1

Assessment

Where's the Bar for Learning?

AN OPENING STORY

In one of the first years of implementation of data-driven instruction in our North Star middle school, we had a principal intern who was supervising one of our sixth-grade math teachers. One morning, he came to my office and put a sheet of paper in front of me: it was the Do Now worksheet that he had just seen used in the classroom.

"What do you notice?" he asked me. I reviewed the worksheet and saw ten problems on basic addition of fractions.

"This looks like a basic review of fractions," I answered.

"Exactly," the intern replied. "But the interim assessment we just reviewed asks students to solve word problems with fractions, and even the fractions themselves in those word problems are more complex than these ones. Yet the teacher is confident that she's preparing her students to master adding fractions."

These might seem like commonsense conclusions to many of you, but for us at North Star, the principal intern's insight was a watershed moment in identifying the common disconnect between what the teacher was teaching and what the interim assessment was measuring. What ensued was a deeper look at assessment at all levels of teaching.

Data-Driven Success Story

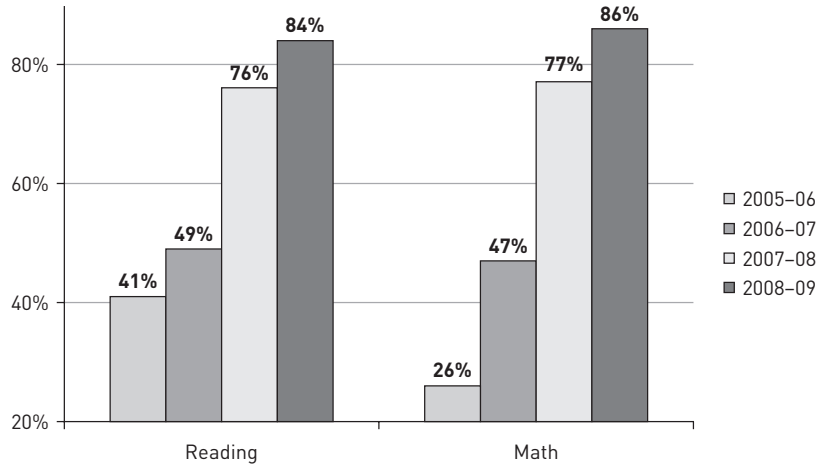
Morrell Park Elementary/Middle School: Data in the War Room

The Results

Maryland State Assessment: Percentage at or Above Proficiency, Third Through Eighth Grades

Math	3	4	5	6	7	8
2005–06	41%	37%	46%	24%	6%	2%
2008–09	**100%**	**93%**	**89%**	**83%**	**74%**	**79%**
3-year gains	**+59**	**+56**	**+43**	**+59**	**+68**	**+77**
Reading	**3**	**4**	**5**	**6**	**7**	**8**
2005–06	49%	49%	41%	38%	37%	32%
2008–09	**89%**	**93%**	**86%**	**79%**	**77%**	**79%**
3-year gains	**+40**	**+44**	**+45**	**+41**	**+40**	**+47**

Figure 1.1 Morrell Park Students on the Maryland State Assessment: Percent Proficient in Third Through Eighth Grades.



The Story

When Sean Conley began his principalship, Morrell Elementary/Middle School was a historically failing school: student achievement had remained consistently below 50 percent proficient, with scores often in the teens or twenties. The school was *zero-based,* which meant every teacher had to reapply and Conley was able to decide which teachers to hire. Conley quickly identified the best teacher in the school and asked her to let other teachers see what she was doing. By asking her to be an instructional coach, Conley sent a message that teachers needed to share their best practices with each other. The teachers started by creating common monthly math assessments, and they analyzed the results to determine which skills needed whole-class instruction, small-group re-teaching, or individual support. As they analyzed the results as a grade-level team, if one teacher had better results on one standard and another did better on a different standard, the team would regroup *all* the students from the grade into groups that were taught by the teacher most skilled at that particular standard. While they weren't able to launch a common assessment in literacy in 2006–07, they created a child-friendly writing rubric and had students analyzing writing responses and editing their own to meet the rubric.

In 2007–08, Conley focused his efforts particularly on literacy. He incorporated leveled texts and individual reading plans for students based on their reading assessment results. He increased instructional time and opportunities for teachers to pull out students who needed extra support. The main office was turned into a ''war room'' where all assessment information was posted so that teachers and parents could see it. Every faculty meeting started with celebrations and some sort of data about the students. The conversation among faculty members shifted from ''The test is not fair'' to ''What do we have to do to move students?'' Homework was differentiated to what students needed, and the district's Mathworks program continued to provide opportunities for teachers to plan the teaching of a standard aligned to the rigor of the state tests.

After moderate gains in 2006–07, results skyrocketed in 2007–08 and 2008–09. A school that had once had fewer than one in five students proficient had tripled its performance and achievement. Conley had transformed a failing building into a true school of learning.

Key Drivers from Implementation Rubric

- *Common interim assessments:* By creating common math assessments, grade-level teams were able to analyze results together and establish common goals and lesson plans.
- *Teacher action plans:* The strategic decision to re-teach difficult standards according to teachers' strengths was a creative approach to making teaching more effective.

