Ten Principles of New Era Assessments
David T Conley, PhD

Incomplete developmental draft
To appear in:
New Era for Educational Assessment
Harvard University Press
Available Spring, 2018

Principle 1: New era assessment are focused on learner needs first and foremost. Students are actors, not objects.

This principle certainly receives lip service from test makers and educators, but the reality may be somewhat different than the ideal. Achievement-test designs view learners fundamentally as objects to be acted upon. Student interest in being tested and in the results from the test are largely secondary to the uses adults have for the information. Students may or may not be able to interpret and use the results.

Think how different things would be if the test itself were developed from the perspective of learners as individuals whose very individuality affects the way they experience the assessment task. Instead of trying to somehow sublimate that individuality, what if assessments sought to maximize it, to use the assessment process to gain insight not only into the knowledge and skills of individuals, but of the ways that knowledge and those skills relate to individual interests and the reasons for the performance that the assessment elicited? What if assessments were of interest to the student and if the student could use the results in some meaningful way to become better at something of value? In this alternative formulation of assessments, the learner’s needs are the initial driver and organizer for the assessment, and students have opportunities to be active participants who shape the assessment process, and not merely passive objects responding in ways that yield information that allows comparisons but may not be an accurate reflection of the value of the material to the test taker and other factors that shaped the test taker’s mindset and behaviors.

Here are three rules to use to judge the degree to which an assessment views the learner as an active participant and not an object.
Rule 1: The assessment does not stand alone. The information from the assessment is combined with other information to provide greater insight into the individual. This is the notion of a system of assessments. What it does is provide greater insight into the reasons students performed as they did. No one score is judged in isolation from all other information.

For example, a student might score poorly on a standardized achievement test, but the reason for doing so might be that the students is alienated from school and simply didn’t care. In this situation, the test score is fundamentally invalid, something that would not be known without the additional information. Similarly, a student who scores well on a standardized test might do quite poorly on a task that requires deeper understanding of the concepts and principles that underlie the content tested. The student was a savvy test taker, or the test items were relatively easy to decode, and the student did well based as much on those factors as on actual mastery of content at a deep level.

In these cases, additional information is necessary to explain the standardized test score. Putting this rule into practice would be easier if every standardized test score had to be accompanied by an explanation of why each test taker performed as she or he did. This would necessitate collecting additional information, including student self-explanations of performance and perhaps teacher predictions of student scores, and combining them into a score appendix that expressed the confidence that the score was an accurate reflection of student knowledge of test content. Ideally, students would then be provided an opportunity to respond to the proposed explanation of their performance. In doing so, this entire process engages the student as actor and generates invaluable diagnostic information that helps students, teachers, and parents develop plans to improve performance on subsequent exams.

Rule 2: Test takers have options. Those options might include the prompt to which they choose to respond or the way they formulate their strategy to solve a problem. For selected-response tests, learners might be presented with a wide range of items and be allowed to select the ones to which they wished to respond, in essence creating their own adaptive test. It’s difficult to understand why test takers are subjected to the humiliation of being unable to answer numerous questions when they themselves already know they cannot answer them and would choose not to do so if given the option.

The options extend beyond in-test choices. Students being assessed need to have choices regarding the means of assessment that best reflect their skills, abilities, and interests. In essence, with multiple means of assessment available, students would
be able to work with their teachers to devise more individualized plans of assessment consistent with their strengths, aspirations, and interests. Certain safeguards would be required to prevent teachers from guiding students inappropriately, and all students would need to be accumulating enough information to allow overall determinations regarding academic competency. This topic will be explored further in chapter xx.

Rule 3: Test takers get to express their opinion about the assessment. Currently, test developers, when creating new tests, may run cognitive labs where students answer questions and explain why they answered the way they did. A few might even hold a student focus group or two when introducing a brand-new assessment. But none include an option for students to provide direct feedback to the test developer and to those who use the test scores about their opinion of the validity of the assessment, the value of the knowledge and skills being tested, and the explanation for the reasons they performed as they did on the test. If this information had been gathered systematically during the development of the two consortium tests of the Common Core, PARCC and Smarter Balanced, it is possible that resistance the tests encountered might have been lessened or, at the least, that test developers and policy makers might have been better prepared for some of the reactions the tests engendered.

Patient-centered health care

The notion of focusing on the learner is akin to the idea of patient-centered health care, a concept that has gained a great deal of traction in the medical community over the past few decades. Until relatively recently, patients were viewed as objects to be acted upon by medical professionals. Little attention was paid to the role patients played in their own health and wellness, their values, preferences, and expressed needs. People who are ill feel powerless to begin with. Medical care that treats them as objects to be acted upon does little to engage patients in the healing process.

Patient-centered health care begins with the premise that patients know best how well health care providers are meeting their needs, that even though health care providers may deem a procedure or process a success, that judgment must be tempered by the experience of the patient and the patient’s final evaluation of its success and utility. Researchers¹ found no relationship between how expert the health care provider’s opinion was and the degree of positive results from the patient’s perspective. Patient-centered care resulted in quicker recovery and better emotional

health with less use of diagnostic tests and additional referrals. This type of care seeks to move from “what’s the matter with the patient” to “what matters to the patient.”

Similarly, new era assessment seeks to put the learner in the center of the teaching/learning/assessment process and consider what assessment looks like and how it feels from the learner perspective. This acknowledges the fact that the learner’s perspective on being assessed is an important facet of performance. Learner-centered assessment is not just a nice thing to do; it enhances the validity of results and has the potential to improve student performance on assessments.

Principle 2: New era assessments generate actionable information and in addition to any status measures or categorization.

