DEEPER LEARNING RESEARCH SERIES

A NEW ERA FOR EDUCATIONAL ASSESSMENT

By David T. Conley
October 2014

EDITORS' INTRODUCTION TO
THE DEEPER LEARNING RESEARCH SERIES

In 2010, Jobs for the Future—with support from the Nellie Mae Education Foundation—launched the Students at the Center initiative, an effort to identify, synthesize, and share research findings on effective approaches to teaching and learning at the high school level.

The initiative began by commissioning a series of white papers on key topics in secondary schooling, such as student motivation and engagement, cognitive development, classroom assessment, educational technology, and mathematics and literacy instruction.

Together, these reports—collected in the edited volume *Anytime, Anywhere: Student-Centered Learning for Schools and Teachers*, published by Harvard Education Press in 2013—make a compelling case for what we call "student-centered" practices in the nation's high schools. Ours is not a prescriptive agenda; we don't claim that all classrooms must conform to a particular educational model. But we do argue, and the evidence strongly suggests, that most, if not all, students benefit when given ample opportunities to

> Participate in ambitious and rigorous instruction tailored to their individual needs and interests

> Advance to the next level, course, or grade based on demonstrations of their skills and content knowledge

> Learn outside of the school and the typical school day

> Take an active role in defining their own educational pathways

Students at the Center will continue to gather the latest research and synthesize key findings related to student engagement and agency, competency education, and other critical topics. Also, we have developed—and will soon make available at www.studentsatthecenter.org—a wealth of free, high-quality tools and resources designed to help educators implement student-centered practices in their classrooms, schools, and districts.

Further, and thanks to the generous support of The William and Flora Hewlett Foundation, Students at the Center is now expanding its portfolio to include a second, complementary strand of work.

With the present paper, we introduce a new set of commissioned reports—the Deeper Learning Research Series—which aims not only to describe best practices in the nation's high schools but also to provoke much-needed debate about those schools' purposes and priorities.

In education circles, it is fast becoming commonplace to argue that in 21st century America, "college and career readiness" (and "civic readiness," some add) must be the goal for each and every student. But as David Conley explains in these pages, a large and growing body of empirical research shows that we are only just beginning to understand what "readiness" really means.

In fact, the most familiar measures of readiness—such as grades and test scores—tend to do a very poor job of predicting how individuals will fare in their lives after high school. While one's command of academic skills and content certainly matters, so too does one's ability to communicate effectively, to collaborate on projects, to solve complex problems, to persevere in the face of challenges, and to monitor and direct one's own learning—in short, the various kinds of knowledge and skills that have been grouped together under the banner of "deeper learning."

What does all of this mean for the future of secondary education? If "readiness" requires such ambitious and multi-dimensional kinds of teaching and learning, then what will it take to help students become genuinely prepared for college, careers, and civic life?

Over the coming months, many of the nation's leading education researchers will offer their perspectives on the specific kinds of policies and practices that will be needed in order to provide every student with the opportunity to learn deeply.

We are delighted to share this first installment in our new Deeper Learning Research Series, and we look forward to the conversations that all of these papers will provoke.

To download the papers, introductory essay, executive summaries, and additional resources, please visit the project website: www.studentsatthecenter.org/topics.

Rafael Heller, Rebecca E. Wolfe, Adria Steinberg

Jobs for the Future

Introducing the Deeper Learning Research Series

Published by Jobs for the Future | New and forthcoming titles, 2014-15

A New Era for Educational Assessment
David T. Conley, EdImagine Strategy Group and the University of Oregon

Digital Tools for Deeper Learning
Chris Dede, Harvard Graduate School of Education

The Meaning of Work in Life and Learning
Nancy Hoffman, Jobs for the Future

Equal Opportunity for Deeper Learning
Linda Darling-Hammond, Stanford University & Pedro Noguera, Teachers College, Columbia University

English Language Learners and Deeper Learning
Guadaulpe Valdés, Stanford University & Bernard Gifford, University of California at Berkeley

Deeper Learning for Students with Disabilities
Louis Danielson, American Institutes for Research & Sharon Vaughn, University of Texas

Deeper Learning for Engaged Citizenship
Kei Kawashima-Ginsberg & Peter Levine, Tufts University

Ambitious Instruction
Magdalene Lampert, Boston Residency for Teachers and the University of Michigan

District Leadership for Deeper Learning
Meredith Honig & Lydia Rainey, University of Washington

Profiles of Deeper Learning
Rafael Heller & Rebecca E. Wolfe, Jobs for the Future

Reflections on the Deeper Learning Research Series
Jal Mehta & Sarah Fine, Harvard Graduate School of Education

JOBS FOR THE FUTURE

Jobs for the Future works with our partners to design and drive the adoption of education and career pathways leading from college readiness to career advancement for those struggling to succeed in today's economy. We work to achieve the promise of education and economic mobility in America for everyone, ensuring that all low-income, underprepared young people and workers have the skills and credentials needed to succeed in our economy. Our innovative, scalable approaches and models catalyze change in education and workforce delivery systems.

WWW.JFF.ORG

Students at the Center—a Jobs for the Future initiative—synthesizes and adapts for practice current research on key components of student-centered approaches to learning that lead to deeper learning outcomes. Our goal is to strengthen the ability of practitioners and policymakers to engage each student in acquiring the skills, knowledge, and expertise needed for success in college, career, and civic life. This Jobs for the Future project is supported generously by funds from the Nellie Mae Education Foundation and The William and Flora Hewlett Foundation.

WWW.STUDENTSATTHECENTER.ORG

ACKNOWLEDGEMENTS

ABOUT THE AUTHOR

Dr. David T. Conley is the founder, chief executive officer, and chief strategy officer of the Educational Policy Improvement Center (EPIC). Dr. Conley also serves as President of EdImagine Strategy Group, Professor of Educational Policy and Leadership, and founder and director of the Center for Educational Policy Research at the University of Oregon. Dr. Conley serves on numerous technical and advisory panels, consults with national and international educational agencies, and is a frequent speaker at meetings of education professionals and policymakers. He is the author of several books and numerous articles on the topic of college and career readiness, including his newest book *Getting Ready for College, Careers, and the Common Core: What Every Educator Needs to Know* (Jossey-Bass 2013).

Dr. Conley developed and implemented the nation's first proficiency-based college admission system, PASS, from 1993-1999. It was used by the Oregon University System, then field tested at 52 Oregon high schools, and continues to be used by students as a means to demonstrate college readiness. In 2003, Dr. Conley completed Standards for Success, a groundbreaking three-year research project to identify the knowledge and skills necessary for college readiness.

Dr. Conley received a BA with honors in Social Sciences from the University of California, Berkeley. He earned his Master's degree in Social, Multicultural, and Bilingual Foundations of Education and his doctoral degree in Curriculum, Administration, and Supervision at the University of Colorado, Boulder.

This report was funded by The William and Flora Hewlett Foundation.

Suggested citation: Conley, D.T. 2014. *A New Era for Educational Assessment*. Students at the Center: Deeper Learning Research Series. Boston, MA: Jobs for the Future.

Cover photography © iStockphoto/mediaphotos 2012

TABLE OF CONTENTS

INTRODUCTION

Imagine this scenario: You feel sick, and you're worried that it might be serious, so you go to the nearby health clinic. After looking over your chart, the doctor performs just two tests—measuring your blood pressure and taking your pulse—and then brings you back to the lobby. It turns out that at this clinic the policy is to check patients' vital signs and only their vital signs, prescribing all treatments based on this information alone. It would be prohibitively expensive, the doctor explains, to conduct a more thorough examination.

Most of us would find another health care provider.

Yet this is, in essence, the way in which states gauge the knowledge, skills, and capabilities of students attending their public schools. Reading and math tests are the only indicators of student achievement that "count" in federal and state accountability systems. Faced with tight budgets, policymakers have demanded that the costs associated with such testing be minimized. And, based on the quite limited information that these tests provide, they have drawn a wide range of inferences, some appropriate and some not, about students' academic performance and progress and the efficacy of the public schools they attend.

One would have to travel back in time to the agrarian era of the 1800s to find educators who still seriously believe that their only mission should be to get students to master the basics of reading and math. During the industrial age, the mission expanded to include core subjects such as science, social studies, and foreign languages, along with exploratory electives and vocational education. And in today's postindustrial society, it is commonly argued that *all* young people need the sorts of advanced content knowledge and problem solving skills that used to be taught to an elite few (Conley 2014b; JFF 2005; SCANS 1991). So why do the schools continue to rely on assessments that get at nothing beyond the "Three R's"?[1]

That's a question that countless Americans have come to ask. Increasingly, educators and parents alike are voicing their dismay over current testing and accountability practices (Gewertz 2013, 2014; Sawchuk 2014). Indeed, we may now be approaching an important crossroads in American education, as growing numbers of critics call for a fundamental change of course (Tucker 2014).

In this paper, I draw upon the results from research conducted by my colleagues and me, as well as by others, to argue that the time is ripe for a major shift in educational assessment. In particular, analysis of syllabi, assignments, assessments, and student work from entry-level college courses, combined with perceptions of instructors of those courses, provides a much more detailed picture of what college and career readiness actually entails—the knowledge, skills, and dispositions that can be assessed, taught, and learned that are strongly associated with success beyond high school (Achieve, Education Trust, & Fordham Foundation 2004; ACT 2011; Conley 2003; Conley, et al. 2006; Conley & Brown 2003; EPIC 2014a; Seburn, Frain, & Conley 2013; THECB & EPIC 2009; College Board 2006). Advances in cognitive science (Bransford, Brown, & Cocking 2000; Pellegrino & Hilton 2012), combined with the development and implementation of Common Core State Standards and their attendant assessments (Conley 2014a; CCSSO & NGA 2010a, 2010b), provide states with a golden opportunity to move toward the notion of a more comprehensive system of assessments in place of a limited set of often-overlapping measures of reading and math.

Over the next several years, as the Common Core State Standards are implemented, will educational stakeholders be satisfied with the tests that accompany those standards, or will they demand new forms of assessment? Will schools begin to use measures of student learning that address more than just reading and math? Will policymakers demand evidence that students can apply knowledge in novel and non-routine ways, across multiple subject areas and in real-world contexts? Will they come to recognize

the importance of capacities such as persistence and information synthesis, which students must develop in order to become true lifelong learners? Will they be willing to invest in assessments that get at deeper learning, addressing the whole constellation of knowledge and skills that young people need in order to be fully prepared for college, careers, and civic life?

The goal of this paper is to present a vision for a new system of assessments, one designed to support the kinds of ambitious teaching and learning that parents say they want for their children. Thankfully, the public schools do not have to create such a system from scratch—many schools already exhibit effective practices upon which others can build. For that to happen though, educators, policymakers, and other stakeholders must be willing to adopt new ways of thinking about the role of assessment in education.

In order to help readers understand how we got to the current model of testing in the nation's schools, I begin the paper with a brief historical overview. I then describe where educational assessment appears to be headed in the near term, and discuss some long-term possibilities, concluding with a series of recommendations as to how policymakers and practitioners can move toward a better model of assessment for teaching and learning.

» The goal of this paper is to present a vision for a new system of assessments, one designed to support the kinds of ambitious teaching and learning that parents say they want for their children.

HISTORICAL OVERVIEW

Ironically, due to the decentralized nature of educational governance in the United States, the nation's educators already have access to a vast array of assessment methods and tools that they can use to gain a wide range of insights into students' learning across multiple subject areas. Those methods run the gamut from individual classroom assignments and quizzes to capstone projects to state tests to admissions exams and results from Advanced Placement® and International Baccalaureate® tests. Many measures are homegrown, reflecting the boundless creativity of American educators and researchers. Others are produced professionally and have long histories and a strong commercial presence. Some measures draw upon and incorporate ideas and techniques from other sectors—such as business and the military—and from other countries, where a wider range of methods have solid, long-term track records.

The problem is that not all, or even most, schools or states take advantage of this wealth of resources. By focusing so intently on reading and math scores, federal and state policy over the past 15 or so years has forced underground many of the assessment approaches that could be used to promote and measure more complex student learning outcomes.

