practice, real-time corrective feedback, intensive practice, distributed practice and self-directed practice.

By adopting these seven principles, it is possible that we can break the mold of conventional assessment models. While much work still needs to be done, at least there is hope that the next fifty years of testing will look significantly different from the last fifty years. If he were alive today, perhaps, Buros would be more optimistic about the future of assessing students' abilities in reading, writing and mathematics.

REFERENCES

Boyle, J. (2009). *The public domain: Enclosing the commons of the mind*. New Haven: Yale University Press.

Buros, O.K. (1977). Fifty years in testing: Some reminiscences, criticisms, and suggestions. *Educational Researcher*, 6(7), 9-15.

Ericsson, K., Charness, N., Feltovich, P., & Hoffman, R. (Eds.). (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press.

Glaser, R. (1996). Changing the agency for learning: Acquiring expert performance. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 303-311). Mahwah, NJ: Lawrence Erlbaum Associates.

Kellogg, R. T. (2006). Professional writing expertise. In K.A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 389-402). New York, NY: Cambridge University Press.

Shea, J. B., & Paull, G. (1996). Capturing expertise in Sports. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 321-335). Mahwah, NJ: Lawrence Erlbaum Associates.

Wagner, R K., & Stanovich, K. E. (1996). Expertise in reading. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 189-325). Mahwah, NJ: Lawrence Erlbaum Associates.

Williamson, G.L. (2008). *A Text Readability Continuum for Postsecondary Readiness,* Volume 19, Issue 4. Summer 2008 (pp. 602-632). Storrs, CT: Journal of Advanced Academics.

MetaMetrics Position Paper | 1290L
October 10, 2009

Education Reform: Making this the 'Best of Times'

by MetaMetrics President and Co-founder Malbert Smith III, Ph.D.

When I read and hear about all of the activities associated with the American Recovery and Reinvestment Act (ARRA) and Race to the Top, my thoughts go back to my ninth grade reading assignment of Charles Dickens' "A Tale of Two Cities:"

> *It was the best of times, it was the worst of times; it was the age of wisdom, it was the age of foolishness... we were all going directly to Heaven, we were all going the other way.*

With many of the issues before us today—including health care, immigration and education—there are those who believe our country is poised to enter a new and enlightened period. On the other hand, there are those who think our nation is on the edge of the precipice toward foolishness and despair. As an educator, I am excited to see education be a significant part of ARRA and am equally excited to hear President Obama challenge our students and parents to pursue educational excellence. I subscribe to the notion that our biggest assets are our children, and those who are entrusted to their education. However, I am concerned that the increased attention and financial resources committed to education could, in the long term, harm our efforts to achieve equity and excellence in education if we do not proceed with some level of caution.

Of primary concern is that we establish the proper framework and expectations of what these additional resources are likely to yield in terms of educational gains. Too often we look for quick fixes, silver bullets and short-term turnarounds when we should seek a more strategic and long-term perspective. We need to be honest and realistic about where we are, the steepness of the climb to the top, and the end goal. That is, how do we define success, how long will it take to achieve that success, and how will we determine if we are making progress along the way?

If we fail to be proactive in describing what is realistic and how we will measure progress, I fear that the dark side of Dickens' quote will be our fate. If we fast forward four years, I can foresee many pundits pointing to our failure to improve and that we spent too much money on public education and that the return on investment was not worth it. I think it is likely that even if we are hugely successful and judicious stewards of the dollars committed to education, the gains we make may not be quick or dramatic enough to silence the critics. We should model Wall Street and issue guidance as to what to expect in the years to come.

How will we judge our success? The basic goals are twofold: first, we need to ensure that all high school students graduate college- and career-ready; and second, we need to improve our performance relative to other countries as measured by international benchmarks. For the first goal, will we look to gains on ACT or College Board's SAT? How soon should we expect to see gains? And how will we measure preparation for college and the workplace? For international benchmarks, will we look at the results

from the 2012 PISA, 2011 PIRLS or 2011 TIMSS? Is success based on whether U.S. students rise in the international rankings? Will we look at the 2012 NAEP results and if we move ten percent more of our fourth, eighth and twelfth graders to proficiency is that enough? Are all of the above dates too soon to make judgments about the efficacy of our investments? Should we wait until the next testing cycle of these assessments (PISA 2015, PIRLS 2016, TIMSS 2015)?

Again, we need to be realistic about the amount of change and over what period of time to expect the requisite change. In his highly acclaimed book, "The Seven Habits of Highly Effective People," Steven Covey points out that we should always begin with the end in mind. In "Good to Great," Jim Collins points out the importance of being honest about your current status as one embarks on interventions and the importance of "The Flywheel," the additive effect of many small initiatives. We, as educators and policy makers, should draw heavily on the advice of these authors.