Large scale assessments over the past century have been used extensively to determine where individual test takers stand in relation to the distribution of all test takers (norm-referenced) and to a designated standard (criterion-referenced). Both are useful things to know. This information alone, though, only provides the most general insight into what the learner needs to do next to improve position or status. New era assessments provide learners with value-added information that guides them on what they should do, not only how to better their performance on the same or similar tests next time, but on how they can learn better or think more deeply about what is being tested.

This principle extends to classroom-based assessment as well, much of which has been equally unhelpful in guiding learners. Whether it’s a letter grade or a point total, most of the testing and grading of assignments that goes on regularly in classrooms isn’t very specific about what learners need to do next. Sometimes this advice comes from the teacher individually and idiosyncratically to the student. Other times, students have to retake the test (or are given an opportunity to do so) with the assumption that performance will improve, once again with little guidance on how to master what they didn’t the first time.

Students get the scores they do on tests and assignments for a variety of reasons. Without knowing more about the reason they score as they do, it is difficult to provide guidance on how to improve. The assumption on the part of test makers is that any incorrect answer is a function of a lack of student understanding of the content or concept underlying the question. However, as most teachers know from experience, the reasons students perform as they do can be as varied as the students, and may or may not have anything to do with student understanding of the content itself.
Let’s examine the experiences of three students with differing scores on the same standardized test to see how valid a reflection those scores might be of the student’s knowledge and skills, and, more importantly, what those scores tell us about what the students need to do to improve their scores. The thing to watch for in these examples is the degree to which the assessment is generating actionable information and the degree to which each student’s subsequent test performance is likely to improve based on the information they get back from the test on their performance.

First a bit about the test itself. The test was constructed meticulously following all the current rules and principles of test design. From a technical point of view, the test meets all applicable standards. The feedback it provides is both descriptive and analytic. Test takers see where their scores place them relative to all other test takers in terms of a percentile rank and whether their performance ranks them as emerging, proficient, or advanced. In addition, they see each question they answered incorrectly, the correct answer, and what it was testing. Finally, the report students receive contains recommendations for content they should learn better, based on items missed and performance level achieved. Will this information be helpful to our three students?

Maria scored well on tests. She always took them seriously, even when she didn’t know what the purpose was. She often made it into a game. She liked to figure out what test-makers were thinking when they put the items together. When she got her scores back, she would review them, largely to see if they lined up with her perception of how well she did. She rarely, if ever, did anything as a result of the scores, but she did enjoy getting them back because she almost always did well.

Let’s compare Maria’s experience with Brayden’s. Brayden did reasonably well in school. His hallmark was doing everything fast. This meant he had time to goof off or talk with friends, which he enjoyed. The result was that he often made careless errors in his work. His teacher took this into account when grading homework and class tests. However, when it came time to take a standardized test, Brayden’s penchant for speed often tripped him up. It didn’t help that students were allowed to put on their headphones and listen to music after completing the test. He would be among the first to complete the test and then sit back and enjoy his tunes. His scores were always decent but not outstanding. He, like Maria, did little or nothing differently as a result of the scores. He would look at the error analysis report and say, “Aww, I know that…” His teachers tended to agree, and took into account that part of the explanation was Brayden’s tendency to make errors from not reading questions carefully or selecting an incorrect answer that was very similar to the correct answer and differed only on a small detail, such as placement of a decimal point or exponent.
Finally, we consider one final student, Amber, who was academically capable but never performed well on tests. Amber did not like her teacher or most of her classmates very much. She remained withdrawn and sullen much of the time. She thought school was “dumb.” She did not have any specific aspirations or goals in her life except not to be in school. When she took tests, she only looked for obviously correct answers. If a question took her more than ten seconds or so to figure out, she just made her best guess and moved on. When she would get her scores back and they were not good, she would glance at them and then unceremoniously dispose of the score report without seriously considering how to improve her scores. After all, they were just another part of “school,” which had scant legitimacy in her eyes to begin with.

Based on the description of these students, how likely is it that they will use the information provided in the score reports to change their performance? How useful is that information to them? How relevant and explanatory of performance is the information? What is the likelihood of reaching a false conclusion about the student’s potential for future academic success in the subject area based on this information alone and in isolation from everything else a teacher might know about these students?

If we were to judge each of these students on their scores on a single achievement test and never meet them, we would conclude that they had three very different levels of achievement and, by inference, different levels of aptitude. Their teachers might have known this was not necessarily true. They may have observed Maria doing quite ordinary work in class but demonstrating real savvy about how schooling works and how to do well in the classroom without really trying too hard. They may have realized Brayden had a streak of real creativity and insight in him, that he grasped almost everything quickly, and that he had superb people skills. They may have recognized that Amber had solid foundational skills and was perfectly capable of learning what she was being taught but was choosing not to. It was clear that Amber needed a reason to do schoolwork, something Maria did not need. She also needed to care about the results enough to want to do something to improve her scores.

Each of these three example students received scores that either did not reflect well their true level of knowledge and skill or that did not provide any insight into why they scored as they did. In addition, none of these students found any great reason to be motivated to improve in areas where they did not perform well on the test. Their teachers understood that the test results provided only very limited glimpses into their mastery of the curriculum, and that simply addressing the items they got wrong would not necessarily improve their performance next time.
New report types

Is this the fault of the test? Not at all. A test of this nature is not designed to or capable of ascertaining student motivations, test-taking strategies, or attitudes toward authority figures and schooling in general. Is it possible to get to truly actionable information off of the results of a single test? It’s not easy. What follows is a brief, hypothetical description of a system that would potentially provide more such information to students and teachers. Would it be perfect and work for all students? No. Can any single test ever generate all necessary information to fully explain the reasons test takers score as they do? No. That is why the notion of multiple measures is so central to a new era system of assessments model.