A HISTORICAL TENDENCY TO FOCUS ON BITS AND PIECES

The current state of educational assessment has much to do with a longstanding preoccupation in the U.S. with reliability (the ability to measure the same thing consistently) over and above concern with validity (the ability to measure the right things). To be sure, psychometricians—the designers of educational tests—have always considered validity to be critical, at least in theory (AERA, APA, & NCME 2014). In practice, though, they have had far more success in assuring the reliability of individual test forms than in dealing with messier and more complex questions about what should be tested, for what purposes, and with what consequences for the people involved.[2]

Over the past several decades, this emphasis on reliability has led to the creation of tests made up of lots of discrete questions, each one pegged to a very particular skill or bit of knowledge—the more specific the skill, the easier it becomes to create additional test items that get at the same skill at the same level of difficulty, which translates to consistent results from one test to the next.

This focus on particulars has had a clear impact on instruction. In order to prepare students to do well on such tests, schools have treated literacy and numeracy as a collection of distinct, discrete pieces to be mastered, with little attention to students' ability to put those pieces together or to apply them to other subject areas or real-world problems.

Further, if the fundamental premise of educational testing in the U.S. is that any type of knowledge can be disassembled into discrete pieces to be measured, then the corollary assumption is that, by testing students on just a sample of these pieces, one can get an adequate representation of the student's overall knowledge of the given subject.

It's a bit like the old connect-the-dots puzzles, with each item on a test representing a dot. Connect enough items and you get the outline of a picture or, in this case, an outline of a student's knowledge that, via inference, can be generalized to untested areas of the domain to reveal the "whole picture."

This certainly makes sense in principle, and it lends itself to the creation of very efficient tests that purport to generate accurate data on student comprehension of the given subject. But what if these assumptions aren't true in a larger sense? What if understanding the parts and pieces is not the same as getting the big picture that tells whether students can apply knowledge, and, perhaps most important, can transfer knowledge and skills from one context to an entirely new situation or different subject

area? If it's not possible to do these critical things, then current tests will judge students to be well educated when, in practice, they cannot use what they have been taught to solve problems in the subject area (what is known as "near transfer") or to problems in novel contexts and new areas (known as "far transfer").

ASSESSMENT BUILT ON INTELLIGENCE TESTS AND SOCIAL SORTING MODELS

Another reason for this focus on measuring literacy and numeracy in a particularistic fashion has to do with the unique evolution of assessment in this country. Interestingly, a very different approach, what would now be called "performance assessment" (referring to activities that allow students to show what they can do with what they've learned) was common in schools throughout the early 1900s, although not in a form readily recognizable to today's educator. Recitations and written examinations (which were typically developed, administered, and scored locally) were the primary means for gauging student learning. In fact, the College Board (originally the College Entrance Examination Board) was formed in 1900 to standardize the multitude of written essay entrance examinations that had proliferated among the colleges of the day.

These types of exams were not considered sufficiently "scientific," an important criticism in an era when science was being applied to the management of people. Events in the field of psychological measurement from the 1900s to the 1920s exerted an outsized influence on educational assessment. The nascent research on intelligence testing gained favor rapidly in education at a time when the techniques of scientific management had near-universal acceptance as the best means to improve organizational functioning (Tyack 1974; Tyack & Cuban 1995). Further, tests administered to all World War I conscripts seemed to validate the notion that intelligence was distributed in the form of a normal curve (hence "norm-referenced testing") among the population: immigrants and people of color scored poorly, whites scored better, and upper-income individuals scored the best. This seemed to confirm the social order of the day (Cherry 2014).

At the same time, public education in the U.S. was experiencing a meteoric increase in student enrollment, along with rising expectations for how long students would stay in school. Confronted with the need to manage such rapid growth, schools applied the thinking of the day, which led them to categorize, group, and distribute students according to their presumed abilities (Tyack 1974). Children of differing ability should surely be prepared for differing futures, the thinking went, and "scientific" tests could determine abilities and likely futures cheaply and accurately. All of this would be done in the best interest of children to help them avoid frustration and failure (Oakes 1985).

Unfortunately, the available testing technologies have never been sufficiently complex or nuanced enough to make these types of predictions very successfully, and so assessments have been used (or misused, really) throughout much of the past century to categorize students and assign them to different tracks, each one associated with a particular life pathway.[3]

> Public education in the U.S. was experiencing a meteoric increase in student enrollment, along with rising expectations for how long students would stay in school. Confronted with the need to manage such rapid growth, schools applied the thinking of the day, which led them to categorize, group, and distribute students according to their presumed abilities.

Moreover, additional problems with such norm-referenced testing—designed to see how students stack up against one another—are readily apparent. In the first place, it is not clear how to interpret the results. By definition, some students will come out on top and others will rank at the bottom. But this is no reason to assume that the top-scorers have mastered the given material (since they may just have scored a little less poorly than everybody else). Nor can it be assumed that the low-scorers are in fact less capable (since, depending on where they happen to go to school, they may never have had a chance to study the given material at all). And, finally, even if they could be trusted to sort students into winners and losers, such tests would still fail to provide much actionable information as to what those students need to learn or do to improve their scores.

ASSESSMENT TO GUIDE IMPROVEMENT

Since the late 20th century, the use of intelligence tests and academic exams to sort students into tracks has been largely discredited (Goodlad & Oakes 1988; Oakes 1985). In today's economy, when everyone needs to be capable of learning throughout their careers and lives, it would be especially counterproductive to keep sorting students in this way—far better to try to educate all children to a high level than to label some as losers and anoint others as winners as early as possible.

The first limited manifestation of an alternative approach was the mastery learning movement of the late 1970s (Block 1971; Bloom 1971; Guskey 1980a, 1980b, 1980c). Consistent with prevailing approaches to assessment, mastery learning focused entirely on basic skills in reading and math, and it reduced those skills down to the smallest testable units possible, rather than measuring students' capacity to integrate or apply their new knowledge and skills. At the same time, however, mastery learning represented a real departure from the status quo, since it argued that students should continue to receive instruction and opportunities to practice until they mastered the relevant content. In theory, everyone could succeed. The purpose of assessment was not to put students into categories but, simply, to generate information about their performance, in order to help them improve.

One of the problems with mastery learning, though, was that it was limited to content that could be broken up into dozens of distinct subcomponents that could be tested in detail (Horton 1979). As a result, educators and students were quickly overwhelmed trying to keep track of progress on all the elements. Equally vexing was the fact that mastering those elements didn't necessarily lead to proficiency in the larger subject area, or the ability to transfer what has been learned to new contexts (Horton 1979). Students could pass the reading tests only to run into trouble when they encountered new and different kinds of material, and they could ace the math tests only to be stumped by unfamiliar problems. To critics of mastery learning, the approach highlighted the limitations of shallow-learning models (Slavin 1987), a problem that "criterion-referenced" testing was designed to address.

Whereas norm-referenced tests aim to show how students stack up against each another, criterion-based assessments are meant to determine where students stand in relation to a specific standard.[4] Like mastery learning, the goal is not to identify winners and losers but, rather, to enable as many students as possible to master the given knowledge and skills. However, while mastery learning uses tests to help students master discrete bits of content, criterion-based assessments measure student performance in relation to specific learning targets and standards of performance.

EARLY STATEWIDE PERFORMANCE ASSESSMENT SYSTEMS

Initially referred to as outcomes-based education, the first wave of academic standards emerged in the late 1980s and early 1990s (Brandt 1992/1993). While borrowing from mastery learning in the sense that students were supposed to master them, these standards were more expansive and complex, designed to produce a well-educated, well-rounded student, not just one who could demonstrate discrete literacy and numeracy skills. Thus, for example, they included not just academic content knowledge, but also outcomes that related to thinking, creativity, problem solving, and the interpretation of information.

These more complex standards created a demand for assessments that went well beyond measuring bits and pieces of information. Thus, the early 1990s saw the bloom of statewide performance assessment systems that sought to gauge student learning in a much more ambitious and integrated fashion. In those years, states such as Vermont and Kentucky required students to collect their best work in "portfolios," which they could use to demonstrate their full range of knowledge and skills. Maryland introduced performance assessments (Hambleton et al. 2000), California implemented its California Learning Assessment System—CLAS—and Oregon created an elaborate system

that included classroom-based performance tasks, along with certificates of mastery at the ends of grades 10 and 12, requiring what amounted to portfolio evidence that students had mastered a set of content standards (Rothman 1995).

These assessments represented a radical departure from previous achievement tests and mastery learning models. And they were also quite difficult to manage and score—requiring more classroom time to administer, more training for teachers, and more support by state education agencies—and they quickly encountered a range of technical, operational, and political obstacles.

Vermont, for example, ran into problems establishing reliability (Koretz, Stecher, & Deibert 1993), the holy grail of U.S. psychometrics, as teachers were slow to reach a high level of consistency in their ratings of student portfolios (although their reliability did improve as teachers became more familiar with the scoring process). In California, parents raised concerns that students were being asked inappropriately personal essay questions (Dudley 1997; Kirst & Mazzeo 1996). (Also, one year, the fruit flies shipped to schools for a science experiment died en route, jeopardizing a statewide science assessment). In Oregon, some assessment tasks turned out to be too hard, and others were too easy. And everywhere, students who had excelled at taking the old tests struggled with the new assessments, leading to a backlash among angry parents of high achievers.

In the process, a great deal was learned about the dos and don'ts of large-scale performance assessment. Inevitably, though, political support for the new assessments weakened, and standards were revised once again in a number of states, resulting in a renewed emphasis on testing students on individual bits and pieces of academic content, particularly in reading and mathematics. And while a number of states continued their performance assessments systems throughout the decade, most of these systems came under increasing scrutiny due to their costs, the challenges involved in scoring them, the amount of time it took to administer them, and the difficulties involved in learning to teach to them.

The final nail in the coffin for most large-scale state performance assessment systems was the federal No Child Left Behind legislation passed in 2001, which mandated testing in English and mathematics in grades 3-8 and once in high school. The technical requirements of NCLB (as interpreted in 2002 by Department of Education staff) could only be met with standardized tests using selected-response (i.e., multiple-choice) items almost exclusively (Linn, Baker, & Betenbenner 2002; U.S. Department of Education 2001).

The designers of NCLB were not necessarily opposed to performance assessment. First and foremost, however, they were intent on using achievement tests to hold educators accountable for how well they educated all student populations (Linn 2005; Mintrop & Sunderman 2009). Thus, although the law was not specifically designed to eliminate or restrict performance assessment, this was one of its consequences. A few states (most notably Maryland, Kentucky, Connecticut, and New York) were able to hold on to performance elements of their tests, but most states retreated from almost all forms other than multiple-choice items and short essays.

Fast forward to 2014, however, and things may be poised to change once more. As I will discuss in the next section, this trend may now be on the verge of changing direction for a variety of reasons, not the least of which is a relaxing of NCLB requirements.

WHY IT'S TIME FOR ASSESSMENT TO CHANGE

An important force to consider when viewing the current landscape of assessment in U.S. schools is the rising weariness with test-based accountability systems of the type that NCLB has mandated in every state. Although the expectations contained in NCLB were both laudable and crystal clear—that all students become competent readers and capable quantitative thinkers—the means by which these qualities were to be judged led to an overemphasis on test scores derived from assessments that inadvertently devalued conceptual understanding and deeper learning. Even though student test scores improved in some areas, educators were not convinced that these changes were associated with real improvements in learning (Jennings & Rentner 2006). A desire to increase test scores led many schools to a race to the bottom in terms of the instructional strategies employed, which included an outsized emphasis on test-preparation techniques and a narrowing of the curriculum to focus, sometimes exclusively, on those standards that were tested on state assessments (Cawelti 2006).

But in addition to the public and educators tiring of NCLB-style tests (as well as the U.S. Department of Education's apparent willingness to allow states to experiment with new models), at least two other important reasons help explain why the time may be ripe for a major shift in educational assessment:

First, the results from recent research that clarifies what it means to be college and career ready make it increasingly difficult to defend the argument that NCLB-style tests are predictive of student success.

Second, recent advances in cognitive science have yielded new insights into how humans organize and use information, which make it equally difficult to defend tests that treat knowledge and skills as nothing more than a collection of discrete bits and pieces.

WHAT DOES IT MEAN TO BE COLLEGE AND CAREER READY?

The term "college and career ready" itself is relatively recent. Up until the mid-2000s, education as practiced in most high schools was geared toward making at least some students eligible to attend college, but not necessarily to make them ready to succeed.

