

next-generation assessments

by Malbert Smith III, Ph.D., President, MetaMetrics, Inc.



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, cultural awareness, digital literacy), 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: “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 mis-

guided. 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 Macauley and Thomas Jefferson (J. Boyle, 2009). In most cases, these individuals were never “taught” cultural awareness or critical thinking. Rather, 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 (see Figure 1). 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 (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.

An important question for teachers and policy makers to address is: How can this intricate process be applied in the classroom to promote the development of expertise in reading, writing and mathematics?

Placing students and instructional resources on the same scales, coupled with the research on optimal skill development (Ericsson, Charness, Feltovich, & Hoffman, 2006), has resulted in MetaMetrics®, Inc., developer of the Lexile and Quantile Frameworks, building a suite of next-generation Web utilities: MyReadingWeb, MyWritingWeb and MyMathWeb.

MyReadingWeb, powered by The Lexile Framework for Reading, is a Web-based application that matches students with text based on a student's Lexile measure and text difficulty.

MyReadingWeb enables a student to choose the most appropriate text targeted to his or her reading ability and the topic being taught. One of the application's most important features is its ability to constantly monitor and update a student's Lexile measure after reading an article. Literally, the student is assessed while he or she is reading the text. MyReadingWeb develops multiple-choice comprehension cloze items on the fly as the article is being read. As the student's reading ability "grows," MyReadingWeb "feeds" the student more difficult articles. Educators can be confident about the precision and

FIGURE 1: The Lexile and Quantile Frameworks



The Lexile Framework for Reading provides a common, developmental scale for matching reading ability and text difficulty. Lexile® measures enable educators, parents and students to select targeted materials that can improve reading skills and to monitor reading growth across the curriculum, in the library and at home. Recognized as the most widely adopted reading measure, Lexile measures are used at the school level in various capacities in all 50 states and abroad. Each year, more than 28 million Lexile measures are reported from state and national assessments, classroom assessments and reading programs, representing about half of U.S. students. More than 115,000 books, 80 million articles and 60,000 Web sites have Lexile measures, and the number of resources with Lexile measures continues to grow.



The Lexile Framework for Writing measures student writing ability on the same Lexile scale as reading ability, providing educators with a consistent and straightforward method to monitor student performance in both reading and writing and reinforcing the importance of reading in the development of writing skills. Like Lexile reader measures, students receive Lexile writer measures from standardized tests and other assessment tools. Educators use Lexile writer measures to target instruction and monitor growth in writing ability over time and across the curriculum.



The Quantile Framework for Mathematics provides a common, developmental scale for measuring student mathematics achievement, the difficulty of mathematical skills and concepts, and the materials for teaching mathematics. By placing the curriculum, teaching materials and students on the same scale, Quantile® measures enable educators to describe which mathematical skills and concepts a student is ready to learn and those that will require additional instruction so that students can be matched with resources that meet their learning needs. Parents use Quantile measures to support students' mathematical development by connecting them with targeted mathematics activities at home.

utility of the student's Lexile reader measure—without borrowing instructional time to administer tests that assess reading ability.

MyWritingWeb, powered by The Lexile Framework for Writing, is a unique, Web-based application that integrates the monitoring of student growth in writing ability with writing activities targeted to each student's Lexile writer measure. MyWritingWeb is designed to enhance writer ability, convention ability and device fluency. The activities were created using instructional design principles proven to develop writer ability: 1) immersion of students in professionally authored text; 2) automatic machine scoring; 3) immediate feedback; and 4) direct instruction tailored to each student's unique profile of strengths and areas in need of improvement.

MyMathWeb, powered by the Quantile Framework, blends mathematics instruction with assessment to connect day-to-

day classroom learning with year-to-year growth. The student-centered, Web-based application provides students with targeted practice in each area of mathematics. Each student receives real-time, corrective feedback on intensive daily mathematics activities, including progress monitoring, computational fluency and mathematical games.

Currently, more than 50,000 students use these Web utilities. Unlike traditional assessment methods, the utilities do not rely on separate times for assessment and instruction. Rather, assessment data is mined from the instructional experience. As students' skills improve in one of the constructs, more challenging, but targeted, material is presented. The psychometric principles embedded in the assessment and instructional side allow for the optimal growth of the student.

MyReadingWeb, MyWritingWeb and MyMathWeb share many of the following features and principles:

1. *Assessment and instruction are blurred*, proving that it is possible to “mine the exhaust” of the instructional experience for assessment data as the student engages in instructional tasks.
2. *Computer-adaptive engines are 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 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 single administration of a high-stakes assessment.
4. *Test items are 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 are immediate* for students, teachers, parents and policymakers. The learning experience and the assessment data mined is not constrained

by calendar, time or location. Delivery is accessible 24/7 via the Web.

6. *Perspectives and monitoring are longitudinal* across the developmental lifespan of the student for each skill. 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 is measured and expressed with an 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.

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