But what could actionable information look like for these students? To begin with, the score report itself may have to be designed as a separate learning experience and score reports may have to be distributed online in order to do what is being described next. To begin with, after completing the test, students would respond to a set of questions designed to gauge their self-perception of the reasons for their performance. This would be completely confidential and not for high stakes. Results from this self-report activity could be aggregated anonymously to garner some insight into test validity, but the primary purpose would be to let test takers think about the reasons for their performance. Had they been taught the material? Had they understood the material in other contexts but didn’t recognize it when placed into the format used on the test? Did they have sufficient time to complete the test, or did they feel rushed? Did they know why they were being given the test and what the consequences, if any, there were for their score? Did they try their best? Were they just unable to find any relevance to their goals and aspirations in the material being tested? Were they feeling ill on the day of the test? Did something happen in their personal lives that made it difficult for them to concentrate on the test at the time it was given? These are examples of questions that would elicit greater insight into the mindset of test takers, which in turn helps explain any limitations of the validity of the results (without calling into question the technical quality of the test itself).

When students received their results, they would also get an analysis of personal factors affecting test performance. In combination, these two sets of data would form the basis for recommendations on how to address these factors and improve future test performance. Examples of recommendations might be to talk with a counselor or teacher about anything that tends to inhibit their ability to perform well on tests of this nature, and to get help in the future in these areas to be as prepared as possible for the next test. Another recommendation might be simply to slow down and check their
work more carefully. Other suggestions could be to talk with teachers about how to connect what they are learning to their areas of interest so that testing seems more relevant to them and to do all they can to ensure they are ready on the day of the test next time, if they didn’t feel ready this time.

If they indicated they had not been taught the material or that they did not understand it well, this is information of importance to the teacher.

Additional information can be collected during the testing session itself in the case of online administrations. For example, how long did the test taker spend on each item? Does this provide any insight for the student into which topics were understood well and which items might have been answered correctly because of a lucky guess? Similarly, on which items were responses changed, and were the changes better than the original choice? Knowing this helps test takers check their own instincts and become better at discerning when they actually know something they think they know.

Tests with writing samples are more challenging to provide better feedback to students than the kind of rubric-based feedback they receive currently. The costs are generally prohibitive. One simple way to illustrate how to improve writing to the next level on the rubric is to include in the score report examples of either complete pieces of work at the next rubric level above where the test taker scored, or to present paragraphs that illustrate specific improvements in areas of the essay such as introductions and conclusions.

With this information in hand, the student report would be able to consist of recommendations that extend beyond specifications of topics to re-learn. It could include tips on how to understand material at a deeper conceptual level (Maria), learning strategies and techniques to improve accuracy (Braylen), and for Amber, some encouragement to engage more fully in school and to make connections between her interests and the content being tested. For other students, feedback might tell them whether they know important foundational material at an automatic level, even if they got it right. Can they compute fast enough or ready efficiently enough to be able to meet subsequent academic demands, even if they are getting answers right?

Clearly, this type of report is far more detailed and difficult to produce than anything test makers offer today, but this is the whole idea of a new era of educational assessment—trying new things and taking advantage of new technologies and new views of learners and how they learn. This new era score report does not demand major changes to large-scale, standardized tests. It does require much more thoughtful analysis of student performance that includes student input along with more
information from the item responses and response patterns than is extracted currently. Score reports have changed little in the past 50 years. They are an area that has been neglected as a tool to enhance student ownership of learning and engagement in self-improvement.

**Principle 3: New era assessments have high cumulative validity.**

As noted in chapter xx, most assessment is a tug-of-war between reliability and validity. It’s difficult to achieve high levels of both on any single assessment. For new era assessments, validity is enhanced by being cumulative in nature. In other words, all of the assessments and measures of student performance are valid reflections of important things students are expected to know and be able to do, and their validity is enhanced and amplified as results are accumulated across multiple measures and over time.

This idea isn’t particularly new or novel. Unfortunately, the educational system has not done a good job of triangulating the results from various measures of performance nor of gauging student development over time. Neither have a variety of assessment measures been used. The result is a tendency to have measures cluster at each extremes: higher reliability and relatively lower validity (in terms of measuring what students are being taught and experiencing in their classrooms), or relatively lower reliability (for measures such as teacher-created quizzes, exams, assignments and projects) and higher validity.

Standardized achievement tests are perhaps the most reliable measures available in their ability to garner comparable student performances across multiple administrations of an instrument. Their validity has been called increasingly into question, particularly as standards systems attempting to specify more complex phenomena such as college and career readiness versus basic skills incorporate concepts and skills that may not be gauged well or at all by standardized tests. As you review these example standards from the Common Core, think about what it would take to assess them validly:

- Conduct research and synthesizing information
- Develop and evaluating claims
- Read critically reading 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

The Common Core and many state systems also include standards for speaking and listening, which, while not impossible to test in a standardized fashion, do sacrifice a modicum of validity in order to do so at scale. Additionally, many college and career readiness standards along with the Common Core emphasize research skills. While it is not impossible to include items on a standardized test that get at student knowledge of the research process, this is another skill that is best measured as a whole, in the form of a research project.

What, then, does a new era model have to say about these more complex, difficult-to-measure skills? In short, they require careful attention to their nature. Performance needs to be captured in context. Complex skills generally have to be measured multiple times over a period of time, both to support their evolution and to confirm that the performance measured reflects the actual mastery level.

Educators and policy makers often don’t like to hear or acknowledge this. At the same time, they adopt standards that emphasize these exact complex skills as the core goals of schooling. This tension is explored in greater depth in chapter xx. It is acknowledge here only to make it clear that some of the more valid forms of assessment should not be dismissed out of hand because they take more time or are more difficult to score. Validity is a cornerstone of the new era of assessment, and it will be necessary to become better at administering, scoring, and recording information from a wider range of valid assessments in order to encourage students to develop complex skills that are key to success in a knowledge-oriented economy.