TOWARD MEANINGFUL RIGOR: CREATING THE ROAD MAP

Assessment is the first core principle of data-driven instruction. Whether it is praised for emphasizing a ''results orientation'' or condemned for ''teaching the test,'' the practices of data-driven instruction are inextricably bound up with the process of assessment.

Ask a teacher to define what the students should be learning, and chances are the teacher will talk about conforming to school, district, or state standards (or to the standards embedded in a mandated textbook or curriculum). Yet while meeting such standards is necessary, it is nearly impossible to measure a teacher's success simply based on a list of standards. To illustrate this, consider a basic standard taken from middle school math:

> *Understand and use ratios, proportions and percents in a variety of situations.*
>
> —New Jersey Core Curriculum Content Standards for Mathematics, Grade 7, 4.1.A.3

To understand why a standard like this one creates difficulties, consider the following premise. Six different teachers could each define one of the following six questions as a valid attempt to assess the standard of percent of a number. Each could argue that the chosen assessment question is aligned to the state standard and is an adequate measure of student mastery:

Six Assessment Questions ''Aligned'' to the Same Standard

1. Identify 50% of 20:
2. Identify 67% of 81:
3. Shawn got 7 correct answers out of 10 possible answers on his science test. What percent of questions did he get correct?
4. J.J. Redick was on pace to set an NCAA record in career free throw percentage. Leading into the NCAA tournament in 2004, he made 97 of 104 free throw attempts. What percentage of free throws did he make?

5. J.J. Redick was on pace to set an NCAA record in career free throw percentage. Leading into the NCAA tournament in 2004, he made 97 of 104 free throw attempts. In the first tournament game, Redick missed his first five free throws. How far did his percentage drop from before the tournament game to right after missing those free throws?
6. J.J. Redick and Chris Paul were competing for the best free-throw shooting percentage. Redick made 94% of his first 103 shots, while Paul made 47 out of 51 shots.
   a. Which one had a better shooting percentage?
   b. In the next game, Redick made only 2 of 10 shots while Paul made 7 of 10 shots. What are their new overall shooting percentages?
   c. Who is the better shooter?
   d. Jason argued that if Paul and J.J. each made the next ten shots, their shooting percentages would go up the same amount. Is this true? Why or why not?

Though these six questions differ tremendously in scope, difficulty, and design, all of them are aligned to the state standard. Indeed, even if the standard was made more specific and called for "higher-order problem-solving skill," choices four, five, and six would still all be plausible options. If teachers were given this standard without clarification and commentary, no one could fault them for teaching only the skills needed for a question like number four, even if the end-of-year state test demanded the skills needed to answer a question like number six.

From this, one can grasp an important truth: Standards are meaningless until you define how you will assess them.

CORE IDEA

- Standards are meaningless until you define how you will assess them.

The level of mastery that will be reached by a given class is determined entirely by what sort of questions students are expected to answer. This turns conventional wisdom on its head: instead of standards defining the sort of assessments used, the assessments used define the standard that will be reached.

Although this is initially counterintuitive, it's a principle that is constantly visible in the world around us. No one would start building a house without creating the blueprints, or training for the Olympics without identifying what

CORE IDEA

- Assessments are not the end of the teaching and learning process; they're the starting point.

benchmarks define success. Likewise, we should not first teach and then write an assessment to match; instead, we should create a rigorous and demanding test and then teach to meet its standards.

In effective data-driven instruction, the most important tests are *interim assessments:* formal written examinations taken at six- to eight-week intervals during the school year. Interim assessments give standards a clear definition of the level of rigor needed to succeed. Rather than have each teacher choose a level of rigor in response to vaguely written standards, the effective data-driven school leader or teacher works to create challenging interim assessments that set a high bar for student achievement.

As one of the better-known facets of data-driven instruction, interim assessments have been the focus of a great deal of academic research. Thus far, the evidence strongly suggests that, when properly applied, interim assessments are among the most powerful drivers of academic excellence.[1] Why?

Some Advantages of Interim Assessments

- *Road map for instruction:* This point cannot be made too emphatically—rigorous interim assessments define the standards and provide a road map to rigorous teaching and learning. When educators know precisely what skill level their students must reach on each standard, they will have a clear framework for creating a challenging and dynamic curriculum. Traditional curriculum scopes and sequences do not do this on their own.
- *Improvement in teaching:* Well-designed interim assessments serve to identify weaknesses during the course of the school year. Meticulous attention to results and a constant feedback loop allow teachers to improve their craft, changing strategies in response to changing needs.
- *Targeted Focus:* By creating concrete benchmarks, interim assessments allow for classroom strengths and weaknesses to be clearly identified and

systematically targeted. In providing a baseline standard for comparison, interim assessments offer a comprehensive checkpoint of where a class needs to go and what it will take to get each of the students to that level.