For students hoping to attend a selective college, eligibility was achieved by taking required courses, getting sufficient grades and admission test scores, and perhaps garnering a positive letter of recommendation and participating in community activities. And for most open-enrollment institutions, it was sufficient simply for applicants to have earned a high school diploma, then apply, enroll, and pay tuition. Whether students could succeed once admitted was largely beside the point. Access was paramount.

> » A desire to increase test scores led many schools to a race to the bottom in terms of the instructional strategies employed, which included an outsized emphasis on test-preparation techniques and a narrowing of the curriculum to focus, sometimes exclusively, on those standards that were tested on state assessments.

> **Researchers have been able to identify a series of very specific factors that maximize the likelihood that students will make a successful transition to college and perform well in entry-level courses at any of a wide range of postsecondary institutions.**

The new economy has changed all of that. A little college, while better than none, is nowhere near as useful as is a certificate or degree. Being admitted to college does not mean much if the student is not prepared to complete a program of study. Further enhancing the value of readiness and the need for students to succeed is the crushing debt load ever more students are incurring to attend college now. A college education essentially has to improve a student's future economic prospects, if for no other reason than to enable debt repayment.

Why have high school educators been focused on students' eligibility for college and not on their readiness to succeed there? A key reason is that they weren't entirely sure what college readiness entailed. Until the 2000s, essentially all the research in this area used statistical techniques that involved collecting data on factors such as high school grade point average, admission tests, and the titles of high school course taken, and then trying to determine how those factors related to first-year college course grades or retention in college beyond the first term.[5] These results were useful in many ways, identifying certain high school experiences and achievements that correlated to some measures of college success. However, such research could not zero in on what, specifically, enabled some students to succeed while others struggled.

In recent years, however, researchers have been able to identify a series of very specific factors that, in combination, maximize the likelihood that students will make a successful transition to college and perform well in entry-level courses at any of a wide range of postsecondary institutions. In comparison to what was known just 15 years ago, we now have a much more comprehensive, multi-faceted, and rich portrait of what constitutes a college-ready student.

This research includes numerous studies, including many that I conducted with my colleagues, designed to identify the demands, expectations, and requirements that students tend to encounter in entry-level college courses (Brown & Conley 2007; Conley 2003, 2011, 2014b; Conley, Aspengren, & Stout 2006; Conley, et al. 2006a, 2006b; Conley, et al. 2011; Conley, McGaughy, et al. 2009a, 2009b, 2009c; Conley, et al. 2008; EPIC 2014a; Seburn, et al. 2013; THECB & EPIC 2009). These studies have analyzed course content including syllabi, texts, assignments, and instructional methods and have also gathered information from instructors of entry-level courses to determine the knowledge and skills students need to succeed in their courses.

This body of research has reached remarkably consistent conclusions about what it means to be ready to succeed in a wide range of postsecondary environments. And the key finding is one that has far-reaching implications for assessment at the high school level: In order to be prepared to succeed in college, students need much more than content knowledge and foundational skills in reading and mathematics.

On its face, this may not seem all that surprising. Yet, the prevailing methods of college admission in this country, and much research on college success, largely ignore just how critical it is for aspiring college students to develop a wide range of cognitive strategies, learning skills, knowledge about the transition to higher education, and other aspects of readiness.

For clarity's sake, I have organized these factors into a set of four "Keys" to college and career readiness. Before introducing this model, though, it's worth noting that other researchers have offered conceptual models of their own,

choosing to arrange these factors into other categories, using different terminology than I present here. Ultimately, though, it doesn't really matter whether one prefers my model or somebody else's. On the most important points—having to do with the range of factors that contribute to college readiness—researchers have reached a strong consensus. Different models represent different ways of carving up the pie, but the substance is the same.
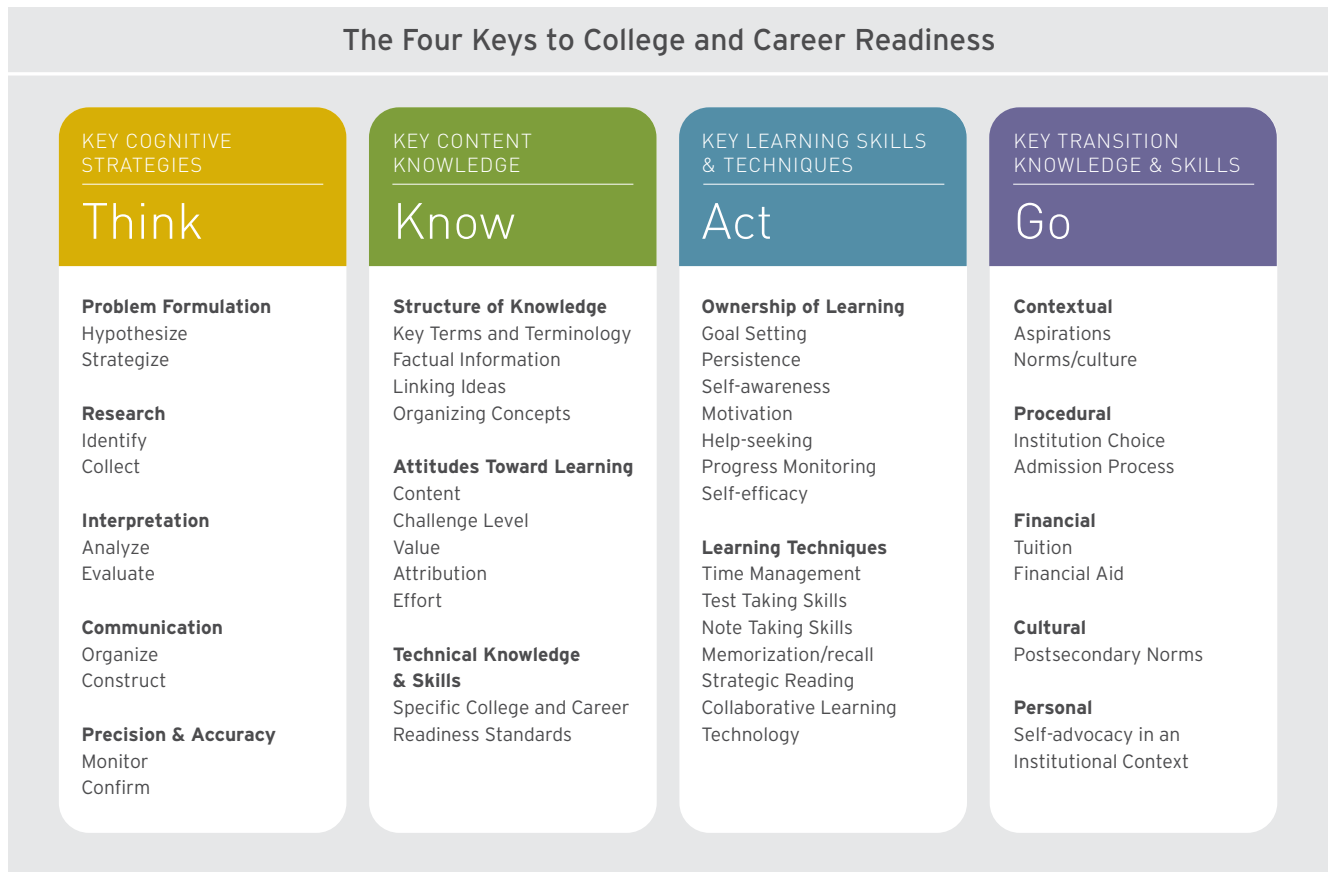
That said, the Four Keys model derives from research on literally tens of thousands of college courses at a wide range of postsecondary institutions. The model highlights four main factors that contribute to college readiness:

> **Key Cognitive Strategies.** The thinking skills students need to learn material at a deeper level and to make connections among subjects.

> **Key Content Knowledge.** The big ideas and organizing concepts of the academic disciplines that help organize all the detailed information and nomenclature that constitute the subject area along with the attitudes students have toward learning content in each subject area.

> **Key Learning Skills and Techniques.** The student ownership of learning that connects motivation, goal setting, self-regulation, metacognition, and persistence combined with specific techniques such as study skills, note taking, and technology capabilities.

> **Key Transition Knowledge and Skills.** The aspiration to attend college, the ability to choose the right college and to apply and secure necessary resources, an understanding of the expectations and norms of postsecondary education, and the capacity to advocate for one's self in a complex institutional context.

In turn, each of these Keys has a number of components, all of which are actionable by students and teachers—in other words, these are things that can be assessed, taught, and learned successfully. (On that score, note that the model does not include certain factors, such as parental income and education level, that are strongly associated statistically with college success but which *are not actionable* by schools, teachers, or students. The point here is to highlight things that can be done to prepare students to succeed, not to list the things that cannot be changed.)

Figure 1.

The Four Keys to College and Career Readiness

KEY COGNITIVE STRATEGIES

Think

Problem Formulation
Hypothesize
Strategize

Research
Identify
Collect

Interpretation
Analyze
Evaluate

Communication
Organize
Construct

Precision & Accuracy
Monitor
Confirm

KEY CONTENT KNOWLEDGE

Know

Structure of Knowledge
Key Terms and Terminology
Factual Information
Linking Ideas
Organizing Concepts

Attitudes Toward Learning
Content
Challenge Level
Value
Attribution
Effort

Technical Knowledge & Skills
Specific College and Career Readiness Standards

KEY LEARNING SKILLS & TECHNIQUES

Act

Ownership of Learning
Goal Setting
Persistence
Self-awareness
Motivation
Help-seeking
Progress Monitoring
Self-efficacy

Learning Techniques
Time Management
Test Taking Skills
Note Taking Skills
Memorization/recall
Strategic Reading
Collaborative Learning
Technology

KEY TRANSITION KNOWLEDGE & SKILLS

Go

Contextual
Aspirations
Norms/culture

Procedural
Institution Choice
Admission Process

Financial
Tuition
Financial Aid

Cultural
Postsecondary Norms

Personal
Self-advocacy in an Institutional Context

>> Students' attitudes toward learning academic material turns out to be at least as important as their aptitude.

ADVANCES IN BRAIN AND COGNITIVE SCIENCE

Recent research in brain and cognitive science provides a second major impetus for shifting the nation's schools away from a single-minded focus on current testing models and toward performance assessments that measure and encourage deeper learning.

Of particular importance is recent research into the malleability of the human brain (Hinton, Fischer, & Glennon 2012), which has provided strong evidence that individuals are capable of improving many skills and capacities that were previously thought to be fixed. Intelligence was long assumed to be a unitary, unchanging attribute, one that can be measured by a single test. However, that view has come to be replaced by the understanding that intellectual capacities are varied and multi-dimensional and can be developed over time, if the brain is stimulated to do so.

One critical finding is that students' attitudes toward learning academic material is at least as important as their aptitude (Dweck, Walton, & Cohen 2011). For generations, test designers have used "observed" ability levels ascertained from test scores to steer them into academic and career pathways that match their natural talents and capabilities. But the reality is that, far from helping students find their place, such test results can also serve to discourage many students from making the sorts of sustained, productive efforts that would allow them to succeed at a more challenging course of study.

Recent research also challenges the commonly held belief that the human brain is organized like a library, with discrete bits of information grouped by topic in a neat and orderly fashion, to be recalled on demand (Donovan, Bransford, & Pellegrino 1999; Pellegrino & Hilton 2012). In fact, evidence reveals that the brain is quite sensitive to the importance of information, and it makes sense of sensory input largely by determining its relevance (Medina 2008). Thus, the longstanding American preoccupation with breaking subject-area knowledge down into small bits, testing students' mastery of each one, and then teaching those bits sequentially, may in fact be counterproductive. Rather than ensuring that students learn systematically, piece by piece, this approach could easily deny them critical opportunities to get the big picture and to figure out which information and concepts are most important.

When confronted by a torrent of bits and pieces presented one after the other, without a chance to form strong links among them, the brain tends to forget some, connect others in unintended ways, experience gaps in sequencing, and miss whatever larger purpose and meaning might have been intended. Likewise, when tests are designed to measure students' mastery of discrete bits, they provide few useful insights into students' conceptual understanding or their knowledge of how any particular piece of information relates to the larger whole.