Projects are often presented as a counterpoint to standardized assessments, particularly projects developed by teachers to capture what they have been teaching in their classrooms. Projects as assessments pose many technical and administrative challenges. Developing high quality projects is not easy, and it can take a great deal of time and testing to perfect a single project. Reliable scoring can be achieved, but it is not easy to do. Perhaps most challenging is that projects take time and cannot be done at a single sitting. This eliminates them as a direct replacement for standardized on-demand tests. They have other limitations as well that will be considered in chapter xx. For the time being it is worth noting that projects are not the be-all and end-all, and they do not represent in and of themselves the new era of educational assessment.
They have an important role to play, but only within a larger system of assessments context.

The importance and value of student attitudes and beliefs have recently been emphasized to a greater degree. These go under many names, such as social emotional learning, soft skills, metacognitive skills, success skills, and non-cognitive skills. All seek to develop student capabilities as learners. Some also seek to cultivate the capabilities and characteristics necessary for students to be successful in life beyond school. Measuring these skills is challenging but not impossible. They have high levels of validity, and it is even possible to establish reliability by triangulating self-reports against other sources of information and performance.

The challenge has been in using them for high stakes purposes, which tends to raise concerns about faking. For the time being, it is perhaps best to consider self-reports as useful supplemental information that has great value for ongoing improvement of student learning skills and that leads to enhanced self-awareness. The information can also be useful to teachers to help explain performance issues for individuals or to develop key learning skills for all students.

Advocates of standardized tests and of classroom-based approaches tend not to consider how to combine them. They also rarely incorporate other readily available measures such as attendance, student goals and aspirations, and personal management skills such as being on time to class and turning in assignments on time. Some schools and districts have developed “dashboards” that display multiple indicators, but few have connected measures in causal chains that explain and predict student performance. The profile approach to integrating data in ways that reveal more about the student than do any individual score will be considered in more detail in chapter xx. For the moment, it is enough to note that they will be challenging to implement, but confronting those very challenges is what will enable schools, districts, and states to take a significant step toward a new era model of assessment.

Principle 4: New era assessments view student development along a novice-expert continuum.

Current era assessments have struggled to place students on scales that cut across grade levels. It’s a truly daunting task, and many competent thinkers and practitioners in the field believe it’s not really possible to do. They assert that students change so much cognitively that using content-knowledge mastery as an indicator of progress on the scale fails to recognize the changes in content complexity. Additionally, some argue that content as taught in most schools isn’t really a neat,
logical progression of knowledge, but rather a series of loosely connected topics, concepts, and techniques that could be taught in any of a number of different orders or sequences, and is in fact taught in different order in other countries’ educational systems. In other words, it’s hard to make a case for a natural, immutable progression of content knowledge that would form the structure for a continuous scale of student scores such that results on tests of the subject from each grade level could be neatly situated on the same numeric scale, thereby showing progress in equal increments across a student’s educational history and being able to gauge how far ahead or behind a student is by using the scale score as an indicator of where the student is in terms of grade-level performance.

Needless to say, others strongly believe that scales can be created that reflect the reality of student performance across grade levels, more or less regardless of the content itself. High quality scales of this nature have been created and are in general use. They meet a number of needs and are very quick and handy ways to size up where a student stands relative to mastery of a body of knowledge that is taught across multiple grades and age levels.

New era assessments take a slightly different approach to the notion of a cross-grade scale. This notion need not replace the content-based scales in use, but can help provide context-based information about the level of student mastery of a subject area.

These scales are based on the progression of the learner from novice to expert. Students can move from showing the types of skill associated with a novice at each appropriate age and developmental level. A novice third grader is different from a novice 11th grader in terms of the content they are learning, but the behaviors associated with a novice are the same. Here is an example of a novice-expert scale that can be used to track student mastery of content regardless of the specific content or the order in which it is taught:

- **Emerging novice**: Follows wrong rules or no rules at all; is highly inefficient, redundant, confused; shows little evidence of proper use of conventional ideas; does not incorporate concepts and/or does not explain findings coherently; shows almost no integration or connection among elements; fails to show a solution strategy.

- **Novice**: Follows subject-area rules literally; is inefficient in the treatment of the task; shows some evidence of use of conventional ideas; presents findings without using concepts significantly; elements are not well integrated or connected; falls well short of a solution.
• Accomplished novice: Follows subject-area rules correctly but demonstrates limited insight into the subject area; has areas of efficiency along with significant inefficiencies; shows consistent evidence of proper use of conventional ideas; uses and incorporates concepts in a limited fashion to organize and explain findings; shows limited evidence of integration or connection among all elements and has at least one clear problem area; falls short of a complete solution strategy.

• Emerging competent: Shows some evidence of applying subject-area rules with insight beyond literal application; shows efficiency in completing task but with several areas that could be improved; shows consistent evidence of proper use of conventional ideas with at least some evidence of original ideas or variations on conventional ideas; uses and incorporates concepts to explain findings with minor inconsistency; shows evidence of integration or connection among elements with some problem areas remaining; comes very close to a complete solution strategy.

• Competent: Shows evidence of applying subject-area rules in an insightful fashion beyond literal application; is predominantly efficient with some inefficiency remaining; shows strong evidence of conventional ideas used properly and some novel or original ideas; incorporates a set of core concepts to explain findings; has convincing evidence of integration or connection among its elements; shows evidence of a full and complete solution strategy.

• Accomplished competent: Shows evidence of a more intuitive use of subject-area rules to gain insight beyond literal application; treats task efficiently with only minor inefficiencies; shows strong evidence of novel or creative use of conventional ideas and/or evidence of original ideas; is clearly structured around a set of core concepts to explain findings; is well integrated and connected; shows evidence of a cogent, coherent solution strategy.