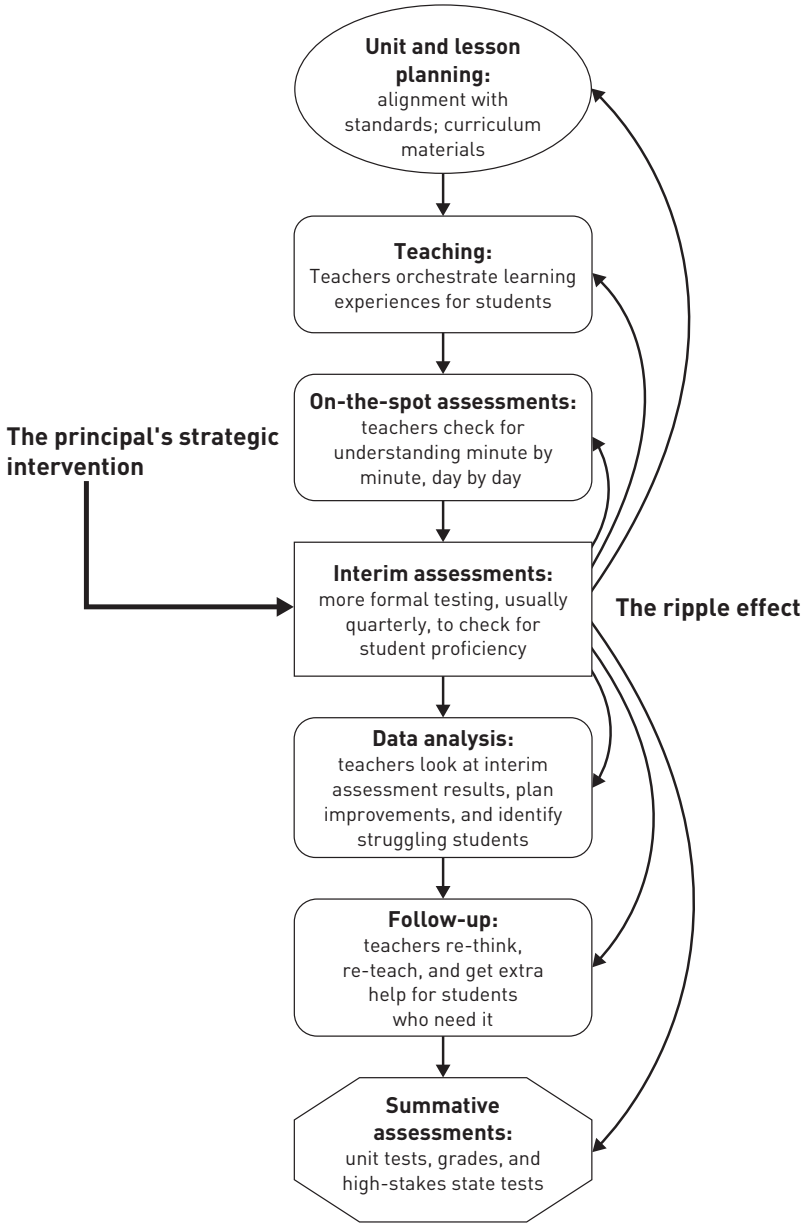
- *Accountability:* The cumulative nature of interim assessments helps hold teachers and principals accountable for student learning results throughout the year. Rather than waiting for a year-end result, interim assessments make it possible to identify failed teaching strategies while there is still time to fix them.
- *Visibility:* Interim assessments allow for performance to be charted graphically so that school leaders and staff may see visual evidence of improvement.
- *Checking for understanding without teacher support:* Because of their formal written nature, interim assessments measure student understanding without what is often called "scaffolded" support (teacher hints and guidance in problem solving), which can often reveal great differences between student output when supported by the teacher and when not!
- *Preparing students for high-stakes assessment:* The written format can also be used to simulate the high-stakes tests by which states and many businesses measure academic achievement. Unlike other types of assessments, interim examinations can adopt the structure and content of end-goal tests to determine whether students have precisely the skills they need.

INTERIM ASSESSMENTS VERSUS IN-THE-MOMENT ASSESSMENTS

A body of research (primarily championed by Paul Black and Dylan Wiliam) asserts that *in-the-moment assessments*—checking for student understanding in the very moment something is learned—have even more power than interim assessments in building student achievement.[2] I agree that, done skillfully, real-time assessments do have a powerful effect on improving teaching—they give teachers immediate data on which students aren't learning and why.

What in-the-moment assessments lack is a sense of the larger year-end goal. By being the starting point, interim assessments have the ability to create what Kim Marshall terms the "ripple effect": they influence every component of

Figure 1.2 The Ripple Effect.



Source: Kim Marshall, 2006.

Planning a Road Trip: The Need for Interim Assessments

To use a simple analogy, imagine a road trip. When you check to make sure you have enough gasoline, your tires have enough air, and nothing is going wrong with the motor while you're driving, you're doing in-the-moment assessments. If you do not do these things, you'll never make it to your destination, so in-the-moment assessments are critical!

However, if you don't have a road map, it doesn't matter how well your car is running: you could be headed to the wrong destination! Interim assessments provide the road map. Within that context, in-the-moment assessment becomes a very powerful, necessary tool. I discuss these further as a follow-up strategy in Chapter Three, on action.

the teaching process. Figure 1.2 shows how interim assessments can drive unit planning, lesson planning, teaching, in-the-moment assessment and follow-up, and finally improved year-end results.

Quality interim assessments have the power to fundamentally improve every aspect of academic performance. To realize this potential, however, you need a sound framework for interim assessments.

THE BUILDING BLOCKS OF EFFECTIVE ASSESSMENT

All these principles contribute to the practice of writing a quality assessment. If standards are meaningless until you define how to assess them, and assessments are the starting point rather than the end, then a few basic building blocks emerge for writing quality assessments:

- Assessments must be the starting point.
- Assessments must be transparent.
- Assessments must be common.
- Assessments must be interim.

