

Assessment for Education

What word comes to mind first when you think of the word *school*?

If the word *test* came to mind, it would certainly be understandable. Dozens—maybe even hundreds—of tests are administered to students over the course of their academic career. Included are teacher-made tests, state-mandated tests, psychologist-recommended tests, and assorted other tests. Why so many tests?

JUST THINK . . .

How many tests would you estimate you have taken since you first entered preschool or elementary school?

The Role of Testing and Assessment in Education

Educators are interested in answers to diverse questions as students progress through school. A small sampling of those questions might be:

How well have students learned and mastered the subject matter they were taught?

To what extent are students able to apply what they have learned to novel circumstances and situations?

Which students have demonstrated the ability or skills necessary to move on to the next level of learning?

Which students have demonstrated the ability or skills necessary for independent living?

What are the challenges or obstacles that are preventing an individual student from meeting educational objectives, and how can those obstacles best be overcome?

How effective are teachers in assisting students to master specific curriculum goals?

Do passing test scores on a curriculum-specific test genuinely reflect the fact that the testtakers have mastered the curriculum?

Do failing test scores on a curriculum-specific test really reflect the fact that the testtakers have not mastered the content of the curriculum?

In recent years, many such questions have been raised with regard to the states' kindergarten-through-12th-grade (K-through-12) educational systems. Some observers cite generally disappointing answers to such questions. Blame for such dubious educational outcomes has been placed by many on federal legislation such as *No Child Left Behind* (NCLB) legislation. Perhaps the most vocal outcries have been reserved not for NCLB as

whole, but certain provisions of NCLB, such as its insistence that educational progress be gauged by standardized tests.

The Case for and Against Educational Testing in the Schools

For all too many parents, teachers, and opinionated third-parties, the “root of all evil” in contemporary education is the standardized test; the blame for any failure of the educational system lies not with the student, the teacher, or the school administration, but with standardized testing (see, for example, García & Thornton, 2015). The logic behind such arguments, questionable as they are, is rooted in the fact that the NCLB legislation mandated improvement in reading and math scores on standardized tests. If an individual state did not meet federal guidelines in this regard, then the result could be the imposition of penalties, such as the withholding of federal funds for education. In the view of some then, the net result of NCLB was undue pressure on teachers to make certain that their students performed satisfactorily on standardized tests. This obligation was seen as forcing teachers to spend valuable classroom time “teaching to the test.” The argument of many anti-test advocates could essentially be summed-up as “If there was no pressure to raise students’ performance to some federally prescribed level, then teachers would be free to teach in ways designed to promote better, more permanent educational outcomes.”

Counter to the arguments of those who rail against standardized tests as the culprit in education is another, more moderate view. Simply stated, this alternative, time-tested position is that standardized tools of assessment, including tests, are not only a desirable element of an educational system, but an absolutely necessary one at that. According to this latter position, tests may serve a variety of critically important needs. For example, standardized educational tests are used for screening purposes. Tests routinely used for screening purposes can alert educators to students who may be at risk for negative, education-related outcomes—outcomes that may be preventable with early, effective intervention. In addition, data from standardized tests may be mined for important diagnostic findings. Diagnostic data may be used to identify areas of weakness that require remediation or other educational intervention. Similarly, diagnostic data may be useful in identifying areas in which a student excels. Knowledge of where a student excels is essential if education resources are to be properly allocated, and not squandered with efforts to teach students what they already know. From the perspective of the student, knowledge derived from diagnostic educational testing is necessary if students are to be increasingly challenged by new and intellectually stimulating subject matter.

In addition to screening and diagnostic purposes, standardized tests are indispensable for purposes of comparison. Data from such tests help educators understand and gauge the rate of progress of a single student, dozens of students, hundreds of students, or thousands of students—all using the same tool of measurement (rather than one idiosyncratic to the preference of an individual teacher or even a particular school district). Students learn at different paces and in different ways, and it is through such comparisons of progress that educators can begin to identify what teaching methods work best for which individual.