• Emerging expert: Shows strong evidence of an intuitive and insightful use of subject-area rules; treats the task highly efficiently; shows novel or creative use of conventional ideas and/or unique or innovative ideas; consciously organizes and explains findings around a set of core
concepts; integrates and connects all task elements highly effectively; shows strong evidence of a cogent, coherent solution strategy.

- Expert: Understands subject-area rules deeply and applies them intuitively, even to the point of breaking rules when it makes more sense to do so in order to achieve a purpose; questions and challenges conventional ideas and has a point of view about controversial topics in the subject area; presents new interpretations of core concepts of the subject area; naturally integrates task elements as a framework for making a larger point or reaching a more substantial conclusion; presents novel and insightful solutions to complex and core problems and questions of the subject area.

The primary prerequisite to being able to applying a scale like this is having learning tasks that elicit the levels of thinking specified in the scale. When curriculum and instruction cause students to engage with content in ways that get them to process what they are learning, mastery can be specified with a cognitive model that is based on degree of expertise demonstrated. This is important because more sophisticated thinking about content almost always represents better mastery.

Virtually no students can expect to function at an expert level while in primary or secondary school. Students simply don’t have enough time and experience with the material to develop the kind of thinking an expert demonstrates. The goal should not be to get all students to the expert level by the end of high school. The target should be somewhere between emerging competent and competent in core subject areas. If this can be achieved, students will enter college and career preparation with strong foundations in the way they think about the knowledge and concepts that comprise academic disciplines. Such a foundation is what is necessary to develop, with practice and with new and novel opportunities, the kind of thinking that characterizes an expert in an area. Reaching the level of expert in at least one area in their lives will be important for success in today’s economy and society, keeping in mind that expertise does not necessarily mean specialization. An expert can be someone who understands how to solve problems in an area, not just how to apply knowledge to specific problems.

The novice-expert framework provides a way to ascertain how well student thinking skills are progressing, but within the context of the content they are learning, not independent of it. It does necessitate being able to say that a student has not reached the highest level on the continuum, something that may be difficult for many teachers (and parents) in environments where students are accustomed to received A’s
on all assignments and where they strive to maintain a 4.0 grade point average. The novice-expert scale is such that a student can be competent in sixth grade, for example, and remain at the competent level through twelfth grade, and still be learning much new material and developing a greater understanding of the subject area. Competence is determined in reference to the material being learned, how complex and challenging it is relative to the learner’s developmental stage and capabilities. This is a very different notion than a scale built on sequential mastery of content knowledge.

Principle 5: New era assessments ascertain learner understanding of the structure of knowledge in addition to mastery of factual information.

Most current era assessments are based on the assumption that if a test taker can demonstrate knowledge of a series of pieces of knowledge from a larger domain, then that test taker understands the domain. This inference jumps over the substantial problem that a test taker can know the parts but not be able to assemble them into a whole. Doing so requires deep understanding of the structure of knowledge in a subject area or academic discipline. Such understanding goes well beyond being able to decipher the intentions of item writers. It requires understanding of concepts and of how bits of information are organized and categorized in the subject area, and then how they relate to the foundational concepts and ideas that combine to create the discipline.

Grant Wiggins and Jay McTighe\textsuperscript{2} in their 1998 work, *Understanding by Design*, lay out a knowledge taxonomy that demonstrates the relationship between bits of factual knowledge and larger concepts and ideas. For example, they describe how subject areas are organized around foundational “understandings.” These lead to “essential questions” that cause the learner to engage in ways that lead to greater insight into the understandings of the discipline. For example, one of the understandings of geography is the relationship between humans and the physical characteristics and features of the areas in which they choose to live or choose not to live. This understanding leads to questions about the trade-offs of living in certain places versus others, to the effects of forces such as natural disasters on human settlements, and to the factors that make it easier or harder for humans to thrive in particular locations. This foundational understanding then frames the acquisition of factual information such as the names of geographical features such as mountains,


plains, peninsulas, archipelagos, bays, and deltas. These in categories in turn are the organizers for the names of specific places, such as San Francisco Bay or the Ural Mountains. If geography is thought of only as the ability to locate places on a map, students do not acquire insight into or understanding of the reasons they are learning the names of places and the categories of geographical features. Including the “big ideas” and foundational understandings of the subject area enable learners’ brains to create the schema, or structures of knowledge, necessary to retain the factual information.

To get a sense of the power of how understanding the structure of knowledge makes a difference, think about grammar, the bane of many an otherwise well-educated person. Why do so many people struggle with the grammar of the English language (beside the fact that the language is composed of parts of several different languages welded together 800 years ago)? Much of it has to do with how people typically learn grammar. They are taught the parts of speech and what the parts of speech do. Nouns name things, verbs are action words, etc. Then, they are expected to put them together into sentences and paragraphs and to analyze written material through this lens.

What is lacking for many people is an understanding of the structure of the language, how words combine into sentences that can contain phrases, how parts of a sentence can be in the form of clauses, and how sentences combine into paragraphs. This isn’t as easy as it sounds because sometimes the rules can be violated and still work, and because some words serve multiple purposes depending on their context. Finally, language is dynamic, and its use in context is also important. In short, grammar is merely a means to an end, and comprehending the nature of the written or spoken piece influences what’s “correct” and “incorrect.” Understanding the structure of grammar takes time, thought, and an expert’s mindset. For someone trying to learn grammar through the rule-based methods of a novice, it never really makes much sense, or it is endlessly frustrating. Teaching the concepts that inform the creation of the structure of the language helps learners organize the myriad terms, terminology, and rules they are taught.