Assessments Must Be the Starting Point

The first step to creating interim assessments is knowing when to start writing them. Traditionally, assessments are written near the end of the semester or

quarter they are reviewing. In this arrangement, the material taught in class determines the standards to which students will be held on the interim examination. In contrast, for data-driven instruction to be effective, this process must be reversed, meaning *interim assessments should be created before teaching ever begins.* In data-driven instruction, the rigor of the actual assessment items drives the rigor of the material taught in class. As explained earlier in this chapter, when assessments are written before teaching begins, teachers can adjust the curriculum and lessons to make sure all necessary skills are addressed. (This requires teacher accountability, also discussed in Chapter Three.)

Assessments Must Be Transparent

As part of this approach, assessments must be transparent and available to teachers and school leaders from the start of the school year. More broadly speaking, teachers, students, parents, and community members should all know exactly what skill level the students will reach and what steps they will take to get there. As demonstrated earlier, *standards are not sufficient to drive teaching to appropriate rigor.* Of course, this does not mean giving out copies of test answers to students on the first day of class, but it does mean publicly posting the exact sorts of skills needed so that every member of the school community knows what to expect. By making assessment expectations transparent and clear, schools can take control of their curriculum and guide learning based on their vision. *Since this is the area where districts fail more than any other, Chapter Five includes a whole section of coping mechanisms to employ when forced to use secretive district-mandated assessments.*

Assessments Must Be Common

It is essential to use the same examinations across all classes in a given grade-level and content area. If individual teachers develop and administer their own assessments, this generates problems similar to the six different levels of mastery demonstrated in the ''Percent of a Number'' questions. Assessments administered by individual teachers also make it nearly impossible to meaningfully track test-to-test progress or to coordinate fully shared standards across the entire student body. Furthermore, the process of creating and sharing a common assessment is itself a valuable opportunity for faculty to share ideas and collaborate to create the best curriculum possible for all students.

Assessments Must Be Interim

While schools and teachers assess in some form all the time, the key assessments—the ones driving change in schools making dramatic gains in achievement—are interim assessments. Interim assessments need to happen at least quarterly, and should ideally be given every six to eight weeks at the middle and high school levels.[3] If assessments are administered less frequently, then weaknesses will go unrecognized until it is too late to correct them. If assessments are administered far more frequently, then teachers cannot do the depth of analysis described in Chapter Two without burning out.

WRITING OR SELECTING THE RIGHT INTERIM ASSESSMENT

Once the fundamental logistics are in place, one can turn to the task of writing or selecting the test itself. The success stories I've included in this book show that effective results can emerge either from creating rigorous interim assessments at the school level or selecting already available assessments. Either choice can lead to success so long as one applies the core principles listed here.

Start from the End-Goal Exam

All public schools (and even most private schools) face the high stakes of end-goal tests by which student achievement is measured. At the primary school level, such assessment often includes statewide or districtwide exams; at the secondary level, it could include SAT/ACT scores or AP/ IB assessment results. In any case, when

CORE IDEAS: Interim Assessment

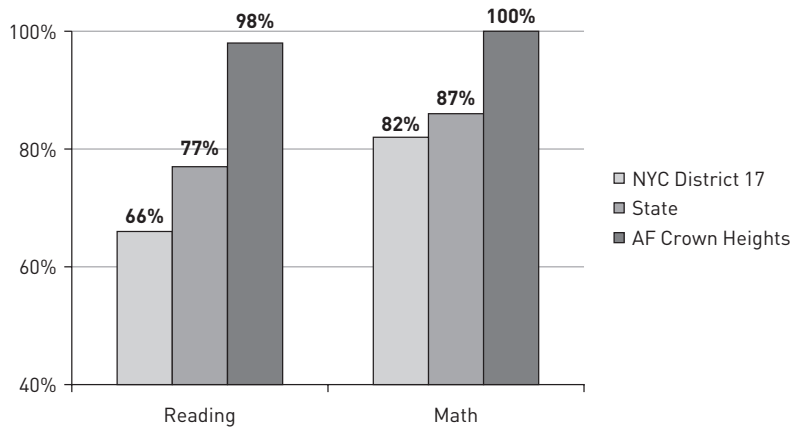
- Start from the end-goal exam.
- Align the interim assessments to the end-goal test.
- If acquiring assessments from a third party, be sure to see the test.
- Assess to college-ready standards.
- Design the test to reassess earlier material.
- Give teachers a stake in the assessment.

Data-Driven Success Story

Achievement First: A System That Makes the Grade

The Results

Figure 1.3 New York State Assessment:
Percent Proficient at Fourth-Grade AF Crown Heights 2009.



The Story

Some schools are data-driven from the start and have built a model for other schools to follow. Achievement First schools, founded in 2005, built on the data-driven legacy of Doug McCurry and Amistad Academy in New Haven, Connecticut. In 1999, McCurry was a lead teacher at Amistad, and he began experimenting with assessments in math and grammar as a way to address his own struggle to know whether the students were learning or not. He looked at state standards, made a sequence, and started building prototype interim assessments. Right away he saw that assessments needed to move away from covering specific units and become cumulative for the year, to allow for retesting of standards. By 2000, McCurry had moved into a leadership role at Amistad and launched interim assessments in math across the whole school, but there were no systems to guarantee analysis or action. A number of teachers were resistant to using interim assessments, and this created more of a ''he said, she said'' defensiveness between teachers and leaders around the results on each assessment. Despite the resistance, however, certain teachers made remarkable gains from one cycle to the next, and that shifted the dynamic in the school. Since then, Amistad has consistently outperformed the statewide proficiency average in Connecticut.