>> Rather than being taught skills and facts in isolation, high school students should be deepening their mastery of key concepts and skills they were taught in earlier grades, learning to apply and extend that foundational knowledge to new topics, subjects, problems, tasks, and challenges.

> » Opportunities for students to demonstrate their conceptual understanding, to relate smaller ideas to bigger ones, and to show that they grasp the overall significance of what they have learned.

The net result is that students struggle to retain information (NRC 2002). Having received few cues about the relative importance of the given content, and having few opportunities to fit it into a larger framework, it's no wonder that they often forget much of what they have learned, from one year to the next, or that even though they can answer detailed questions about a topic, they struggle to demonstrate understanding of the larger relevance or meaning of the material. Indeed, this is one possible explanation for why scores at the high school level on tests such as the National Assessment of Educational Progress, or NAEP—which gets at students' conceptual understanding, along with their content knowledge—have flat-lined over the past two decades, a period when the emphasis on basic skills increased dramatically.

Ideally, secondary-level instruction guides students through learning progressions that build in complexity over time, moving toward larger and more integrated structures of knowledge. Rather than being taught skills and facts in isolation, high school students should be deepening their mastery of key concepts and skills they were taught in earlier grades, learning to apply and extend that foundational knowledge to new topics, subjects, problems, tasks, and challenges.

And in order to provide this sort of instruction, teachers require tests and tools that allow them to assess far more than just the ability to recall bits and pieces of content. What is needed, rather, are opportunities for students to demonstrate their conceptual understanding, to relate smaller ideas to bigger ones, and to show that they grasp the overall significance of what they have learned.
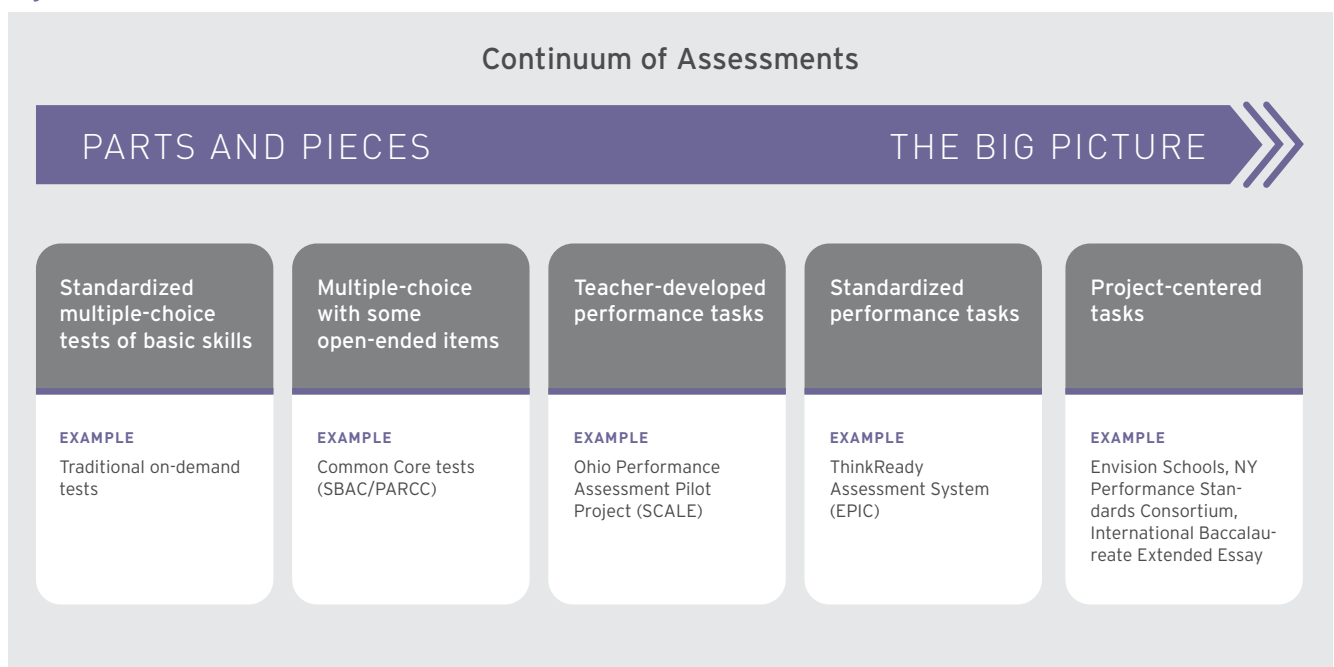
MOVING TOWARD A BROADER RANGE OF ASSESSMENTS

Assessments can be described as falling along a continuum, ranging from those that measure bits and pieces of student content knowledge to those that seek to capture student understanding in more integrated and holistic ways (as shown in Figure 2). But it is not necessary or even desirable to choose just one approach and reject the others. As I describe in the following pages, a number of states are now creating school assessment models that combine elements from multiple approaches, which promises to give them a much more detailed and useful picture of student learning than if they insisted on a single approach.

TRADITIONAL MULTIPLE-CHOICE TESTS

Traditional multiple-choice tests have come under a great deal of criticism in recent years, but whatever their flaws, they are a mature technology that offers some distinct advantages. They tend to be reliable, as noted. Also, in comparison to some other forms of assessments, they do not require a lot of time or cost a lot of money to administer, and they generate scores that are familiar to educators. Thus, it's not surprising that a number of states,

when given the option of using the tests of the Common Core developed by the two state consortia–Partnership for the Assessment of Readiness for College and Careers or Smarter Balanced Assessment Consortium–have instead chosen to reinstitute multiple-choice tests with which they are already familiar. It is likely that multiple-choice tests will continue to be widely used for some time to come, as evidenced by the fact that the Common Core assessments continue to include items of this type in addition to some new item types.

Figure 2.

Continuum of Assessments				
PARTS AND PIECES			THE BIG PICTURE	
Standardized multiple-choice tests of basic skills	Multiple-choice with some open-ended items	Teacher-developed performance tasks	Standardized performance tasks	Project-centered tasks
EXAMPLE Traditional on-demand tests	EXAMPLE Common Core tests (SBAC/PARCC)	EXAMPLE Ohio Performance Assessment Pilot Project (SCALE)	EXAMPLE ThinkReady Assessment System (EPIC)	EXAMPLE Envision Schools, NY Performance Standards Consortium, International Baccalaureate Extended Essay

One recent advancement in this area is the design and use of computer-adaptive tests, which add a great deal of efficiency to the testing process. Depending on the student's responses, the software will automatically adjust the level of difficulty of the questions it poses (after a number of correct answers, it will move on to harder items; too many incorrect responses, and it will move back to easier ones), quickly zeroing in on student's level of mastery of the given material. Further, the technology makes it a simple matter to include items that test content from previous and subsequent grades, which allows measurement of a very wide distribution of knowledge and skills (from below grade level to far above it) that might exist in any given class or testing group.

COMMON CORE TESTS

Two consortia of states have developed tests of the Common Core State Standards, and both of them—the Partnership for the Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC)—have been touted for their potential to overcome many of the shortcomings of NCLB-inspired testing.

These exams will test a range of Common Core standards at grades 3-8 and once in high school, using a mix of methods including, potentially, some performance tasks that get at more complex learning. However, the tests still rely predominantly on items that gauge student understanding of discrete knowledge and, hence do not address a number of key Common Core standards that require more extensive cognitive processing and deeper learning.

This is a critical point, and it bears repeating: While the PARCC and SBAC assessments have been designed specifically to measure student progress on the Common Core standards, in point of fact they address only some of those standards.

Many of the skills that the Common Core defines as necessary preparation for college and careers are ones that can only be tested validly through a wider range of methods than either PARCC or SBAC currently employs. For example, the standards specify that, by the time students graduate from high school, they should be able to:

> Conduct research and synthesize information

> Develop and evaluate claims

> Read critically and analyze complex texts

> Communicate ideas through writing, speaking, and responding

> Plan, evaluate, and refine solution strategies

> Design and use mathematical models

> Explain, justify, and critique mathematical reasoning

In short, many of the standards contained in the Common Core call upon students to demonstrate quite sophisticated knowledge and skills, requiring more complex forms of assessment than PARCC and SBAC can reasonably be expected to provide from a test that will be administered over several hours on a computer.

>> Many of the standards contained in the Common Core call upon students to demonstrate quite sophisticated knowledge and skills, requiring more complex forms of assessment than PARCC and SBAC can reasonably be expected to provide from a test that will be administered over several hours on a computer.

That's not to denigrate those assessments but, rather, to argue that they are not, in and of themselves, sufficient to meet the Common Core's requirements. If states mean to take these learning goals seriously, then they will have to consider a much broader continuum of options for measuring them, including assessments that are now being developed and used locally, in networks, and, in some cases, by states on a limited basis. Such assessments—including performance tasks, student projects, and collections of evidence of student learning—are both feasible and valid, but they also present challenges of their own.

PERFORMANCE TASKS

Performance tasks have been a part of state-level and school-level assessment for decades. They encompass a wide range of formats, requiring students to complete tasks that can take anywhere from twenty minutes to two weeks, and that require them to engage with content that can range from a two-paragraph passage to a whole collection of source documents. Generally speaking, though, most performance tasks consist of activities that can be completed in a few class periods at most, and which do not require students to conduct extensive independent research.

A number of prominent examples deserve mention:

In 1997, the New York Performance Standards Consortium, a group of New York schools with a history of using performance tasks as a central element of their school-based assessment programs, sued the State of New York, successfully, to allow the use of performance tasks to meet state testing requirements (Knecht 2007). Most notable among these schools was Central Park East Secondary School, which had a long and distinguished history of having students present their work to panels consisting of fellow students, teachers, and community members with expertise in the subject matter being presented. Most of these schools were also members of the Coalition of Essential Schools, which also advocated for these types of assessment at its over 600 member schools.

More recently, my colleagues at the Educational Policy Improvement Center (EPIC) and I developed ThinkReady, an assessment of Key Cognitive Strategies (Baldwin, Seburn, & Conley 2011; Conley 2007; Conley, et al. 2007). Its performance tasks—which take anywhere from a few class periods to several weeks (with out-of-class work) to complete—require students to demonstrate skills in problem formulation, research, interpretation, communication, and the use of precision and accuracy throughout the task. Teachers use a common scoring guide that tells them where students stand, on a progression from novice to emerging expert, on the kind of thinking associated with college readiness. The system spans grades 6-12 and is organized around four benchmark levels that correspond with cognitive skill development rather than grade level.

The Ohio Performance Assessment Pilot Project was conceived of as a pilot project to identify how performance-based assessment could be used in Ohio (Ohio Department of Education 2014a, 2014b). Teachers developed tasks at grades 3-5 and 9-12 in English, mathematics, science, social studies, and career and technical pathways. The tasks were field tested and piloted and then refined. Tasks were scored online and at in-person scoring sessions.

New Hampshire is in the process of developing common statewide performance tasks that will be included within a comprehensive state assessment system along with SBAC assessments (New Hampshire Department of Education 2014). Each performance task will be a complex curriculum-embedded assignment involving multiple steps that require students to use metacognitive learning skills. As a result, student performance will reflect the depth of what students have learned and their ability to apply that learning as well.

The tasks will be based on college and career ready competencies across major academic disciplines including the Common Core State Standards-aligned competencies for English Language Arts & Literacy and Mathematics, as well as New Hampshire's K-12 Model Science Competencies recently approved by the New Hampshire Board of Education (New Hampshire Department of Education 2014). Performance tasks will be developed for elementary, middle, and high school grade spans. They will be used to compare student performance across the state in areas not tested by SBAC, such as the ability to apply learning strategies to complex tasks.

New Hampshire also partnered with the Center for Collaborative Education and the National Center for the Improvement of Educational Assessment to develop the Performance Assessment for Competency Education, or PACE, designed to measure student mastery of college and

career ready competencies (New Hampshire Department of Education 2014). PACE includes a web-based bank of common and locally designed performance tasks, to be supplemented with regional scoring sessions and local district peer review audits.

Colorado, Kansas, and Mississippi have partnered with the Center for Education Testing & Evaluation at the University of Kansas to form the Career Pathways Collaborative. The partnership's Career Pathways Assessment System—cPass— is designed to measure high school student readiness for entry into college and/or the workforce (CETE 2014). It uses a mix of multiple-choice questions and performance tasks both in the classroom and in real-world situations to measure the knowledge and skills necessary for specific career pathways.