New era assessment acknowledges the importance of factual knowledge as a means to an end, not an end in itself. That end is the deeper understanding of the subject area and, as we will see in the next principle, the ability to apply that knowledge in meaningful and constructive ways to interesting and challenging problems.
Principle 6: New era assessments provide insight into the integration of knowledge by learners and allow learners to apply knowledge in context.

Closely related to showing an understanding of the structure of knowledge is the ability to integrate knowledge across subject areas and then apply that knowledge in new and novel ways to interesting and challenging problems. Cognitive scientists suggest that this ability to transfer knowledge from one setting to another is among the most challenging tasks for a learner. It may help explain why, when the science teacher tells students that their lab reports will be graded on spelling and grammar in addition to content that students complain about the utter unfairness of it all because, “This isn’t English class.”

Mathematics is another area where instruction and learning is so context-bound that students consequently struggle to apply what they learn in other settings. College science instructors recount the difficulty of getting first-year students to apply their knowledge of algebra to generate hypotheses and analyze data. Perhaps most classically, students who earn A’s in their foreign language courses are completely unable to communicate with a speaker of the language when encountered outside of the classroom.

This is because application and transfer aren’t easy. Building those capabilities requires assessments that get the learner to apply what they know in one area in the service of addressing a task or solving a problem in another area. This creates a dilemma for test makers because they now have what they describe as a multidimensional assessment. In other words, the assessment is not measuring just one thing (content knowledge taught in that subject area); it is also gauging the application of knowledge acquired elsewhere. So, the science task that requires algebra to solve it is also testing how well the test taker acquired algebra skills in a mathematics class previously.

While multidimensional assessments are by no means impossible, they are more complicated to administer and score. They also create problems when used in a high-stakes situation such as contributing to a teacher’s evaluation. Are teachers being judged on what was taught in their classrooms or in someone else’s?

New era assessments intentionally create this problem of multidimensionality. They emphasize the transfer of knowledge and its application to problems that may not have one right answer. Test developers are finding ways to incorporate these types of items into standardized tests (examples). In addition, classroom teachers often use these types of tasks (examples).
Principle 7: New era assessments inform instruction and learning more directly.

Assessment information must be useful and valuable to teachers for improving instruction and schools for improving the instructional program. As much as test providers would like to think otherwise, teachers rarely make good use of data from large-scale assessments. It’s not that this information is not useful. It’s that the assessment was never designed to provide diagnostic classroom information to teachers. Many tests have added such features after the fact, and the best ones are making some progress on this front. The issue will always be that the tests themselves are not constructed in ways that can reflect with fidelity what teachers are doing in their classrooms. Teachers are gathering and using data on student performance all the time, through informal observations, verbal interactions, in-class activities, assignments, quizzes, and tests. This information, however, is not congruent with the results of external exams.

Other countries have developed examination systems that reflect the curriculum, in part because the curriculum is standardized nationally. This will never occur in the U.S., so the solution for how to create alignment between teaching and testing will need to take other forms, some of which will be explored in subsequent chapters. The principle will be the same: making assessment information useful to teachers and reflective of what they are teaching.

One of the perpetual problems that policy makers face is that they do not have good measures of what is happening in the educational system. Over the past 15 years, they have come to rely increasingly on math and reading scores from standardized tests. While this information is certainly useful, it is not particularly helpful in figuring out the system changes that need to be made to bring about significant, sustained improvement in student performance. As a result, most schools, particularly high schools, are about where they were 15 years ago in terms of student achievement. Some have improved and then fallen back again, but it is the rare school indeed that dramatically improved student performance as measured on standardized tests and then sustained that improvement (without some external factor such as a change in student demographics).

True systems improvement will require a combination of information, much external to the classroom, that helps establish the conditions of education necessary for success and the presence or absence of such conditions. Then, information on what’s happening in the classroom can be of greater use and value. This is the point at which such information will need to have already met student and teacher needs in
some form. This type of information will have demonstrated its validity and value, and will be suitable to include in accountability systems or, as an alternative, systems improvement models that do not entail punitive actions toward schools or educators.

**Principle 8: New era assessments promote student ownership of learning and help students identify interests and develop self-knowledge.**

It’s perhaps jarring at first to think of assessments as a means to promote student ownership of learning and increase self-knowledge. Yes, a score (or a grade) tells students something about their knowledge and skills, but this is in relation to other students who took the test, to the teacher or test-maker’s standards, or to set of defined criteria, as in the case of a rubric or scoring guide. What this principle addresses is the ability of assessment to let students take greater control over what they learn and how they learn it and to provide personal insight into the interests of and opportunities for individual students.

None of this works if the curriculum, standards, and instructional methods the student experiences do not promote these goals. So, new era assessments are dependent in this case to the environment in which they are situated. Schools and classrooms in which this type of assessment fits are not organized around content transmission and acquisition primarily, although students are expected to achieve high levels of content mastery. Such classrooms are devoted to helping students learn more about themselves and to explore their possible selves as well as the possibilities of the world they will be entering.

This conceptualization of the purpose of schooling and the role of the learner is a radical reconceptualization of the current era model where students are consumers of content and standards identify all useful and desirable outcomes. This new model views schooling’s most important goal to be to tie student knowledge acquisition to their discovery of how knowledge can be used to pursue goals and achieve aims. Furthermore, learning is constantly inviting learners to consider the futures that are possible for them.

This is not the same as choosing a career pathway as a ninth-grader and then taking an elective course that loosely relate to that area. This approach infuses development of self-knowledge, not just career selection, into all aspects of learning. As learners progress through school, they engage in assignments and activities that cause them to apply what they are learning in various contexts, some of their choosing,
some of their teacher’s choosing. Each time they investigate an area or apply knowledge purposefully, assessments generate information about the student’s experience with that area. Students add to this by recording their own reactions and insights into the fit between the act of applying what they are learning and their sense of self. What did this learning experience tell them about themselves and about their interests?