McCurry founded Achievement First (AF) with Dacia Toll in 2002, and they created an infrastructure to wield across multiple schools. Over the next seven years, AF launched additional systems, creating a process of teacher review of the interim assessments to build in transparency. McCurry built more training around what good analysis looked like, and AF moved from paper-and-pencil analysis to Excel spreadsheets to Athena, an online interim assessment platform designed to save teacher time and customize analytics. It also increased the amount of time in the school calendar given to conversations around data. By 2008, it had also built Data Days—four annual full professional development days devoted to data analysis and discussion with sample agendas and outcomes for those days. In doing so, it moved from surface-level analysis ("I'm going to re-teach that") to super-clear outcomes and a data-driven plan.

Achievement First Bushwick Elementary School is one example of this success. Principal Lizette Suxo began each Data Day with a schoolwide reflection: what were the common challenges across all grades? Are there things we should address schoolwide? Do we need intervention groups? If so, what students should be in each group, and what standards will each group be taught to master? This conversation led directly into grade-level planning. Teachers planned out the standards to be taught for the next six weeks (both new and re-teach standards) and then planned the specific lessons for the next nine days. Teachers and leaders alike used daily exit tickets and three-week targeted assessments to make in-course adjustments. The results tell the story: achievement is truly first at Bushwick and all its sister schools.

Key Drivers from Implementation Rubric

- *Ongoing professional development:* Four full Data Days set the tone for a laser-like focus on data.
- *Effective data reports:* One of the key drivers was the use of Athena, McCurry's automated results analysis system. Commercial interim assessments and analysis services are often problematic because they do not facilitate teacher-level planning, so AF built its platform with the end goal of data analysis meetings in mind, and that analysis is explicitly linked to the creation of a teacher action plan.
- *Implement action plans:* Action plans were explicitly connected to teaching, and everyone in the building monitored student learning to assess the quality of the action.

designing or selecting interim assessments, it is critical that decisions be made in reference to the specific demands of the end goal and not to vague, ill-defined academic standards, as discussed earlier in this chapter.

Align Interim Assessments to End-Goal Test

Once the specific sorts of questions that are employed by the end-goal test are noted, schools should work to create or select interim assessments that are aligned to the specific demands of the end-goal examination. This alignment should not be limited to content but should also follow the format, length, and any other replicable characteristic of the end-goal test.

Be Sure to See Any Third-Party Test

Test sales representatives have a very simple goal: to sell more tests. Because of this, they will do anything in their power to convince schools that their exam is aligned and will meet their needs. Don't take their word for it. Instead, school leaders and teachers should personally inspect actual copies of the product to see how well it lines up with the end goals in question—remembering that no third-party test is perfect. Find something that seems close, then push to modify the examination to exactly align with your school's academic goal. *This is one of most overlooked steps in schools and districts that do not have well-aligned interim assessments.*

Assess to College-Ready Standards

At every level, it's important to realize that the skills needed to pass state tests are often insufficient to ensure success in college or other postsecondary environments. As such, a final goal of well-written interim assessments is that they prepare students not only for a state test but also for college and beyond. High schools have a clearer path to do this, as they can look to align with the SAT/ACT, AP/ IB, or the demands of a college research paper. Elementary and middle schools require more creative thinking. Here are some examples of college-ready rigor for those grade spans:

Elementary School Reading Set higher reading-level expectations: At the elementary level, an easy way to push for greater rigor is to evaluate students' progress at meeting above-grade reading levels. For example, rather than expect

kindergartners to meet the equivalent of Fountas-Pinnell Level B, push for Level C or D by the end of the year. (See Myth 3 in the upcoming section for additional ideas on early elementary reading rigor.)

Elementary School Math Set higher grade-level expectations—for example, prompt each grade level to accomplish a certain percentage of the standards for the subsequent grade level. For example, second graders can accomplish all the operations standards associated with third grade. North Star Elementary School (see Success Story) has done this aggressively from Kindergarten. Using the TerraNova as a guide, they have established interim assessments for Kindergarten that measure all the K standards and half the first-grade standards. First grade then measures all first- and second-grade math standards, and second graders prepare to master all standards on the third-grade state test. Imagine how pleased third- and fourth-grade teachers will be receiving all their students already grade-level proficient at the start of the year!

Middle School Math Embed algebra in every strand. Most eighth-grade state tests have a rudimentary inclusion of basic linear equations or expressions, but few measure all of the rigor of a high school Algebra I curriculum. Middle schools can quickly increase the college-ready rigor in their classrooms by exploring algebraic applications for each mathematical strand that they teach. For example, a fifth-grade teacher presenting addition of fractions could add a question like the College-Ready Example into the class Do Nows or in-class activities:

College-Ready Example—Algebra in Fifth-Grade Math

Write an expression for the following:

Mr. Smith has b books in his classroom. He gives three of them to students. Then he splits the rest of them evenly on his two bookshelves. How many books are on each shelf? Justify your answer.