So, although there are those who stubbornly cling to what Gerson (2012) called “an ideological opposition to testing as the enemy of educational creativity,” standardized testing programs, when judiciously administered, would seem to hold out the promise of more benefit than harm. The logic of those in the education community who focus on testing as the downfall of education seems, at best, flawed. Commenting on teachers who profess such a position, Gerson (2012) reflected, “They love the intangible joys of the profession, without the inconvenience of demonstrating that their work has any effect.”

To be fair, when certain negative conditions relevant to the ecology of testing exist, sentiments against tests and testing may well have merit. For example, if a standardized testing program was designed exclusively for alignment with a controversial teaching curriculum, then the controversy surrounding the teaching curriculum could reasonably be expected to “rub off” on (or generalize to) the tests used in conjunction with that curriculum. Also, if individual test items in a standardized test are constructed in a controversial way (such as being widely perceived as age- or grade-inappropriate in terms of content), then—no matter what educational objectives are being served—the standardized test or entire testing program could well be the object of legitimate questions, concerns, and criticism. In fact, each of these examples typify why a relatively new, nationwide program for kindergarten-through-12-grade students is so controversial. A brief explanation of what this program is, as well as why it is so controversial, follows.

The Common Core State Standards

NCLB required, among other mandates, that all children within a state perform at grade-level in reading and math by the end of 2014. However, each of the states set their own definition of what constituted proficiency at each grade level. Given this lack of uniformity in proficiency standards among the states, it would be entirely possible for a student in one state to be deemed proficient in a particular subject (or, performing at grade-level according to that state’s standards), while having knowledge and skills that would not be deemed proficient according to another state’s standards. Such discrepancies between state standards, as well as myriad other considerations (such as a general lack of sufficient preparation for college-level work) compelled some to consider ways to bring uniformity to state education curriculums for public schools. The result was the creation of a comprehensive set of standards that affects the everyday classroom life of some 50 million K-through-12 students. These standards, which essentially set objectives for what students should know by the end of each school year, are packaged along with a computer-assisted testing program designed to ensure conformity in teachers’ teaching as well as students’ learning. The program, which currently sets standards for learning in English and math (with standards for more subject areas in development), is called the *Common Core State Standards* (CCSS).

To complement the release of CCSS, in March of 2010, the United States Department of Education offered groups of 15 or more states (referred to as *consortia*) close to \$400 million dollars in grant money to develop new K-through-12 tests to be aligned with CCSS (Doorey, 2012/2013). By September of 2010, the funds had been awarded to two consortia, one known as “PARCC” (the Partnership for the Assessment of Readiness for College and Careers), and the other known as “Smarter Balanced” (the Smarter Balanced Assessment Consortium).¹ PARCC and Smarter Balanced were assigned the task of developing a comprehensive assessment system that was qualitatively different in many ways than anything that had preceded it. Doorey (2012/2013) put in perspective some of the challenges inherent in that undertaking:

K–12 assessment is at the beginning of a sea change. Many of the competencies now considered essential for success in college and the workplace are complex and difficult to measure. The

1. Why did the U.S. Department of Education award grants to *two* consortia instead of just one? Although it cannot be stated with certainty, some believe it was “to allay fears of a ‘national assessment’ and of usurpation of local control over the curriculum” (Doorey, 2012/2013, p. 28).

assessment consortia, caught in the midst of this change, must navigate a series of tough challenges, choices, and trade-offs.

To meet the expanded policy purposes and anticipated uses of the data, these systems of assessments must go far beyond simply determining whether a student has met grade-level standards. They must measure individual growth for all students and provide more accurate information concerning students who perform well above or well below the standards. They must yield fine-grain information that can inform instructional and programmatic decisions. And they must be able to evolve over time to reflect changes in the skills needed in our global marketplace and to incorporate advances in technology, cognitive science, and measurement. (pp. 33–34)

You may have heard of *Common Core* when you sat down to watch your local news and saw a group of teachers and parents protesting against it. So why are so many people so up-in-arms against a program that purports to bring uniformly high standards and a much improved testing program to K-through-12 education? More insight on this important question is as close as this chapter's *Everyday Psychometrics*.