In addition to considering time frames, effect sizes and gain scores, we must think about how and where to spend the additional financial resources to achieve the "Flywheel." While not intended to be exhaustive, I submit the following suggestions on where we can achieve the most returns on our investments.

1. ***Time.*** We need to rethink the amount of time our students spend in school, and find ways to increase the school day, week and year. Students in other countries spend much more time in school: 193 days in OECD countries and up to 240 days in some Asian countries. If the U.S. Olympic team only could practice 180 days a year but had to compete against countries that practiced 240 days, this disadvantage would be reported on the front page of every American newspaper. We need to have the same level of concern for education as we have for sports in our country. For a more complete discussion and analyses of the issue of time, I encourage you to check out the tremendous work of Chris Gabrieli and Jennifer Davis at the National Center for Time and Learning (NCTL) at www.timeandlearning.org. In 1994, NCTL issued the report, "Prisoners of Time." In 2007, Elena Silva of Education Sector pointed out the need to rethink the use of time in "On the Clock: Rethinking the Way Schools Use Time." KIPP schools have figured out how to get more hours of instruction, and their approach to time management is worth emulating.

2. ***Early education.*** Low-income students face tremendous achievement gaps when they enter school. Making investments where practical and feasible is critical to addressing some of these gaps. ETS has published two monographs, "Parsing the Achievement Gap II" and "The Family: America's Smallest School," that provide excellent summaries of the research on achievement gaps and point to

suggestions on addressing them. In "Parsing the Achievement Gap II," Paul Barton and Richard Coley note that the initial achievement gaps reported in 2003 remain "apparent and alarming" (Barton & Coley, 2009). One of the main findings from the National Math Panel Report was the considerable amount of knowledge of numbers and other aspects of mathematics that most children acquire before entering kindergarten. This knowledge becomes the foundation of mathematics that the students will learn for years to come. However, children from low-income backgrounds enter school with far less knowledge than their middle-income peers, introducing
an achievement gap which progressively widens over time.

3. ***Summer loss.*** Many students lose ground over the summer when they go home to "text-free zones." A number of researchers (Cooper et al., 1996) have demonstrated that while all students grow during the school year, low-income students' reading skills decline during the summer. Many of the same researchers have pointed out that there are a number of low-cost options to combat this epidemic. For example, Kim (2006) demonstrated that if students read eight books that match their reading ability and areas of interest over the summer, they can realize gains similar to those students who attended summer school. One example of a state putting this research into action is North Carolina. Governor Bev Perdue, struggling with staggering budget deficits and a commitment to protect education and other core services, seized the opportunity to encourage families to use free resources available through www.Lexile.com to excite students about summer reading. The Governor sent letters and informational flyers to principals and librarians, and parents received letters and similar information about the "Find a Book" search utility with their child's report card. (Details: www.governor.state.nc.us/NewsItems/PressReleaseDetail.aspx?newsItemID=452.)

4. ***Select interventions and programs that are grounded in the components of deliberate practice.*** The instructional principles required to move a student from a novice to an expert in any discipline are well defined. Whether considering an intervention or instructional program, we need to make sure that the publisher has incorporated sound instructional principles. Research suggests that a novice develops into an expert through an intricate process that includes the following components (Glaser, 1996; Kellogg, 2006; Shea & Paull, 1996; Wagner & Stanovich, 1996):

    - *targeted practice* in which one is engaged in developmentally appropriate activities;

- *real-time corrective feedback* that is based on one's performance;
- *intensive practice* on a daily basis that provides results that monitor current ability;
- *distributed practice* that provides appropriate activities over a long period of time (i.e., 5–10 years), which allows for monitoring growth towards expert performance;
- *self-directed practice* for those times when a coach, mentor or teacher is not available;

5. ***Meaningful and actionable assessment data.*** The first breakthrough in a new era of meaningful assessments rests upon the idea that educational constructs, such as reading, writing and mathematics, can be measured on vertical/developmental scales. The construction of common vertical scales for these disciplines facilitates communication and clarity. One purpose of education is to foster growth, and it is time that we measure individual student growth. The second breakthrough is predicated upon the premise that just like we can order students from low to high across the vertical scales of reading, writing and mathematics, we also can order instructional content along the same vertical scales. In the cases of reading, writing and mathematics, these breakthroughs have resulted in the creation of The Lexile Framework for Reading, The Lexile Framework for Writing and The Quantile Framework for Mathematics. Using these frameworks, the educational community has seen many benefits. Test and text publishers can and have linked their products to these underlying scales in such a way that educators now can connect assessment with day-to-day instruction in the classroom. In his essay on national standards, Chester Finn (March, 2009) suggested that the country could take advantage of candidate scales—the Lexile and Quantile scales—that already have been voluntarily adopted by states and educational publishers.