Middle School Reading Push for deeper reading of the text: Adding college-ready rigor to reading can be challenging. Giving students harder texts to read is laudable, but it does not accomplish this task in itself: if the book is well above

the students' reading level, vocabulary knowledge might limit their ability to comprehend the text. Still, an overlooked strategy is choosing text with grade-level vocabulary but complex meaning. This allows the teacher to assess for more critical reading than is often possible with middle school novels. The box gives an example of a poem that has very accessible language but requires deep, critical thinking:

College-Ready Example — Rigor in Middle School Reading

Chicago Poet

by Carl Sandburg

I saluted a nobody.
I saw him in a looking-glass.
He smiled — so did I.
He crumpled the skin on his forehead, frowning — so did I.
Everything I did he did.
I said, "Hello, I know you."
And I was a liar to say so.
Ah, this looking-glass man!
Liar, fool, dreamer, play-actor,
Soldier, dusty drinker of dust —
Ah! He will go with me
Down the dark stairway
When nobody else is looking,
When everybody else is gone.

1. In stanza 1, Sandburg looks into the mirror and says, "Hello, I know you," but then calls himself a liar to say so. How is this possible when he is looking at himself? (Inference)
   A. He is still figuring out his identity.
   B. He is insulting himself because he is angry.
   C. He does not recognize his physical self.
   D. He is being playful and joking.
2. Which of the following questions might Sandburg want to ask of this "looking-glass man"? (Asking questions)
   A. Where are your friends?
   B. Where are you right now?
   C. How are you feeling?
   D. Who do you want to be?

Design the Test to Reassess Earlier Material

Additionally, effective assessments revisit material from earlier in the year. In many fields, such as math, this review is vital to retaining information and learning new concepts. It also ensures that teachers have the opportunity to see if their re-teaching efforts were effective. One such method of review is to make tests longer as the year progresses; a second is to test all material from day one and then track improvement as the students actually learn the concepts being tested. No matter which method is chosen, however, it is important that review of past material is made a central part of interim assessments. *This is a common, critical mistake of schools and districts where assessments fail: they convert interim assessments into unit tests (just covering material in that time period) rather than cumulative assessments.*

Give Teachers a Stake in the Assessment

Finally, when assessments are created or selected, teachers should have meaningful input. This is critical, because it ensures accountability; teachers who are included in the assessment writing or selection process become invested in the assessments' effectiveness. Give teachers a stake in the assessment, and you'll give them a stake in the results.

ASSESSMENT MYTHS DISPELLED

Before discussing the key steps to writing effective interim assessments, it's important to address some of the common misperceptions surrounding testing. Given the widely conflicting attitudes and understandings of assessment, it is unsurprising that many myths and half-truths surround the assessment process. Before going any further, it is useful to dispel some of the more prominent and problematic of these myths:

- Multiple-choice questions just aren't rigorous.
- Tests such as the SAT don't really measure student learning.
- Doing well in early elementary reading assessments will guarantee proficiency in the upper elementary grades and beyond.

MYTH 1: Multiple-Choice Questions Just Aren't Rigorous

Reality: In the eyes of many teachers and school leaders, multiple-choice questions are vapid and low-rigor exercises in test taking, useful only because they are easy to score. From this viewpoint, assessments that employ such questions cannot test more sophisticated concepts and do not require rigorous critical thought. Yet while this observation seems intuitively obvious, it is incorrect. Consider the three questions in the box, each of which pertains to the main idea or theme of the story "Little Red Riding Hood."

Three "Little Red Riding Hood" Assessment Questions

1. What is the main idea of this story? (open-ended response)
2. This story is mostly about:
   a. Two boys fighting
   b. A girl playing in the woods
   c. Little Red Riding Hood's adventures with a wolf
   d. A wolf in the forest
3. This story is mostly about:
   a. Little Red Riding Hood's journey through the woods
   b. The pain of losing your grandmother
   c. Everything is not always what it seems
   d. Fear of wolves

In comparing the last two questions, it is easy to see that question 3 is far more demanding. Clearly, some multiple-choice questions are more rigorous than others. But is question 3 less rigorous than the open-ended question 1? The answer: it depends on the rubric used for question 1. If it merely requires that the student identify a wolf and a girl, then it will be the easiest of the three; if it demands a nuanced, five-paragraph explanation of the symbolism and word choice of the story, then it is the most difficult. Implicit in these questions are two important truths: first, for multiple-choice questions, the options define the rigor.

If the choices present are very similar to one another, as in question 3, then the test will require a far higher degree of knowledge. A high degree of comprehension is also required when the options include complex vocabulary and prior knowledge. Thus, depending on their answer options, multiple-choice questions can offer as much rigor as any open-ended response prompt (simply look at Advanced Placement exams for an example of this). For open-ended questions, by contrast, the rubric defines the rigor. Sometimes these rubrics are explicit and transparent for the students, including anchor papers detailing exactly what "proficient" looks like. In other cases, these rubrics are internal metrics in a teacher's mind that the students have not been explicitly taught. Either way, the rubric drives the rigor.

CORE IDEAS: Multiple Choice

- In a multiple-choice question, the *options* define the rigor.
- In an open-ended question, the *rubric* defines the rigor.
- Any good assessment will combine multiple forms to achieve the best measure of mastery.