From the perspective of students, perhaps the most obvious reason testing occurs is to ensure that knowledge being shared by the teacher (in the classroom, on the Internet, or however and wherever teaching is taking place) has indeed been received and learned by the student. But educators are also interested in helping students better their learning skills. Toward that end, educators may administer tests designed to pinpoint possible areas of learning difficulty. There are other circumstances wherein educators have a compelling interest in knowing the extent to which their students—or prospective students—are prepared to learn more advanced material. In such cases, tests variously referred to as “readiness” or “aptitude” tests may be administered. Yet another reason testing occurs in classrooms is very simple: the testing is required by law. And speaking of testing in the schools and the law, let's consider the (legislation-driven) exploding popularity of RtI. That upper case *R*, lower case *t*, and upper case *I* stands for . . .

Response to Intervention (RtI)

Background In the mid-1970s, the federal mandate to identify and assist children with learning problems defined a learning disability as a “severe discrepancy between achievement and intellectual ability” (*Procedures for Evaluating Specific Learning Disabilities*, 1977, p. 65083). For decades thereafter, a specific learning disability (SLD) was diagnosed if a significant discrepancy existed between the child's measured intellectual ability (usually on an intelligence test) and the level of achievement that could reasonably be expected from the child in one or more areas (including oral expression, listening comprehension, written expression, basic reading skills, reading comprehension, mathematics calculation, and mathematics reasoning). However, all of that changed with a new definition of SLD and new diagnostic guidelines.

As defined in 2007 by Public Law 108-147, a **specific learning disability** is “a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which disorder may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations.” Also, as reauthorized in 2004 and enacted into law in 2006, the Individuals with Disabilities Education Act (IDEA) no longer mandated that state-adopted criteria for defining SLD be made on the basis of a severe discrepancy between intellectual ability and achievement. Rather, it required states to allow “the use of a process based on the child's

JUST THINK . . .

Why do you think there has been so much disagreement and controversy over the definition of the term *learning disability*? Why is it so important to “get it right”?

The Common Core Controversy

A visit to the Common Core State Standards (CCSS) official website at corestandards.org can leave one inspired with regard to the potential of CCSS.

There visitors are advised, for example, that CCSS “focuses on developing the critical-thinking, problem-solving, and analytical skills students will need to be successful” and that the standards “establish clear, consistent guidelines for what every student should know and be able to do in math and English language arts from kindergarten through 12th grade.” So what is so controversial about CCSS?

The *Common Core State Standards* (CCSS) are controversial for a number of reasons. For starters, to characterize CCSS as “a set of standards” is, at one and the same time, both factual and misleading. For sure, CCSS is a list of standards by grade that must be met by the end of the school year. However, CCSS is so much more than that—to the extent that referring to CCSS as “standards” seems a misnomer. Viewing CCSS in broad perspective, it is a comprehensive, K-through-12 program for preparing students for college and work life—a program that includes as an integral component, extensive testing to make sure that CCSS objectives are met. In addition, there is also an extensive record-keeping component to CCSS—this for the purpose of evaluating individual and group outcomes, as well as for making future adjustments to the program.

Beyond the fact that CCSS is really a program that was nominally presented as a list of standards, CCSS represents the expression of a singular vision for what K-through-12 education should be. As Toscano (2013) reflected, the standards “push schools toward a particular understanding of ‘college and career readiness,’ and insist that the curriculum embody that understanding.” The implementation of the CCSS vision comes through the combination of standards and test items, jointly designed to maintain uniformity in terms of what students need to know, which skills students need to learn, and which specific methods students should use to approach math and other academic challenges. In essence, the CCSS standards set forth a vision for every grade level that culminates in a graduate who embodies what CCSS envisions a high school graduate should embody. This vision is “enforced” by, and inextricably linked and aligned to, standardized tests. At various points along the way, tests gauge student progress in acquiring CCSS-prescribed knowledge, skills, and problem-solving approaches. In addition, standardized testing becomes a primary vehicle by

which students (and teachers alike) are rewarded and penalized.