6. ***Teacher quality and school leadership***. It is well known that the distribution and staffing of highly qualified teachers is anything but random. Our high-poverty districts are the hardest to staff and maintain continuity of leadership. In a recent conversation with a state chief, I discovered that the problem with sending in "school improvement" teams was a lack of adequate accommodations (in some cases, the closest hotel was over 90 miles away). We need to utilize digital solutions to level the playing field. The days of the digital divide must give way to the democratization of opportunity that digital resources provide. In "Disrupting Class: How Disruptive Innovation Will Change the Way the World Learns," Clayton Christensen, Curtis W. Johnson and Michael B. Horn provide a rich background upon which we should all consider deploying technology to

better serve students' needs. Web-based resources and digital learning can help our remote and poverty-stricken districts have access to a world-class 21st century education.

This is a very exciting window of time and opportunity. Let's use it in such a way that we will all look back and say, "it was the best of times."

REFERENCES

Barton, P. E. & Coley, R. J. (2009). *Parsing the achievement gap II*. Princeton, NJ: Educational Testing Service.

Barton, P. E. & Coley, R. J. (2007). *The family: America's smallest school*. Princeton, NJ: Educational Testing Service.

Christensen, C. M. (2008). *Disrupting class: How disruptive innovation will change the way the world learns*. New York: McGraw-Hill.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66, 227-268.

Covey, S. R. (2004). *The 7 habits of highly effective people*. New York: Simon and Schuster.

Ericsson, K.A. (1996). *The Road To Excellence: the Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games, Erlbaum Publishers.*

Ericsson, K.A., Charness, N., Hoffman, R.R., and Feltovich, P.J. (2006), The Cambridge Handbook of Expertise and Expert Performance, Cambridge University Press.

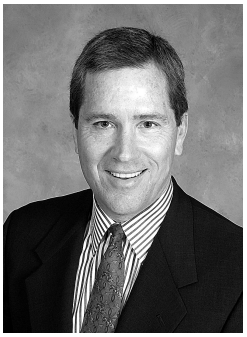
Finn, Jr., Chester E. (2006). *Can we get to national standards, considering the pitfalls? The Education Gadfly,* The Thomas B. *Fordham Institute.*

Kim, J. (2006). Effects of a Voluntary Summer Reading Intervention on Reading Achievement: Results from a Randomized Field Trial. Educational Evaluation and Policy Analysis, Vol. 28, No. 4, 335-355.

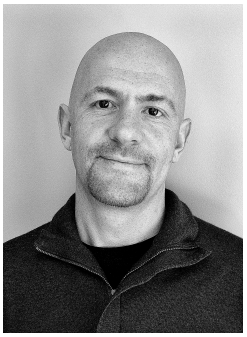
NECTL (1994). Prisoners of Time, Report of the National Education Commission on Time and Learning, http://www.ed.gov/pubs/PrisonersOfTime/index.html

Silva, E. (2007). *On the clock: Rethinking the way schools use time*. Washington, DC: Education Sector.

ABOUT THE AUTHORS



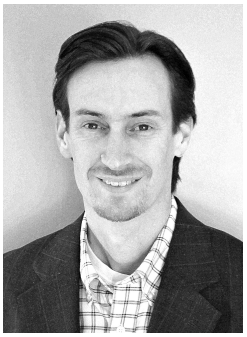
MALBERT SMITH III, PH.D., is president and co-founder of MetaMetrics, where his vision for common metrics and individualized learning has driven extensive partnerships throughout education. Dr. Smith is a senior investigator on a U.S. National Center for Education Statistics research study to examine National Assessment of Educational Progress (NAEP) benchmark scores in relation to university and career readiness. Dr. Smith also serves on the UNC School of Education Foundation Board, the advisory board of Capstone Digital, and is a member of the advisory board for EdSteps, a joint project of The Council of Chief State School Officers and The Bill and Melinda Gates Foundation. Dr. Smith is a research professor at the University of North Carolina at Chapel Hill and has taught graduate seminars in educational research and test design at Duke University and UNC. Widely published, Dr. Smith speaks frequently around the globe on issues related to educational research, measurement and technology.



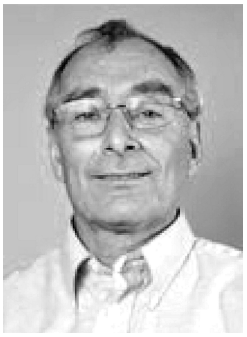
JASON TURNER is the professional development director at MetaMetrics, where he has played a variety of roles. Mr. Turner has managed implementations of and professional development for The Lexile Framework for Reading and The Quantile Framework for Mathematics. Additionally, he has overseen various state and district outreach efforts and developed content and training modules for the Lexile and Quantile Frameworks. Mr. Turner has also worked extensively in the training and implementation of the Lexile and Quantile Frameworks in classrooms across North America.