Even with the greatest rubric or the best wrong answer choices, multiple-choice questions and open-ended questions still measure different aspects of rigor of the same standard. One requires putting thoughts together in your own words; the other involves distinguishing critically between various plausible actions. These two angles simulate real life and also every testing situation students will face in the future. Thus, a truly rigorous test should be created from open-ended responses with challenging rubrics and multiple-choice questions (or other formats) with challenging options. Both test critical skills, and when it comes to creating effective and rigorous assessments, both are necessary, complementary sides of the same coin.

MYTH 2: Tests Such as the SAT Don't Really Measure Student Learning

Reality: A second common misconception holds that standardized assessments like the SAT are simply tricky tests that don't really show student mastery.

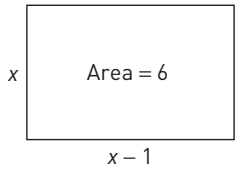
While this is sometimes true, it is not categorically correct. Consider the two questions below:

Two Questions on Quadratic Equations

1. Solve the following quadratic equation:

$$x^2 - x - 6 = 0$$

2. Given the following rectangle with the lengths shown below, find the value of x:



If you solve question 2 algebraically, you arrive at the same quadratic question that is listed for question 1. However, you needed mastery of many additional mathematical concepts to set up the equation: properties of a rectangle, area, distributive property, and more. Question 1 also has two possible answers: -2 and 3. In question 2, however, the student must eliminate -2 as a possible answer, because a rectangle would not have a negative side!

Question 1 is taken straight from an algebra textbook; question 2 is from the SAT. The issue with the SAT question is not that it's trying to trick the student; it's that it requires a deeper, more conceptual understanding of quadratic equations, as well as the ability to apply it in the context of geometric properties. So when a student struggles with the SAT but does well in algebra, the first thing a teacher or leader must consider is whether the rigor of class instruction matched the rigor of the SAT. Defining the rigor of the questions on assessments to the highest bars that students are expected to reach makes sure that students will master any sort of test put in front of them.

SAT Rigor: The Experience of a High School Math Teacher

In the first round of implementation of data-driven instruction at North Star Academy Charter High School, our school leaders designed interim assessments that were aligned both to the New Jersey state test and the SAT. After we implemented the first round of interim assessments, one of the math teachers complained that there were too many SAT prep questions that weren't really connected to teaching high school math. As a school leader, I had a choice to make about how to respond. I could explain to the teacher why I thought the assessment was a valuable tool with rigorous SAT questions embedded, or I could try to get her to reach the conclusion herself. I opted for the latter. I said that I appreciated her concern and invited her to look at the assessment with me to identify the questions that were problematic. We then identified the questions that were aligned to the rigor of the SAT, and with each one I asked her, "Are these skills that your students should know in your class?" After each question, she acknowledged that the question did indeed measure standards she was teaching. We repeated this exercise throughout the entire test, and the teacher slowly realized that her concern was not about SAT prep but the high level of rigor of the questions. Without challenging her directly, our whole conversation had shifted from test prep critique to how much additional work it would take students to reach a higher level of standards mastery. The rest of our time was focused on new strategies and schoolwide systems to support student learning.

Too often our critiques of tests are done from the 20,000-foot aerial view. However, when you get up close and examine actual test items, you start to discover the real issues of rigor (or the lack thereof).

MYTH 3: Doing Well on Early Elementary School Reading Assessments Will Guarantee Proficiency in the Upper Elementary Grades and Beyond

Reality: Although it might seem counterintuitive, schools assess more in early elementary school than at any other grade level: there are individual reading assessments, observation checklists, sight word checks, and so on and on. This is justifiable because the ability to read is one of the critical foundations of

Data-Driven Success Story

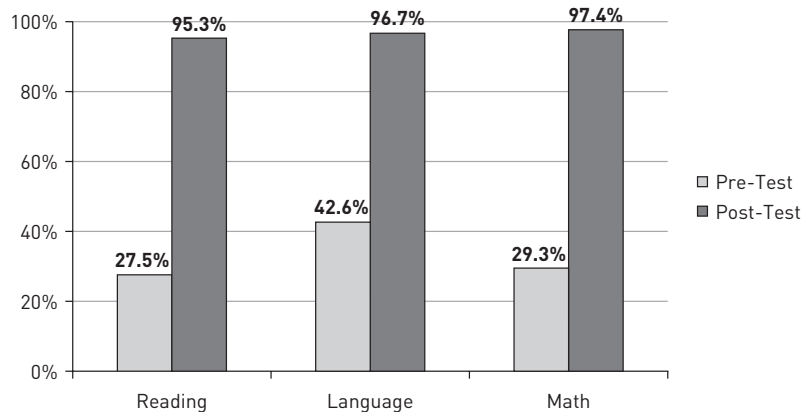
North Star Elementary: Exploding Expectations

The Results

2008–09 TerraNova Exam: Kindergartners' Median National Percentile

Reading			Language			Math		
Pre-test	**Post-test**	**Change**	**Pre-test**	**Post-test**	**Change**	**Pre-test**	**Post-test**	**Change**
27.5	95.3	**+67.8**	42.6	96.7	**+54.1**	29.3	97.4	**+68.1**

Figure 1.4 North Star 2008–09 TerraNova Exam: Kindergarten — Median National Percentile.