Practically speaking, the CCSS program would seem to leave little room in it for non-CCSS activities for the purpose of experiencing the sheer fun of learning or discovery (such as independent study on a topic of the student’s choice). Similarly, the CCSS program would seem to leave little room for teachers to evaluate student learning in ways not specifically prescribed by CCSS (such as organizing and conducting a mock debate on a current event). So, although CCSS is not a national, K-through-12 curriculum, it would certainly seem to have elements resembling one. One could quip that CCSS presents teachers and school districts with a *go figure curriculum*; that is, teachers are told what students need to know and what students will be tested on, but are left to *go figure* how to teach the required subject matter.

Another source of controversy concerns how this program came to be. According to the Common Core website, the program was originally conceived by a number of representatives from different states, including governors and heads of state school systems. Ostensibly supporting that claim is the fact that authorship of the program is nominally attributed to two organizations: The National Governors’ Association Center for Best Practices and the Council of Chief State School Officers (2010). However, critics and skeptics have long questioned whether the whole story of how CCSS came to be has ever come to light, especially since the program was so rapidly adopted by so many states, sight unseen (Toscano, 2013). It is common knowledge that state buy-ins were encouraged, if not compelled, by the incentive of federal funds. But what are the facts regarding the origination of CCSS? For example, to what extent was Microsoft founder Bill Gates involved in the establishment of CCSS, and was the motivation for the involvement purely philanthropic? Through their charitable organization, The Bill and Melinda Gates Foundation, the Gates donated some 2.5 billion dollars to develop CCSS. Commenting on that fact, Thorner (2014) noted, “there is no way Common Core could have been brought into the nation’s schools given that it was the product of a small group of activists supported by billionaire Bill Gates.” Bill Gates has envisioned CCSS being administered not only to millions of students in the United States, but to many more millions of students throughout

(continued)

The Common Core Controversy (*continued*)

the world ("Bill and Melinda Gates," 2015). Given the fact that participation in CCSS requires school systems to purchase and perpetually update expensive computer systems and software, the question of whether Gates' contribution is more philanthropy or an investment has been raised (Beck, 2013; "The Case Against Common Core," 2015).

Another reason CCSS is controversial is due to the *en masse*, blind buy-in of so many states with absolutely no evidence that the program works. It seems fair to surmise that the participating states were primarily induced to participate as a result of federal funding incentives for doing so—over four billion dollars in grants was disbursed to the states in return for their participation in CCSS. Thus, although the federal government did not create CCSS, it would seem to own sole responsibility for the states' rapid adoption of the program. By the way, that fact is, in itself, controversial because responsibility for education is Constitutionally left to the states. Holding out huge sums of federal grant money in exchange for the institution of a nationwide set of uniform educational standards has been viewed by some as a quasi-legal way for the federal government to assume greater responsibility and control for education in the states that accept the deal.

Also controversial are the educational credentials of the folks who authored the standards. The Common Core website tells us that the program was developed by teachers and educators from around the country. However, critics of CCSS have disputed that assertion ("The Case Against Common Core," 2015). According to Thornton, CCSS was "written by a small group of individuals and then copyrighted by two Washington lobbyist groups" (the lobbyist groups being The National Governors' Association Center for Best Practices and the Council of Chief State School Officers). Two prominent members of the Common Core Validation Review Panel were James Milgram, professor of mathematics at Stanford University (the panel's math expert), and Sandra Stotsky, Professor Emerita at the University of Arkansas and former Senior Associate Commissioner of the Massachusetts Department of Education (the panel's English expert). According to Thornton (2014), both of these experts "refused to give Common Core Math and English standards, respectively, a good recommendation... Both have gone on to testify with a warning voice to state legislatures and school boards about the inadequacy of the standards."

It is also a matter of controversy, if not outright concern, that the CCSS contains recommendations for educational

practices that are not widely supported by the scholarly literature. One such practice is **cold reading**; that is, reading without the benefit of background information or context. So, for example, a student might be asked to study the Gettysburg Address by reading it "cold" (i.e., without the teacher setting the reading material in context or providing any background information on it at all). Commenting on this practice, Anderson (2014) pointed out, "Obviously additional information would help the student become more proficient in the subject, yet Common Core disallows it. The standards cite no research supporting such a practice."