The goal is that as students progress through school, they become much more self-aware and demonstrate more intentionality in the learning process. At the same time, they think more about their future, what they need to do to be ready for that future, and what avenues best help get them there post-high school. The goal is to have students leave high school with much greater clarity about what they want to do next.

Some will interpret this as forcing children to choose their life’s work or occupation, which is not at all what is intended. Instead, new era assessment can empower students to make more decisions about the path their learning takes and where it leads them. This builds ownership of learning, the idea that I, the student, have ultimate responsibility for what I get out of the learning experience, and that it’s worth my time and effort to get all I can out of each learning experience that I choose to own.

Ownership of learning is an incredibly powerful motivating force and something unfamiliar to most students, who adopt a compliance-based approach as they march from grade to grade. Tell me what I need to do to get an A/get out of this class, and I’ll do it, assuming it’s not too onerous. All the while, I’ll try to get you to lower your standards and expectations for me, and I’ll invest as little as I need to in this enterprise. Test scores and marks toward my grade will be viewed transactionally in terms of what they get me in proportion to how much effort I have to expend. I will not expect to learn much about myself, my interests, or what I want to do with my life. But that’s okay; it’s not the goal in the first place. Most of those things happen outside of school for me anyway, increasingly in online environments that I can choose and where I can learn about myself.

Principle 9: New era assessments use data to create profiles of student as learners in ways that aid individual development and goal attainment.

Current models for collecting information on student performance are firmly rooted in the early 20th century. The standard student transcript is hobbled by its lack of flexibility, adaptability, and versatility. Some of the reason for this is that schools
have a difficult time managing information about student performance and getting it into a transcript in the first place. Another major reason is that the recipients of transcripts, generally postsecondary institutions, do not ask for or expect much beyond what they receive currently. Colleges have been efficient at extracting as much information as they can from transcripts and using this for a few basic, but important, functions, including as a key contributing factor in admission decisions at institutions with selective admission processes.

As noted earlier, information about individuals is being assembled to create more complex and complete pictures of them. Predictably, the leaders in doing so are retailers who can benefit financially by knowing much more about their shoppers’ habits, attitudes, preferences, and backgrounds. They encourage people to give them this information through loyalty and discount programs as well as by buying data. They build complex predictive models that enable them to offer the right product to the right person at the right price at the right time. These models will become even more powerful in the future as more of human activity occurs online where it can be captured and analyzed.

Not everyone thinks all of this is necessarily a good thing, but it is happening nonetheless. Privacy advocates attempt to prevent abuses and to guarantee safeguards for data collected. Some people opt out or find ways to avoid having their data collected. But for many, this information in the fashion it is organized is a help and an assistance to them.

The intent here is not to settle this debate. It is to suggest that schools have fallen far behind other segments of society in assembling information into a useful form that is beneficial to those who generate the data in the first place. With proper safeguards and options to opt out in place, data systems that collect and organize a wider range of information about students can be highly beneficial tools to identify patterns, predict problems, spot opportunities, and paint a fuller picture of the individual that can be used, with the individual’s consent, for a wider range of purposes including providing postsecondary institutions with more useful information to help students succeed in college, not just to admit them.

The profiles of students that new era systems of assessment generate hold traditional information on academic performance and choices. They also aggregate test scores and contain more information on the components of teacher-generated grades. They may also include performance data from outside of school, such as proficiency standards met through organizations such as Girl Scouts, Boy Scouts, Eagle Scouts, certifications attained such as first aid training completed, Special Olympics
awards, sports achievements, spelling bees, as well as participation recognition such as volunteering in the local community or serving as a student helper. Many of these elements are included in the college applications and resumes of high performing students, but they are not typically collected for the general student population.

Many vendors and a number of schools have created electronic portfolios in which students collect and track their schoolwork over time and across grades. These e-portfolios may even have room in them for some of the extra-curricular elements described above. Generally, however, they concentrate on student academic work and become repositories of primary documents demonstrating academic development and skill. This information is also central to a student’s profile, and it can be considered more in context when viewed alongside other evidence of accomplishments and interests.

A third element of the profile is student-generated information about themselves. This includes self-reports, journals, reflections, personal plans, goals and aspirations, interests, and self-analyses. This information is designed to frame information on academic achievements and extra-curricular accomplishments.

In total, the profile is designed to provide insight into the nexus between an individual student’s progress and their aspirations and destination. This process highlights the personal growth and development that is occurring along the way. In some cases, it also calls attention to a lack of focus and aspiration, which is also important to know, recognize, and address.

Ultimately, the profile is a guide to how well students are achieving their potential, an elusive yet near-universal goal of education. The profile encourages students to review their progress regularly and assess their behavior in relation to their goals and aspirations. Are their actions helping them achieve their goals? Are their goals realistic and achievable? Do they need to reassess their goals in relation to their behaviors? The profile review process can be highly affirming and also an exercise in self-examination. Students have a role in figuring out how well they are achieving their potential and what they need to do (and how their education can help them) to fulfill their potential.

This type of engagement encourages increased ownership of learning, which fosters better decision making. Clearly, the educational process needs to give students the opportunity to fulfill their potential, which leads to the issue of equity, which is the final principle.
Principle 10: New era assessments are conscious of equity issues and seek to address them.

Historically assessments have not been seen as vehicles to address equity issues or in most cases even to acknowledge equity as a factor that may affect performance or score interpretation. This was the result of several factors discussed earlier including the reliance on the creation of “normal” distributions of scores, an inability to distinguish whether a student did not know something because they did not understand it or they were never taught it in the first place, and, until recently, a lack of score breakdowns by subgroups such as race/ethnicity, gender, special needs status, or parental income. Additionally, until relatively recently, test items were not screened for issues such as cultural assumptions or differential responses by different groups of students.