It is worth noting parenthetically that the Advanced Placement® testing program has long included an open-ended component known as a constructed response item and does allow for other artifacts of learning on a very small number of exams, such as the Studio Art exams portfolio requirement. In addition, the College and Work Readiness Assessment, or CWRA+, combines selected-response items with performance-based assessment to determine student proficiency in complex areas such as analysis and problem solving, scientific and quantitative reasoning, critical reading and evaluation, and critiquing an argument (Council for Aid to Education 2014). When answering the selected-response items, students refer to supporting documents such as letters, memos, photographs, charts, or newspaper articles.

Finally, both PARCC and SBAC include performance assessments in a limited fashion, by requiring students to construct complex written responses to prompts (PARCC 2014; SBAC 2014). The specifics of these tasks, the number that will be required, and their inclusion in calculations of final student scores is all still under consideration, to be decided on a state-by-state basis. However, the tests themselves will incorporate some fairly innovative items that elicit a high level of student engagement and reasoning by requiring them to elaborate upon and provide evidence to support the answers they provide.

PROJECT-CENTERED ASSESSMENT

Much like performance tasks, project-centered assessment engages students in open-ended, challenging problems (Soland, Hamilton, & Stecher 2013). The differences between the two approaches have to do mainly with their scope, complexity, and the time and resources they require. Projects tend to involve more lengthy, multistep activities, such as research papers, the extended essay required for the International Baccalaureate Diploma, or assignments that conclude with a major student presentation of a significant project or piece of research.

For example, Envision Schools, a secondary-level charter school network in the San Francisco area, have made this kind of assessment a central feature of their instructional program, requiring students to conduct semester- or year-long projects that culminate in a series of products and presentations, which undergo formal review by teachers and peers (SCALE 2014). A student or team of students might undertake an investigation of, say, locally sourced food—this might involve researching where the food they eat comes from, what proportion of the price represents transportation, how dependent they are on other parts of the country for their food, what choices they could make if they wished to eat more locally produced food, what the economic implications of doing so would be, whether doing so could cause economic disruption in other parts

>> Both PARCC and SBAC will incorporate some fairly innovative items that elicit a high level of student engagement and reasoning by requiring them to elaborate upon and provide evidence to support the answers they provide.

of the country as an unintended consequence, and so on. The project would then be presented to the class and scored by the teacher using a scoring guide that includes ratings of the students' use of mathematics and economics content knowledge; the quality of argumentation; the appropriateness of sources of information cited and referenced; the quality and logic of the conclusions reached; and overall precision, accuracy, and attention to detail.

Another well-known example is the Summit Charter Network of schools, also located in the Bay Area (Gates Foundation 2014). While Summit requires students to master high-level academic standards and cognitive skills, the specific topics they study and the particular ways in which they are assessed are personalized, planned out according to their needs and interests. The school's schedule provides students ample time to work individually and in groups on projects that address key content in the core subject areas. And in the process, students assemble digital portfolios of their work, providing evidence that they have developed important cognitive skills (including specific "habits of success," the metacognitive learning skills associated with readiness for college and career), acquired essential content knowledge, and learned how to apply that knowledge across a range of academic and real-world contexts. Ultimately, the goal is for students to present projects and products that can withstand public critique and are potentially publishable.

COLLECTIONS OF EVIDENCE

Strictly speaking, collections of evidence are not assessments at all. Rather, they offer a way to organize and review a broad range of assessment results, so that educators can make accurate decisions about student readiness for academic advancement, high school graduation, or postsecondary programs of study (Conley 2005; Oregon State Department of Education Salem 2005).

For example, New Hampshire recently introduced a technology portfolio for graduation, which allows students to collect evidence to show how they have met standards in this field. And the New York Performance Standards Consortium, which currently consists of more than 40 in-state secondary schools, as well as others beyond New York, received a state-approved waiver allowing its students to complete a graduation portfolio in lieu of some of New York's Regents Examination requirements. Students must compile a set of ambitious performance tasks for their portfolios, including a scientific investigation, a mathematical model, a literary analysis, and a history/social science research paper, sometimes augmented with other tasks such as an arts demonstration or analyses of a community service or internship experience. All of these are measured against clear academic standards and are evaluated using common scoring rubrics.

The state of Kentucky adopted a similar approach as a result of its Education Reform Act of 1990, which included KIRIS, the Kentucky Instructional Results Information System (Stecher, et al. 1997). Implemented in 1992, KIRIS incorporated information from several assessment sources, including multiple-choice and short-essay questions, performance "events" requiring students to solve applied problems, and collections of students' best work in writing and mathematics (though students were also assessed in reading, social science, science, arts and humanities, and practical living/vocational studies). The writing assessment, which continued until 2012, was especially rigorous: In grades 4, 7, and 12, students submitted three to four pieces of written work to be evaluated, and in grades 5, 8, and 12 they completed on-demand writing tasks, with teachers assessing their command of several genres, including reflective essays, expressive or literary work, and writing that uses information to persuade an audience.

> » Organize and review a broad range of assessment results, so that educators can make accurate decisions about student readiness for academic advancement, high school graduation, or postsecondary programs of study.

In 2009, the Oregon State Board of Education adopted new diploma requirements, specifying that students must demonstrate proficiency in a number of "Essential Skills." These include goals in traditional subject areas such as reading, writing, and mathematics, but they also address a number of other complex, cross-cutting outcomes, such as the ability to think critically and analytically, to use technology in a variety of contexts, to demonstrate civic and community engagement, to demonstrate global literacy, and to demonstrate personal management and teamwork skills. Basic academic skills will be tested via the SBAC exam, while the remaining Essential Skills will be assessed via measures developed locally or selected from a set of approved methods (Oregon Department of Education 2014).

Such approaches, in which a range of student assessment information is collected over time, permit educators to combine some or all of the elements on the continuum of assessments presented in figure 2 on page 12. Doing so results in a fuller picture of student capabilities than is possible with any single form of assessment. And because this allows for the ongoing, detailed analysis of student work, it gives schools the option to assess their progress on relatively complex cognitive skills, which is very difficult to measure using occasional achievement tests.

OTHER ASSESSMENT INNOVATIONS

Recently, the Asia Society commissioned the RAND Corporation to produce an overview of models and methods for measuring 21st-century competencies (Soland, Hamilton, & Stecher 2013). The resulting report describes a number of models that closely map onto the range of assessments described in figure 2, on page 12. However, it also describes "cutting-edge measures" such as assessments of higher-order thinking used by the Program for International Student Assessment (PISA) and the Graduation Performance System.

Coordinated by the Organization for Economic Cooperation and Development, PISA is a test, first administered in 2000, designed to allow for comparisons of student performance among member countries. Administered every three years to randomly selected 15-year-olds, it assesses knowledge and skills in mathematics, reading, and science, but it is perhaps best known for its emphasis on problem-solving skills and other more complex (sometimes referred to as "hard to measure") cognitive processes, which it gauges through the use of innovative types of test items.

Beginning in 2015, for example, PISA will introduce an online assessment of students' performance on tasks that require collaborative problem solving. Through interactions with a digital avatar (simulating a partner the student has to work with on a project), test-takers will demonstrate their skills in establishing and maintaining a shared understanding of a problem, taking appropriate action to solve it, and establishing and maintaining team organization. Doing so requires a series of deeper learning skills including analyzing and representing a problem; formulating, planning, and executing a solution; and monitoring and reflecting on progress. During the simulation, students encounter scenarios in which the context of the problem, the information available, the relationships among group members, and the type of problem all vary, and they are scored based on their responses to the computer program's scenarios, prompts, and actions. Early evidence suggests that this method is quite effective in distinguishing different collaborative problem-solving skill levels and competencies.

Developed collaboratively by Asia Society and the Stanford Center for Assessment, Learning, and Equity, the Graduation Performance System (GPS) measures student progress in a number of areas, with particular emphasis on gauging how "globally competent" they are—i.e., how knowledgeable about international issues and able to recognize cross-cultural differences, weigh competing perspectives, interact with diverse partners, and apply various disciplinary methods and resources to the study of global problems. The GPS assesses critical thinking and communication, and it provides educators flexibility to make choices regarding the specific pieces of student work that are selected to illustrate student skills in these areas.

Further, national testing organizations such as ACT and the College Board, makers of the SAT, are updating their systems of exams to keep them in step with recent research on the knowledge and thinking skills that students need to succeed in college, although these tests will remain in their current formats and not involve student-generated work products beyond an optional on-demand essay. ACT has introduced Aspire, a series of summative, interim, and classroom exams and optional measures of metacognitive skills, designed to determine whether students are on a path to college and career readiness from third grade on (ACT 2014). The SAT in particular is undergoing a series of changes that require test-takers to cite evidence to a greater degree when making claims, as well as to understand what they are reading more deeply than just being able to identify the sequence of events or cite key

> Metacognition occurs when learners demonstrate awareness of their own thinking, then monitor and analyze their thinking and decision-making processes or—as competent learners often do—recognize that they are having trouble and adjust their learning strategies.

ideas in a passage (College Board 2014). However, these tests will continue to consist primarily of selected-response items, with all of the attendant limitations of this particular testing method. An essay option is available on both tests.

METACOGNITIVE LEARNING STRATEGIES ASSESSMENTS

Metacognitive learning strategies are the things students do to enable and activate thinking, remembering, understanding, and information processing more generally (Conley 2014c). Metacognition occurs when learners demonstrate awareness of their own thinking, then monitor and analyze their thinking and decision-making processes or—as competent learners often do—recognize that they are having trouble and adjust their learning strategies.

Indeed, metacognitive skills often contribute as much or even more than subject-specific content knowledge to students' success in college. When faced with challenging new coursework, students with highly developed learning strategies tend to have an important advantage over peers who can only learn procedurally (i.e., by following directions).

Similarly, assessments designed to gauge students' learning skills offer an important complement to tests that measure content knowledge alone. Ideally, they can provide teachers with useful insights into why students might be having trouble learning certain material or completing a particular assignment.

However, measures of these skills and strategies are subject to their own set of criticisms. For example, many of them rely on student self-reports (e.g., questionnaires about what was easy or difficult about an assignment), which limits their use for high-stakes purposes. Critics also point out that, while they may not be intended for this purpose, they can easily lead teachers to make character judgments about students, bringing an unnecessary source of bias into

the classroom. Finally, the measurement properties of many early instruments in this area have been somewhat suspect, particularly when it comes to reliability. In short, while assessments of metacognition can be useful, educators and policymakers have good reason to take care in their use and in the interpretation of results.

Still, it is beyond dispute that many educators and, increasingly, policymakers are taking a closer look at such measures, excited by their potential to help have an impact on the achievement gap for underperforming students. For example, public interest has surged, of late, in the role that perseverance, determination, tenacity, and grit can play in learning (Duckworth & Peterson 2007; MacCann, Duckworth, & Roberts 2009; Tough 2012). So, too, has the notion of academic mindset struck a chord with many practitioners who see evidence daily that students who believe that effort matters more than innate aptitude are able to perform better in a subject (Farrington 2013). And researchers are now pursuing numerous studies of students' use of study skills, their time management strategies, and their goal setting capabilities.

In large part, what makes all of these metacognitive skills so appealing is the recognition that such things can be taught and learned, and that the evidence suggests that all are important for success in and beyond school.

One of the best-known assessment tools in this area is Angela Duckworth's Grit Index (Duckworth Lab 2014), which consists of a dozen questions that students can quickly complete. These questions can predict the likelihood of their completing high school or doing well in situations that require sustained focus and effort. Another, Carol Dweck's Growth Mindset program (mindsetworks 2014), helps learners understand and change the way they think about how to succeed academically. The program focuses on teaching students that their attitude toward a subject is as important as any native ability they have in the subject.

EPIC's CampusReady instrument is designed to assess students' self-perceptions of college and career readiness in each of the Four Keys described earlier (EPIC 2014b). It touches on many aspects of grit and academic mindset, as well as a number of other attitudes, habits, behaviors, and beliefs necessary to succeed at postsecondary studies.