The very content of the CCSS has drawn a great deal of criticism. Part of the problem here is that some of the standards appear to be age- and grade-inappropriate. For example, according to one of the CCSS standards for fifth-grade English, students must be able to demonstrate an understanding of the influence of the perspective of a story's narrator. Similarly, some math standards are arguably ambitious. Enter the search term "inappropriate Common Core math standard" on Google (or any comparable computer search engine), and the chances are good the screen will be populated by dozens of entries. One such standard for first-graders is "Understand subtraction as an unknown-addend problem" (for context and an explanation, see Strauss, 2013).

Along with a questioning of the content of the CCSS has come a questioning of the test items used to evaluate teaching outcomes. In some cases, test items have been heavily criticized for being age-inappropriate, or otherwise inappropriate in item content. For example, some of the content in some of the reading passages used to evaluate accomplishment in English has been criticized as being age- and grade-inappropriate. What follows is a sample of one such reading comprehension item represented to have been taken from fourth-grade Common Core classwork (Hope, 2014), and posted online by the student's outraged parent:

Ruby sat on the bed she shared with her husband holding a hairclip. There was something mysterious and powerful about the cheaply manufactured neon clip that she was fondling suspiciously. She didn't recognize the hairclip. It was too big to be their daughter's, and Ruby was sure that it wasn't hers. She hadn't had friends over in weeks but there was this hairclip, little and green with a few long black hair strands caught in it. Ruby ran her fingers through her own blonde hair. She had just been vacuuming when she

noticed this small, bright green object under the bed. Now their life would never be the same. She would wait here until Mike returned home.

Why is Ruby so affected by the hairclip?

How has the hairclip affected Ruby's relationship?

Finally, there is the question of the ultimate benefit of subjecting millions of students to an essentially untested program. In this context, critics have raised questions like, "Would it not be more reasonable to model such a new, large-scale, nationwide undertaking after the teaching processes that are known to work?" Why not model the new, nationwide program after a state like Massachusetts, for example?* Why

*In a Wall Street Journal survey, Massachusetts was found to have the best K-through-12 education program (Frohlich & Sauter, 2014).

plunge millions of students and teachers into what is essentially a large-scale educational experiment with an uncertain outcome hanging in the balance?

In sum, since its rather hasty, sight-unseen adoption by so many states, CCSS has been the subject of many important questions, some still in search of satisfactory answers. Skeptics have raised questions having to do with the role of financial profit in the initial development and promulgation of CCSS. Academics have raised questions having to do with the efficacy of the program. Privacy advocates have raised questions about the extensive record-keeping and archiving of educational data that is integral to the administration of CCSS. Until such time that widely acceptable answers to these and related questions have been given, controversy will continue to surround CCSS.

response to scientific, research-based **intervention**" (emphasis added). Although academics had been writing about "response to intervention" in the interim period (see, for example, Dwairy, 2004), all of a sudden the rush to understand and apply this model was on. By 2008 most states had in place legislative or regulatory policies that changed their existing special educations to require RtI (Zirkel & Krohn, 2008).

The RtI model Based on the definition presented on the federally funded website of the National Center on Response to Intervention (2011, p. 2), we may define the **response to intervention model** as a multilevel prevention framework applied in educational settings that is designed to maximize student achievement through the use of data that identifies students at risk for poor learning outcomes combined with evidence-based intervention and teaching that is adjusted on the basis of student responsiveness. A simpler description of this model of teaching and assessment is: (a) Teachers provide evidence-based instruction, (b) student learning of that instruction is regularly evaluated, (c) intervention, if required, occurs in some form of appropriate adjustment in the instruction, (d) reevaluation of learning takes place, and (e) intervention and reassessment occur as necessary.

The model is *multilevel* because there are at least three levels of intervention (or teaching). The first level is the classroom environment wherein all students are being taught whatever it is that the teacher is teaching. The second level of intervention is one in which a small group of learners who have failed to make adequate progress in the classroom have been segregated for special teaching. The third level of intervention is individually tailored and administered instruction for students who have failed to respond to the second level of intervention.

By providing intervention (teaching or remedial instruction, as the case may be) appropriate to the level of the student's needs, the objective of RtI is to accelerate the learning process for all students. In addition, RtI doubles as a process in place that will identify students with learning disabilities. In this sense, RtI is seen by many as superior to the more traditional,