Test developers would argue that their tests were not inequitable; it was the circumstances under which they were administered that was the problem. The developers could not be expected to anticipate or compensate for differential opportunities for students to learn the material on the test, for example. Classroom-based grading systems also had problems with fairness when they were heavily weighted on attendance, for example, or when teachers did not provide proper supports for students who were being taught challenging material.

New era assessments can begin to change this equation. On large-scale assessments, it is possible, as noted in principle xx, to ask students if they had the opportunity to learn the material and then to provide a coverage map back to the school and teacher to show which topics students perceived to be unfamiliar. Grading systems can also be calibrated much more to actual work produced, what some call competency grading, rather than issues such as attendance, tardies, or behavior. Grading systems are not the proper venue for behavior management systems or other means of gaining student compliance.

It is entirely feasible to have a two-dimensional grading system. One dimension addresses student mastery of specified academic content, thinking skills, and learning strategies such as time management and persistence with challenging tasks. The other dimension reports on student behavior and other important social skills such as collaboration, conflict resolution, attendance, and general engagement. The purpose of a two-dimensional grading system would be to send clear signals of the relationship between mastering academic content and the social and personal skills students demonstrate in the classroom. From an equity perspective, this is fairer because students who can show mastery but who have social or behavioral issues are not
punished, and their mark for social/personal behavior is used as an indicator and means to improve in those areas in ways that support improved academic achievement.

Another strategy to help enhance equity for large-scale assessments is in the area of constructed responses. Although constructed responses in general are under pressure from states wishing to reduce the cost and time of assessment, tasks such as writing prompts and multi-step mathematics problems where students show their work do provide valuable information in key skill areas. Test administrators can do as is done in several other countries with national exams and release the general topic for the writing task or mathematics problem in advance, at the beginning of the school year. Doing so improves opportunity to learn because teachers can help students prepare in that general area.

For example, the writing task may focus on the use of description in contemporary fiction, without revealing the exact nature of the prompt to be used. With this information, teachers can focus on literature that exemplifies this technique and help students learn to analyze it. Then, students have to show they can transfer their knowledge from the material they have read to a new, unfamiliar text. This type of transfer and application of knowledge, as has been noted, is in and of itself challenging and a demonstration of deep learning of the content. From an equity perspective, this is fairer because all students have a greater opportunity to learn the skill at hand, rather than the test being the kind of general knowledge exam that has typified large-scale assessment. The assumption is that if students show they can master a complex literary analysis technique, this is an indication of a broader underlying knowledge base in literature and literary analysis and, by extension, of exposure to a challenging curriculum in literature.

This assumption would then have to be validated further by other evidence. This is where additional assessment tasks and classroom assignments come into play. They play the validating role. If the student demonstrates consistent skills in literary analysis across multiple types of assessment, it can be concluded that the student has achieved sufficient mastery in this area. No further need exists to determine if they can analyze every genre of literature or use multiple analysis styles and techniques. These are skills of the emerging expert, while the purpose of new era assessment, as noted, is to see evidence of emerging competence and a movement toward solid competence by the end of 12th grade.

Perhaps most important is the commitment in new era assessment not to use data that promotes inequitable categorizations or conclusions about individuals and
groups of students. While a test can be useful to discover when students are not being given an equal opportunity to succeed, that information becomes successively less useful if nothing is being done to address the issue. Continuing to test students on topics they are not being taught makes little sense. To be an assessment that serves an equity purpose, the results must be used to redesign curriculum, train teachers, equip educational leaders to make necessary changes, and to lead to system redesigns that result ultimately in students being given the opportunity to learn what has been identified as knowledge and skills critical to their future success.

This is an additional equity issue. The material being tested on high-stakes exams and graded in the classroom must rise to the highest levels of value to justify attaching stakes to assessments. New era assessments are closely calibrated to the 21st century knowledge and skills that are essential for success, many of which go beyond traditional content areas to include cognitive strategies, learning skills, and personal management habits and techniques.

Many systems of educational standards exist currently and serve as the basis for specifying the knowledge domain to be measured on standardized tests and the content to be taught and tested in the classroom. These systems represent an amalgam of current era and new era knowledge and skills, and often overlook entirely aspects that go beyond traditional academic knowledge. This is unfortunate from an equity perspective because many students do not need to do well on every standard in order to succeed; they need to address a critical subset that are the most closely associated with college readiness and career success. They also need to be recognized for other skills and capabilities that may serve them well in life beyond formal schooling.

Students well prepared from an equity standpoint have the ability to be adaptable, lifelong learners who can succeed in at least one area of endeavor. Judging them on breadth of knowledge alone overlooks these key criteria. Too much of the current era assessment approach seems designed to prepare students to succeed on the game show Jeopardy rather than to successfully enter a particular career pathway. New era assessments can help identify student strengths in core areas and encourage them to keep building on those strengths while also addressing areas in need of improvement. This ultimately leads to all students being able to succeed in life and to keep learning and acquiring new knowledge and skills throughout their careers and lives.

It’s important to note that this is not a prescription for early specialization or for neglecting certain subjects for certain students who are marked for a particular
occupation. Nothing could be less equitable. In fact, this is what often occurs as a result of current era assessment and the tracking that is often based on the score obtained on a standardized test or the grade achieved in an entry-level class. The idea here is that assessment is based on what's truly important to the future of a wide range of students from varying backgrounds, not an idealized student from a presumed middle-class upbringing. The U.S. student population is becoming ever more diverse, and assessment needs to help create opportunity for all students, not eliminate them. Focusing on what’s truly important is a first step toward making assessment more equitable for all students.