The California Office to Reform Education districts will incorporate metacognitive assessments into their accountability systems, starting in the 2014-15 academic year (CORE 2014). Four metacognitive assessments are currently being piloted across twenty CORE schools. These four metacognitive assessments are designed to measure growth mindset, self-efficacy, self-management, and social awareness. For each metacognitive assessment, one version has been selected from existing measures, while the other version has been developed in partnership with methodological experts in an effort to improve upon existing measures.

While a great deal of attention is currently being paid to these metacognitive measures, they still face a range of challenges before they are likely to be used as widely or for as many purposes as traditional multiple-choice tests. Perhaps the greatest obstacle to their use is the fact that most rely on self-reported information, which is subject to socially desirable bias—in other words, even if no stakes are attached to the assessment, respondents tend to give answers they believe people want to see.

This issue can be addressed to some extent by triangulating responses and scores against other data sources, such as a test score or attendance record, or even other items in an instrument, such as those that ask students how they spend their time. Inconsistencies can indicate the presence of socially desirable responses. Over time, students can be encouraged to provide more honest self-assessments, particularly if they know they will not be punished or rewarded excessively based on their responses.

However, information of this sort is best used longitudinally, to ascertain overall trends and to determine if students are developing the learning strategies and mindsets necessary to be successful lifelong learners. Such assessments can help guide teachers and students toward developing important strategies and capabilities that enhance learner success and enable deeper learning, but they should not be overemphasized or misused for high stakes purposes, certainly not until more work has been done to understand how best to use these types of instruments.

» Metacognitive assessments can help guide teachers and students toward developing important capabilities that enhance learner success and enable deeper learning, but these assessments should not be overemphasized or misused for high-stakes purposes.

TOWARD A SYSTEM OF ASSESSMENTS[6]

As the implementation of the Common Core proceeds, and as a number of states rethink their existing achievement tests, a golden opportunity may be presenting itself for states to move toward much better models of assessment. It may now be possible to create combinations of measures that not only meet states' accountability needs but that also provide students, teachers, schools, and postsecondary institutions with valid information that empowers them to make wise educational decisions.

Today's resurgent interest in performance tasks, coupled with new attention to the value of metacognitive learning skills, invites progress toward what I like to call a "system of assessments," a comprehensive approach that draws from multiple sources in order to develop a holistic picture of student knowledge and skills in all of the areas that make a real difference for college, career, and life success.

The new PARCC and SBAC assessments have an important contribution to make to this effort, in that they offer well-conceived test items along with carefully designed performance tasks that require valuable writing skills and problem-solving capabilities. These assessments should help signal to students that they are expected to engage deeply in learning and to devote serious time and effort to developing higher-order thinking skills. On their own, however, the Common Core assessments are not a system.

A genuine system of assessments would address the varied needs of all of the constituents who use assessment data, including public schools; postsecondary institutions; state education departments, state and federal policymaking bodies, education advocacy groups; business and community groups; and others. It would serve purposes that go well beyond the task of rating schools, judging them to be successes or failures. Most importantly, it would avoid placing too much weight on any single source of data. In short, such a system would produce a nuanced and multilayered profile of student learners.

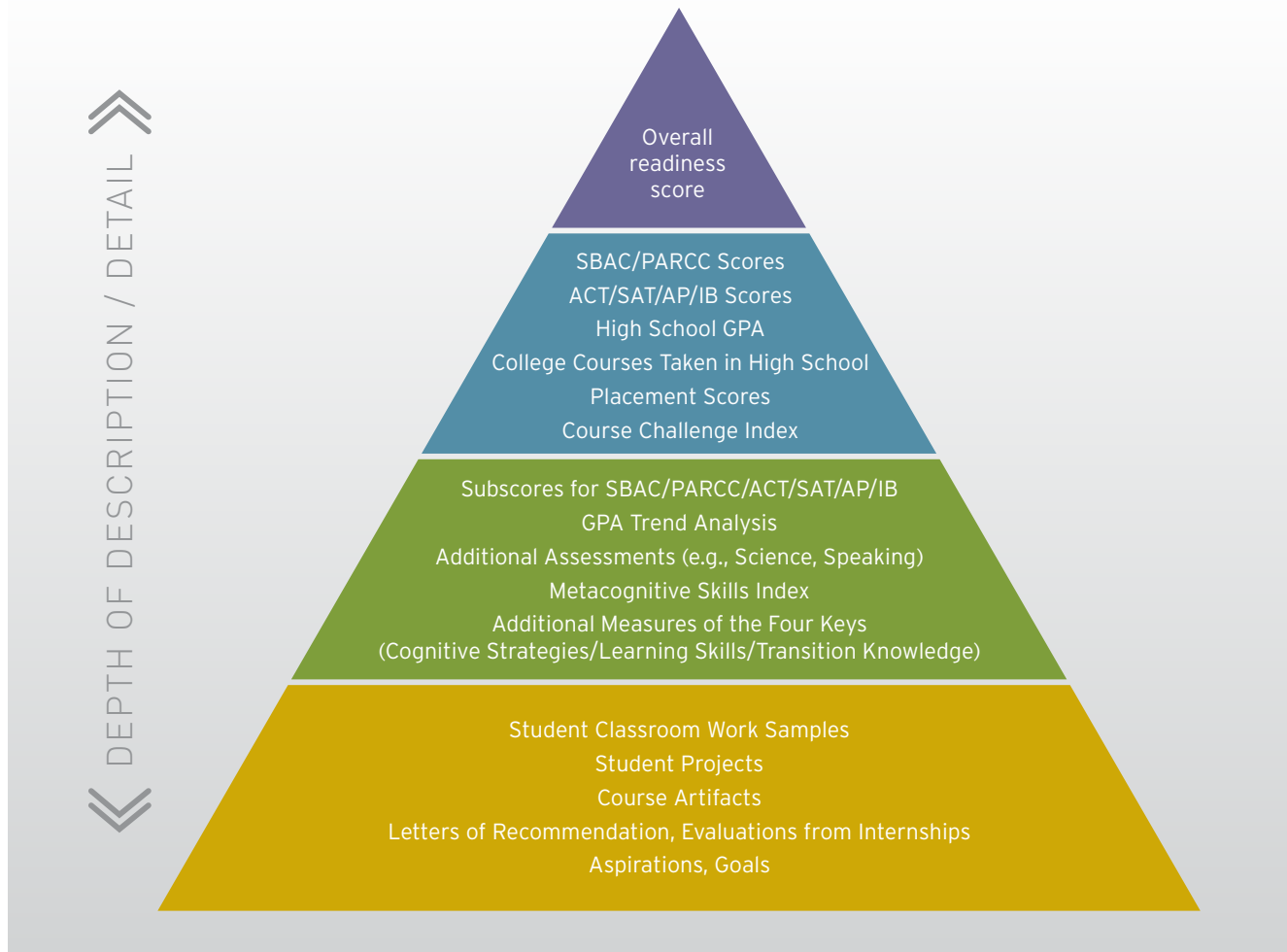## A PROFILE APPROACH TO READINESS AND DEEPER LEARNING[7]

A system of assessments yields many more data points than does a single achievement test. Compared to the familiar connect-the-dots sketch of students' knowledge and skills, it offers a much more precise, high-definition picture of where they are, how far they've come, and how far they have to go in order to be ready for college and careers.

Ultimately, this should allow educators to create profiles of individual students that are far more detailed than the familiar high school transcript, which tends to list just a few test scores and teacher-generated grades. Rather, it should be possible to use a more integrative and personalized series of measures, calibrated to individual student goals and aspirations, which highlights much more of what those students know and are able to do.

> » A genuine system of assessments would address the varied needs of all of the constituents who use assessment data, including public schools; postsecondary institutions; state education departments, state and federal policymaking bodies, education advocacy groups; business and community groups; and others.

Figure 3.

DEPTH OF DESCRIPTION / DETAIL

Overall
readiness
score

SBAC/PARCC Scores
ACT/SAT/AP/IB Scores
High School GPA
College Courses Taken in High School
Placement Scores
Course Challenge Index

Subscores for SBAC/PARCC/ACT/SAT/AP/IB
GPA Trend Analysis
Additional Assessments (e.g., Science, Speaking)
Metacognitive Skills Index
Additional Measures of the Four Keys
(Cognitive Strategies/Learning Skills/Transition Knowledge)

Student Classroom Work Samples
Student Projects
Course Artifacts
Letters of Recommendation, Evaluations from Internships
Aspirations, Goals

Such a profile might have something like a wedding-cake structure, with the more familiar college admission tests and the PARCC and SBAC assessments (or other state tests) on the top levels, and additional information gathered systematically and in greater detail at each subsequent level. For example, it would include familiar data such as high school grades and GPA, but it would also include novel sources of data, such as research papers and capstone projects, students' assessments of their own key learning skills over multiple years, indicators of perseverance and goal focus as evidenced by their completion of complex projects, and teachers' judgments of student characteristics (aggregated so as to eliminate outlier views about the student). Figure 3 offers an illustration of the sorts of information that could be included.

Subordinate levels of the profile would contain additional information including actual student work with insights into the techniques and strategies they used to generate the work. Student work would be sorted and categorized through the use of metadata tags to array it by characteristic that would make it easy and convenient for a reviewer to pull up samples based on areas of interest, such as interpretive thinking or research or mathematical reasoning.

Note, however, that this would not be the same as a portfolio of student work. While portfolios may remain useful within schools, they do not translate well to out-of-school uses. The profile model, rather, could serve not just individual students and their teachers and parents but also a range of potential external users, too, such as college admission officers, advisors, and instructors or potential employers. To be sure, safeguards would have to be in place in order to ensure students' privacy and protect against misuse of their information–just as is true today of student transcripts. But as long as safeguards are in place, then a profile should offer quite useful insights into students' progress and valuable diagnostic information that can be used to help them prepare for college and careers.

> Scoring may be the holy grail of performance assessment of deeper learning. Until and unless designers can devise better ways to score complex student work, either by teachers or externally, the Common Core standards that reflect deeper learning will largely be neglected by the designers of large-scale statewide assessments.

CHALLENGES OF DEEPER LEARNING ASSESSMENT

Today's information technologies are sufficiently sophisticated and efficient enough to manage the complex information generated by a system of assessments. They would, however, still face a series of daunting challenges in order to be implemented successfully and on a large scale.

Although some states, researchers, and testing organizations are seeking to develop new methods to assess deeper learning skills on a large scale, none have yet cracked the code to produce an assessment that can be scored in an automated fashion at costs in line with current tests. Indeed, scoring may be the holy grail of performance assessment of deeper learning. Until and unless designers can devise better ways to score complex student work, either by teachers or externally, the Common Core standards that reflect deeper learning will largely be neglected by the designers of large-scale statewide assessments, at least those used for high-stakes accountability purposes.

As long as the primary purpose of assessments is to reach judgments about students and schools (and, increasingly, teachers), reliability and efficiency will continue to trump validity. Thankfully, though, one important lesson to emerge from No Child Left Behind—and its decade-long rush to judge the quality of individual schools—is that not all assessment are, or should be, summative. In fact, the majority of the assessment that goes on every day in schools is designed not to hold anybody accountable but to help people make immediate decisions about how to improve student performance and teaching practice. Over the past 10 years, educators have learned the distinction between summative and formative assessments, and they know full well that not all measures must be high stakes in nature or that all judgments need be derived from multiple-choice tests.

While it will always be important to know how well schools are teaching foundational skills in English language arts and mathematics, the pursuit of deeper learning will require a much greater emphasis on formative assessments that signal to students and teachers what they must do to become ready for college and careers, including the development of metacognitive learning skills—about which selected response tests provide no information at all.

In fact, skills such as persistence, goal focus, attention to detail, investigation, and information synthesis are more likely to be the most important for success in the coming decades. It will become increasingly critical for young people to learn how to cope with college assignments or

> The pursuit of deeper learning will require a much greater emphasis on formative assessments that signal to students and teachers what they must do to become ready for college and careers, including the development of metacognitive learning skills.

》 More innovative campuses and systems are already gearing up to make decisions more strategically and to learn how to use something more like a profile of readiness rather than just a cut score for eligibility.

work tasks that do not have one right answer, that require them to gather new information and make judgments about the information they collect, and that may have no simple or obvious solution. Such integrative and applied skills can be assessed, and they can be assessed most usefully by way of performance assessments. They neither can nor should be measured at the granular level that is the focus of most standardized tests.