STEVE LATTANZIO is a research engineer at MetaMetrics where he specializes in the development of new algorithms and data analysis. Steve has worked on algorithms that range from growth modeling and forecasting of student ability to automatic essay scoring engines. Steve received his B.S.E. and M.Sc. in Civil and Environmental Engineering from Duke University where he performed research in the area of control theory and developed a passion for stochastic dynamic processes, optimization and machine learning which he has applied to his work at MetaMetrics.

TODD SANDVIK is senior vice president of global services for MetaMetrics, where he heads the company's international development. Since joining MetaMetrics in 2002, Sandvik has served also as director of operations and vice president of corporate development. In 2010, as executive director of global services, he architected the Engaging English® reading service, which attracted online subscribers from more than 160 countries upon release. Sandvik holds a bachelor's degree in English from the University of Montana and a master's degree in business administration from The Fuqua School of Business at Duke University.

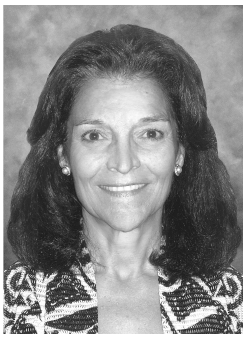


MATT COPELAND is the Lexile curriculum specialist at MetaMetrics. A sixth-generation classroom teacher, Copeland most recently served as a Language Arts and Literacy Consultant for the Kansas State Department of Education where most of his work centered on the implementation of the Common Core State Standards (CCSS). Copeland has presented numerous workshops around the country on the topic of text complexity and the Common Core State Standards. Recognized as a 2006 Distinguished Kansan of the Year in Education, 2005 Milken National Educator and 2003 Kansas Master Teacher, Copeland is a sought-after presenter at local, regional and national conferences. He is the author of Socratic Circles: Fostering Critical and Creative Thinking in Middle and High School. Copeland also served as president and executive board member of the Kansas Association of Teachers of English and co-edited the journal Kansas English. He taught high school English for nine years and is currently a part-time adjunct professor for the University of Phoenix. He holds both a bachelor's degree and a master's degree from the University of Kansas.



DAVID LIBEN is senior content specialist of the Literacy and English Language Arts Team at Student Achievement Partners (SAP). David has over 40 years of experience in education. He has taught elementary, middle school and high school students in public and private schools as well as community college and teacher preparation courses in New York City and Vermont. David and his wife, Meredith, founded two innovative model schools in New York City. In 1985, they joined other teachers to found New York Prep, a junior high school in East Harlem. In 1991, David and Meredith founded the Family Academy. David served as the school's Principal and lead curriculum designer for the first twelve years of the school. Since 2003, David has been consulting with many schools and districts, working with educators on how to teach students to learn independently from complex text. David helped work with the writing

of the CCSS as well as helping to synthesize the research for Appendix A. He is currently the Senior Content Specialist at Student Achievement Partners.



ANNE SCHIANO is the Vice President Government Relations at MetaMetrics. She joined MetaMetrics in 2011 after serving many years as a senior manager responsible for policy development at the New York State Education Department. Most recent]y, she led the Office of Curriculum, Instruction, and Instructional Technology. She has been the recipient of several honors and awards, including NYS Technology Education Association's Appreciation Award, the NYS School Library Association's Advocate Recognition, the NYS Reading Association's Advocate of the Year and Georgetown Washington University's Elliot Fellow Award: Governing in the Global Age.

ELIZABETH LATTANZIO is the marketing Manager at MetaMetrics. She joined MetaMetrics in 2007 and has served many roles in the organization, including product analyst in the Learning Sciences Division and communications director. She shapes messaging for press releases and announcements and provides assistance for state, testing, and book publishing partners' messaging needs. She also works closely with state departments of education to develop and implement their summer reading initiatives. Elizabeth oversees the development of MetaMetrics' white papers, policy briefs, research briefs and position papers. She holds a bachelor's degree from Duke University.

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.

www.MetaMetricsinc.com

www.Lexile.com

www.Quantiles.com

MetaMetrics® is an educational research organization focused on improving education for learners of all ages. For over twenty years, their work has been increasingly recognized for its distinct value in differentiating instruction and personalizing learning. MetaMetrics researchers and psychometricians publish their findings in hopes of better informing the public about important educational issues and topics, in order to potentially influence policy makers.

Visit www.lexile.com/research-and-publications/ for a complete collection of recent publications.

MetaMetrics®
LINKING ASSESSMENT WITH INSTRUCTION