A final, though by no means trivial, question is whether the nation's postsecondary institutions, having relied for so many decades on multiple-choice tests to help them make admission and placement decisions, can or will use information from assessments of deeper learning, if such sources of data exist.

The short and somewhat unsatisfying answer is that most states are not giving much thought to how to provide postsecondary institutions with more information on student readiness or on the deeper learning skills associated with postsecondary success beyond the Common Core assessment results, which will be used to exempt students from remedial education requirements. Consequently, postsecondary institutions are doing little

to signal any interest in more complex information on readiness nor to work with secondary education to develop the data collection and interpretation systems necessary to use results from profiles, portfolios, and performance tasks to gain more interesting and potentially useful insights into student readiness.

Again, it is worth noting that this use of PARCC and SBAC represents a useful step forward. At the same time, though, it should not be mistaken for the kind of bold leap that will be required in order to capture the student knowledge, skills, abilities, and strategies associated with postsecondary readiness and success.

The postsecondary community seems to be spread along a continuum from being resigned to having to accommodate more information to being eager to be able to make better decisions about student readiness. While concerns always exist at larger institutions, especially about how they will process more diverse data for thousands of applicants, the more innovative campuses and systems are already gearing up to make decisions more strategically and to learn how to use something more like a profile of readiness rather than just a cut score for eligibility.

RECOMMENDATIONS

Many issues will need to be addressed in order to bring about the fundamental changes in assessment practice necessary to promote and value deeper learning. The recommendations offered here are meant to serve as a starting point for a process that likely will unfold over many years, perhaps even decades. The question is: Can policymakers sustain their attention to this issue long enough to enact the policies necessary to bring about necessary changes? For that matter, can educators follow through with new programs and practices that turn policy goals into reality? And will the secondary and postsecondary systems be able to cooperate in creating systems of assessments and focusing instruction on deeper learning?

I believe that if we are to move toward these goals, education policymakers will need to:

1. **Define college and career readiness comprehensively.** States need clear definitions of college and career readiness that highlight the full range of knowledge, skills, and dispositions that research shows to be critical to students' success beyond high school (including not only key content knowledge but also cognitive strategies, learning skills and techniques, and knowledge and skills related to the transition to college and the workforce).

2. **Take a hard look at the pros and cons of current state accountability systems.** If they agree that college and career readiness entails far more than just a narrow set of academic skills and knowledge, then policymakers should ask themselves how well—or poorly—existing state and district assessments measure the full range of things that matter to students' long-term success.

   Further, policymakers should take stock of the real-world impacts that the existing assessment models have had on teaching and learning. For well over a decade, proponents of high-stakes testing have asserted that the prevailing model of accountability creates strong incentives for teachers and schools to improve. However, high-stakes testing is past due for an assessment of its own. State leaders should ask themselves: Are the existing tests, and their use in evaluating teacher and school performance, truly having the desired impact? In reality, what changes in instruction do teachers make in response to summative results and their use in evaluating their, and their schools', performance? How much time and money is currently devoted to such tests, and what might be the opportunity costs? That is, to what extent could high-stakes testing be crowding out other, more useful ways of assessing student progress?

3. **Support the development of new assessments of deeper learning.** Across the country, many efforts are now underway to create assessments that address a wide range of knowledge and skills, going well beyond reading and mathematics, and these efforts need to be encouraged and nurtured. However, several key

> » Can policymakers sustain their attention to this issue long enough to enact the policies necessary to bring about necessary changes? For that matter, can educators follow through with new programs and practices that turn policy goals into reality?

problems will need to be resolved if assessments of deeper learning are to be scalable, reliable, and useful enough to justify their expense. In particular, when it comes to measures that require students to report on their own progress—or that require teachers to rate students in some way—means will have to be developed by which to triangulate these reports against other data sources, in order to ensure a reasonable level of consistency. Further, it will be extremely important to institute safeguards to protect students' privacy and ensure that this sort of information is not used inappropriately. And, finally, policymakers and educators will have to be careful to distinguish between assessment tools that are meant to serve low-stakes, formative purposes—generating information that can be used to improve teaching and learning—and those that can fairly be used as the basis for summative judgments about students' learning or teachers' performance.

4. **Learn from past efforts to build statewide performance assessment systems.** States' pioneering efforts to develop performance assessments in the 1990s and early 2000s yielded a wealth of lessons that can inform current attempts to expand assessment beyond a limited set of tests. Most important is the need to proceed slowly at first, in order to develop systems by which to manage the sometimes-complex mechanics of collecting, analyzing, reporting, and using these types of richer information. Educators, especially, must have sufficient time to learn how to work with new assessments, not only how to score them but how to teach to them successfully.

5. **Take greater advantage of advances in information technology.** Many of the challenges that confronted states 25 years ago, when they first adopted performance assessment systems, can be addressed today through the use of vastly more sophisticated technology for information storage and retrieval. Online storage is plentiful and cheap, and it is far easier to move data electronically now than it was then. The technological literacy level of educators is higher, as are the capabilities of postsecondary institutions to receive information electronically. If districts and states take advantage of this new capacity to manage complex data in useful and user-friendly ways, they should find it much easier than in past decades to store student data in digital portfolios and access that information to meet the needs of audiences such as educators, admission officers, parents, students themselves, and perhaps potential employers.

6. **Adapt federal education policy to allow greater flexibility in the types of data that can be used to demonstrate student learning and growth.** The U.S. Department of Education's waiver process has introduced some flexibility with respect to the measures of student learning that states—and, in at least one case, a consortium of school districts—can use to meet federal accountability requirements. However, any reauthorization of the Elementary and Secondary Education Act and its NCLB provisions should go much further to encourage the use of multiple forms of assessment and to make clear to states that such models can pass federal muster.

7. **Consider using the National Assessment of Educational Progress as a baseline measure of student problem-solving capabilities.** The design of NAEP, particularly the fact that not all test-takers are asked to complete the entire battery of NAEP items, allows it to include fairly complex and time-intensive tasks. This design characteristic can be used both to field-test more complex performance items as well as to generate a better national metric of student problem-solving skills in the areas NAEP assesses. Having a baseline that is consistent across states can help determine which states are making the most progress with their statewide systems of assessment of deeper learning. PISA, too, could be used in this fashion, but the implementation challenges would be much greater than building upon NAEP's existing infrastructure.

8. **Build a strong base of support for a comprehensive system of assessments.** The process of developing a more complex system of assessments must not exclude any major group of stakeholders. Teachers in particular need to be centrally involved in designing, scoring, and determining how data from rich assessments of student learning will be used. State policymakers, too, have a compelling interest in finding ways to make sure that those assessments are both valid and reliable. And postsecondary and business leaders must have a seat at the table, as well, if they will be expected to make use of any new sources of information about students' college and career readiness.

9. **Determine the professional learning, curriculum, and resource needs of educators.** Currently, few states do much, if anything, to gauge schools' capacity to provide meaningful opportunities for professional learning. And as a result, most schools are unable to help their teachers acquire new skills. In order to implement any

new assessments successfully, it will be absolutely critical to determine—early on in the process—what resources will be necessary to ensure that all teachers are assessment literate, can use the information generated by multiple sources of assessment, are capable of developing assignments that lead to deeper learning, and can teach the full range of content and skills that prepare students to succeed in college and careers. It is worth noting that few state education departments or intermediate service agencies currently have the capacity to offer the level of guidance and support most schools, particularly those in smaller districts, need to undertake the type of professional learning program necessary to implement and use a system of assessments approach to instructional improvement.

10. **Look for ways to improve the Common Core State Standards and related assessments so that they become better measures of deeper learning.** This may be a tall order at a time when Common Core implementation is undergoing a rocky period. However, the surest way to undermine the credibility of the standards and the assessments would be to refuse to improve them in response to feedback from the field. Such a stance would only lead educators to view them as just another mandate to be complied with, rather than as a source of professional guidance and growth. Already, the standards are almost five years old, and it is past time to begin the lengthy process of designing and initiating a careful and systematic review process. Similarly, even though PARRC and SBAC are only just now completing their field testing, their designers must continue to seek out criticism, keep a close eye on their rollout, communicate more frankly and vocally the limitations of these assessments, while simultaneously suggesting ways to get at the various aspects of college and career readiness that these assessments currently overlook.

Ideally, the educational assessment system of the future will be analogous to a thorough, high-quality medical diagnostic procedure, rather than the cursory check-up described at the beginning of this paper. Educators and students alike will have at their disposal far more sophisticated and targeted tools to determine where they are succeeding, to show where they are falling short, and to point in the direction of how and what to improve. They will receive rich, accurate information about the cause of any learning problems, and not just the symptoms or the effects. Policymakers will understand that improved educational practice, just like improved health, is rarely achieved by compelling people to follow uniform practices or using data to threaten them but, rather, by creating the right mix of incentives and supports that motivate and reward desired actions, and that help all educational stakeholders to understand which outcomes are in their mutual best interests.

Research and experience make it clear that educational systems that can foster deeper learning among students must incorporate assessments that honor and embody these goals. New systems of assessment, connected to appropriate resources, learning opportunities, and productive visions of accountability, comprise a critical foundation for enabling students to meet the challenges that face them throughout their education and careers in the 21st century.

ENDNOTES

[1] It's always worth noting parenthetically that only one of the "Three R's" actually begins with the letter "r."

[2] The just-released version of the Standards for Educational and Psychological Testing takes up the issue of validity in greater depth, but test-development practices for the most part have not yet changed dramatically to reflect a greater sensitivity to validity issues.

[3] See: Wikipedia, "Structural Inequality in Education." Accessed September 9, 2014, from http://en.wikipedia.org/wiki/Structural_inequality_in_education

[4] See: "Criterion-referenced test," April, 30, 2014. Accessed September 9, 2014, from http://edglossary.org/criterion-referenced-test/

[5] These methods are still widely used, particularly by colleges themselves.

[6] Portions of this section are excerpted or adapted from: Conley, D.T. & L. Darling-Hammond. 2013. *Creating Systems of Assessment for Deeper Learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

[7] For a more detailed discussion of profiles, see: Conley, D.T. 2014. "New conceptions of college and career ready: A profile approach to admission." *The Journal of College Admission* (223).

REFERENCES

Achieve, The Education Trust, & Thomas B. Fordham Foundation. 2004. *The American Diploma Project: Ready or Not: Creating a High School Diploma that Counts*. Washington, DC: Achieve, Inc.

ACT. 2011. Accessed on November 23, 2011. http://www.act.org/standard/

ACT. 2014. Accessed on September 5, 2014. http://www.discoveractaspire.org/in-a-nutshell.html

American Educational Research Association (AERA), American Psychological Association (APA), & National Council for Measurement in Education (NCME). 2014. *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

Baldwin, M., Seburn, M., & Conley, David T. 2011. *External Validity of the College-Readiness Performance Assessment System (C-PAS)*. Paper presented at the 2011 American Educational Research Association Annual Conference New Orleans, LA Roundtable Discussion: Assessing College Readiness, Innovation, and Student Growth.

Bill & Melinda Gates Foundation. 2014. Accessed on September 5, 2014. http://nextgenlearning.org/grantee/summit-public-schools

Block, J.H. 1971. *Mastery Learning: Theory and Practice*. New York, NY: Holt, Rinehart, & Winston.

Bloom, B. 1971. *Mastery Learning*. New York, NY: Holt, Rinehart, & Winston.

Brandt, R. 1992/1993. "On Outcome-Based Education: A Conversation with Bill Spady." *Educational Leadership*. Vol. 50, No. 4.

Bransford, J.D., Brown, A.L., & Cocking, R.R., eds. 2000. *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy of Sciences.

Brown, R.S. & Conley, D.T. 2007. "Comparing State High School Assessments to Standards for Success in Entry-Level University Courses." *Journal of Education Assessment*. Vol.12, No. 2.

California Office to Reform Education (CORE). 2014. Accessed on September 5, 2014. http://coredistricts.org

Cawelti, G. 2006. "The Side Effects of NCLB." *Educational Leadership*. Vol. 64, No. 3.

Center for Educational Testing & Evaluation (CETE). 2014. Accessed on September 5, 2014. http://careerpathways.us/news/career-pathways-collaborative-offers-new-assessments-career-and-technical-education

College Board. 2006. Standards for College Success. New York, NY: The College Board.

College Board. 2014. Accessed on September 5, 2014. https://http://www.collegeboard.org/delivering-opportunity/sat/redesign

Conley, D.T. 2003. *Understanding University Success*. Eugene, OR: Center for Educational Policy Research, University of Oregon.

Conley, D.T. 2005. "Proficiency-Based Admissions." In W. Camara & E. Kimmell, eds. *Choosing Students: Higher Education Admission Tools for the 21st Century*. Mahwah, NJ: Lawrence Erlbaum.

Conley, D.T. 2007. *The College-readiness Performance Assessment System*. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T. 2011. *The Texas College and Career Readiness Initiative: Overview & Summary Report*. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T. 2014a. *The Common Core State Standards: Insight into their Development and Purpose*. Washington, DC: Council of Chief State School Officers.

Conley, D.T. 2014b. *Getting Ready for College, Careers, and the Common Core: What Every Educator Needs to Know*. San Francisco, CA: Jossey-Bass.

Conley, D.T. 2014c. *Learning Strategies as Metacognitive Factors: A Critical Review*. In Prepared for the Raikes Foundation, ed. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T., Aspengren, K., & Stout, O. 2006. *Advanced Placement Best Practices Study: Biology, Chemistry, Environmental Science, Physics, European History, US History, World History*. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T., Aspengren, K., Gallagher, K., & Nies, K. 2006a. *College Board Validity Study for Math*. Eugene, OR: College Board.

Conley, D.T., Aspengren, K., Gallagher, K., & Nies, K. 2006b. *College Board Validity Study for Science*. Eugene, OR: College Board.

Conley, D., Aspengren, K., Gallagher, K, Stout, O., Veach, D., & Stutz, D. 2006. *College Board Advanced Placement best practices course study*. Eugene, OR: Center for Educational Policy Research.

Conley, D.T., Brown, R. 2003. *Analyzing State High School Assessments to Determine Their Relationship to University Expectations for a Well-Prepared Student*. Paper presented at the American Educational Research Association, Chicago, IL.

Conley, D.T., Drummond, K.V., DeGonzalez, A., Rooseboom, J., & Stout, O. 2011. *Reaching the Goal: The Applicability and Importance of the Common Core State Standards to College and Career Readiness*. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T., McGaughy, C., Brown, D., van der Valk, A., & Young, B. 2009a. *Texas Career and Technical Education Career Pathways Analysis Study*. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T., McGaughy, C., Brown, D., van der valk, A., & Young, B. 2009b. *Validation Study III: Alignment of the Texas College and Career Readiness Standards with Courses in Two Career Pathways*. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T., McGaughy, C., Cadigan, K., Flynn, K., Forbes, J., & Veach, D. 2008. *Texas College and Career Readiness Initiative Phase II: Examining the Alignment between the Texas College and Career Readiness Standards and Entry-Level College Courses at Texas Postsecondary Institutions*. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T., McGaughy, C., Cadigan, K., Forbes, J., & Young, B. 2009c. *Texas College and Career Readiness Initiative: Texas Career and Technical Education Phase I Alignment Analysis Report*. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T., McGaughy, C., O'Shaughnessy, T., & Rivinus, E. 2007. *College-Readiness Performance Assessment System (C-PAS) Conceptual Model*. Eugene, OR: Educational Policy Improvement Center

Council for Aid to Education. 2014. Accessed on September 5, 2014. http://cae.org/participating-institutions/cwra-overview/

Council of Chief State School Officers (CCSSO) & National Governors Association (NGA). 2010. Accessed on September 6, 2010. http://www.corestandards.org/assets/CCSSI_ELA Standards.pdf

Council of Chief State School Officers & National Governors Association. 2010. Accessed on September 6, 2010. http://www.corestandards.org/assets/CCSSI_Math Standards.pdf

EdGlossary. 2014. Accessed on September 9, 2014. http://edglossary.org/criterion-referenced-test/

Donovan, M.S., Bransford, J.D., & Pellegrino, J.W., eds. 1999. *How People Learn: Bridging Research and Practice*. Washington, DC: National Academy Press.

Duckworth, A.L., Peterson, C. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology*. Vol. 92, No. 6.

Duckworth Lab. 2014. Accessed on September 5, 2014. https://sites.sas.upenn.edu/duckworth/pages/research

Dudley, M. 1997. "The Rise and Fall of a Statewide Assessment System." *English Journal*. Vol. 86, No.1.

Dweck, C.S., Walton, G.M., & Cohen, G.L. 2011. *Academic Tenacity: Mindsets and Skills that Promote Long-Term Learning*. Seattle, WA: Gates Foundation.

Educational Policy Improvement Center (EPIC). 2014. Accessed on September 5, 2014. https://collegeready.epiconline.org/info/thinkready.dot

Educational Policy Improvement Center. 2014. Accessed on September 5, 2014. https://collegeready.epiconline.org/info/thinkready.dot

Farrington, C.A. 2013. *Academic Mindsets as a Critical Component of Deeper Learning*. Chicago, IL: University of Chicago.

Education Week. 2013. Accessed on September 16, 2014. http://blogs.edweek.org/edweek/curriculum/2013/04/halt_high_stakes_linked_to_common_core.html

Education Week. 2014. Accessed on September 16, 2014. http://blogs.edweek.org/edweek/curriculum/2014/03/opting_out_of_testing.html?qs=parents+common+core+testing

Goodlad, J., Oakes, J. 1988. "We Must Offer Equal Access to Knowledge." *Educational Leadership*. Vol. 45, No. 5.

Guskey, T.R. 1980a. "Individualizing Within the Group-Centered Classroom: The Mastery Learning Model." *Teacher Education and Special Education*. Vol. 3, No. 4.

Guskey, T.R. 1980b. "Mastery Learning: Applying the Theory." *Theory Into Practice*. Vol. 19, No. 2.

Guskey, T.R. 1980c. "What Is Mastery Learning? And Why Do Educators Have Such Hopes For It?" *Instructor*. Vol. 90, No. 3.

Hambleton, R.K., Impara, J., Mehrens, W., & Plake, B.S. 2000. *Psychometric Review of the Maryland School Performance Assessment Program*. Annapolis, MD: Maryland State Department of Education.

Hinton, C., Fischer, K.W., & Glennon, C. 2012. *Mind, Brain, and Education*. Boston, MA: Jobs for the Future, Nellie Mae Education Foundation.

Horton, L. 1979. "Mastery Learning: Sound in Theory, but…" *Educational Leadership*. Vol. 37, No. 2.

Jennings, J. & Rentner, D.S. 2006. "Ten Big Effects of the No Child Left Behind Act on Public Schools." *Phi Delta Kappan*. Vol. 82, No. 2.

Jobs for the Future. 2005. *Education and Skills for the 21st Century: An Agenda for Action*. Boston, MA: Jobs for the Future.

Kirst, M.W., Mazzeo, C. 1996. "The Rise, Fall, and Rise of State Assessment in California, 1993-96." *Phi Delta Kappan*. Vol. 78, No. 4.

Knecht, D. 2007. "The Consortium and the Commissioner: A Grass Roots Tale of Fighting High Stakes Graduation Testing in New York." *Urban Review: Issues and Ideas in Public Education*. Vol. 39, No. 1.

Koretz, D., Stecher, B., & Deibert, E. 1993. *The Reliability of Scores from the 1992 Vermont Portfolio Assessment Program*. Santa Monica, CA: Center for the Study of Evaluation, University of California, Los Angeles.

Linn, R.L. 2005. "Conflicting Demands of No Child Left Behind and State Systems: Mixed Messages about School Performance." *Education Policy Analysis Archives*. Vol. 13, No. 33.

Linn, R.L., Baker, E.L., & Betenbenner, D.W. 2002. *Accountability Systems: Implications of Requirements of the No Child Left Behind Act of 2001*. Los Angeles, CA: Center for The Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, & University of California, Los Angeles.

MacCann, C., Duckworth, A.L., & Roberts, R.D. 2009. "Empirical Identification of the Major Facets of Conscientiousness." *Learning and Individual Differences*. Vol. 19, No. 4.

Medina, J. 2008. *Brain Rules: 12 Principles for Surviving and Thriving at Work, Home, and School*. Seattle, WA: Pear Press.

Mindsetworks. 2014. Accessed on September 5, 2014. http://www.mindsetworks.com

Mintrop, H., & Sunderman, G.L. 2009. "Predictable Failure of Federal Sanctions-Driven Accountability for School Improvement—and Why We May Retain It Anyway." *Educational Researcher*. Vol. 38, No. 5.

National Research Council. 2002. *Learning and Understanding: Improving Advanced Study of Mathematics and Science in U.S. High Schools*. Washington, DC: National Academy Press.

New Hampshire Department of Education. 2014. Accessed on September 5, 2014. http://www.education.nh.gov/assessment-systems/

Oakes, J. 1985. *Keeping Track: How Schools Structure Inequality*. New Haven, CT: Yale University Press.

Ohio Department of Education. 2014a. Accessed on September 5, 2014. http://education.ohio.gov/Topics/Testing/Next-Generation-Assessments/Ohio-Performance-Assessment-Pilot-Project-OPAPP

Oregon Department of Education. 2014b. Accessed on September 5, 2014. http://www.ode.state.or.us/search/page/?id=2042

Oregon State Department of Education Salem. 2005. *Career-Related Learning Standards and Extended Application Standard: Guide for Schools to Build Relevant and Rigorous Collections of Evidence*. Salem, OR: Oregon State Dept. of Education.

Partnership for Assessment of Readiness for College and Careers. 2014. Accessed on September 5, 2014. http://www.parcconline.org/sample-assessment-tasks

Pellegrino, J. & Hilton, M., eds. 2012. E*ducation for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: National Academy Press.

Rothman, R. 1995. "The Certificate of Initial Mastery." *Educational Leadership*. Vol. 52, No. 8.

Sawchuk, S. 2014. Education Week. Accessed on September 16, 2014. http://blogs.edweek.org/edweek/teacherbeat/2014/05/new_york_union_encourages_dist.html?qs=new+york+common+core

Seburn, M., Frain, S., & Conley, D.T. 2013. J*ob Training Programs Curriculum Study*. Washington, DC: National Assessment Governing Board, WestEd, Educational Policy Improvement Center.

Secretary's Commission on Achieving Necessary Skills. 1991. *What Work Requires of Schools: A SCANS Report for American 2000*. Washington, DC: U.S. Department of Labor.

Slavin, R.E. 1987. "Mastery Learning Reconsidered." *Review of Educational Research*. Vol. 57.

Smarter Balanced Assessment Consortium. 2014. Accessed on September 5, 2014. http://www.smarterbalanced.org/sample-items-and-performance-tasks/

Soland, J., Hamilton, L.S., & Stecher, B.M. 2013. *Measuring 21st Century Competencies: Guidance for Educators*. New York, NY: Asia Society/RAND Corporation.

Stanford Center for Assessment Learning and Equity (SCALE). 2014. Accessed on September 5, 2014. https://scale.stanford.edu/content/envision-schools-college-success-portfolio

Stecher, B.M., Rahn, M.L., Ruby, A., Alt, M.N., & Robyn, A. 1997. *Using Alternative Assessments in Vocational Education: Appendix B: Kentucky Instructional Results Information System*. Berkeley, CA: National Center for Research in Vocational Education.

Texas Higher Education Coordinating Board (THECB) & Educational Policy Improvement Center (EPIC). 2009. *Texas College and Career Readiness Standards*. Austin, TX: THECB & EPIC.

Tough, P. 2012. *How Children Succeed: Grit, Curiosity, and the Hidden Power of Character*. New York, NY: Houghton Mifflin Harcourt.

Tucker, M.S. 2014. *Fixing our National Accountability System*. Washington, DC: National Center on Education and the Economy.

Tyack, D.B. 1974. The One Best System: *A History of American Urban Education*. Cambridge, MA: Harvard University Press.

Tyack, D.B., Cuban, L. 1995. *Tinkering Toward Utopia: A Century of Public School Reform*. Cambridge, MA: Harvard University Press.

U. S. Department of Education. 2001. *No Child Left Behind*. Washington, DC: U.S Department of Education.

JOBS FOR THE FUTURE