The Story

When founding principal Julie Jackson began planning the launch of North Star Elementary School, almost everyone she talked to argued that data-driven instruction did not really apply to K–2 education. "Save testing for the older grades." "Child development in Kindergarten is so irregular that it's too difficult to measure." "Children need to develop at their own pace." Because of her experience as a high school principal who witnessed firsthand the challenges of first-generation college students in their initial years as undergraduates, Jackson was determined to break the stereotypes of what elementary school students could do. First, she selected a leveled reading assessment that targeted critical reading

comprehension in addition to fluency, accuracy, and decoding—the STEP assessment out of the University of Chicago. By doing so, teachers avoided the trap of only teaching reading for fluency and accuracy and also focused on Kindergartners' ability to make inferences and understand the purposes of stories. Teachers targeted lessons directly toward the areas in reading where the students were struggling, whether it was the vowel sounds, the ability to draw meaning from a picture, or making a connection between the beginning and end of a story.

Jackson's colleague, Christian Sparling, designed interim math assessments that not only covered Kindergarten material, but also half the standards from first grade as well. Unsatisfied with the lack of rigor of most Kindergarten math curricula, they supplemented their math program with story problems and more time to apply the mathematical principles in exercises. They built a comprehensive assessment calendar, trained all teachers in the principles of data-driven instruction, and posted class results in the teachers' room after each round of interim assessments. The Kindergarten teachers collaboratively planned lessons and consistently targeted deficient standards and struggling students throughout the year.

By the end of the first year of the school, Jackson and her team of teachers had accomplished extraordinary results. At the beginning of the year, only three of the seventy-eight Kindergartners were reading. By the end of the year, all but three were reading on the first-grade level. The median national percentile scores on the TerraNova placed North Star students in the upper 10 percent of all students nationwide. For 2008–09, Jackson and the elementary school set the goal to accomplish all second-grade math standards by the end of first grade, and preliminary results strongly suggest that they will meet that goal. By exploding expectations for what is possible in K–2 education, North Star Elementary School is redefining elementary education.

Key Drivers from Implementation Rubric

- *College-ready interim assessments:* Jackson set the interim assessment rigor to match first grade, not just Kindergarten, causing teachers to shoot higher with their students and keep in mind the larger goal.
- *Collaborative lesson planning:* The Kindergarten teachers identified lead planners for each subject, and the planners led the creation of lesson plans that everyone agreed to implement. After teaching, they discussed the strengths and weaknesses of the lesson, and they developed even more effective lessons based on that feedback.
- *Accountability:* Jackson reviews all lesson plans as well as observing classes to look for alignment with the key standards from assessment analysis.
- *Deep analysis:* By creating an innovative analysis spreadsheet for the leveled reading assessment results, teachers were able to identify the precise letters, sounds, and reading strategies that needed attention.

elementary education. The issue at hand, however, is whether schools are using the right assessments to prepare critical readers.

Consider four different early literacy assessments, all testing students at Level C on the Fountas-Pinnell scale (a K–1 reading passage).

Four Early Literacy Assessments

Story text (assume that every line is a separate page and accompanied by a picture):

> When I grow up, I want to put out fires.
> I want to play ball.
> I want to go to the moon.
> I want to teach school.
> I want to fix cars.
> But now, I am happy to be a kid.

Here are the comprehension questions required by four major early childhood assessments:

DRA (Diagnostic Reading Assessment)
Students need to

- Re-tell the story.
- Make a personal connection.

DIBELS (Dynamic Indicators of Basic Early Literacy Skills)

- No test for comprehension given: fluency in reading combined with proper decoding via nonsense words is considered an adequate predictor of comprehension.

Running Record (this is a sample; there are many different versions)

- Students need to: Tell what happened in the story.
- Answer ''right there'' basic comprehension questions.

STEP (Strategic Teaching and Evaluation of Progress) Assessment
Students need to answer the following questions

- What is the first job the girl thinks about doing?
- What job would make her leave earth?
- What does she think she will do as a teacher?
- Why does the girl say that for now she is happy being a kid?

Each test claims to be an accurate, adequate predictor of future reading performance. They all monitor how quickly a student reads, with how many mistakes (and each has unique additional components focused on decoding, spelling, and other areas). However, they vary significantly in how they assess for comprehension.

Considering that most elementary educators consider the assessments discussed in this section to be interchangeable, it is striking how each one requires a radically different level of mastery of reading comprehension. As is true in later grades, the assessment selected goes a long way in determining just what will be learned—and therefore what will be taught. While every elementary teacher that I know strives for deep understanding with each year's students, seeing students pass an assessment that doesn't require critical comprehension can lull any teacher into a false sense of security about the student's progress. Particularly in urban schools, ignoring the importance of selecting rigorous assessments for the early years of education will leave students at a tremendous disadvantage as they move toward later grades. Countless schools have shared their experience of having students with excellent mastery on DRA or DIBELS turn out to struggle with state assessments. It is not that the state assessments are unfair measures of student learning at the early grades; it is that the state assessments ask for critical reading comprehension that the other assessments did not. Choosing an effective early assessment that measures not only fluency, decoding, and basic comprehension but also inferential thinking will push schools and teachers to introduce more rigor and better prepare young students for the challenges ahead.

BOILING DOWN TO THE ESSENCE: THE FIVE CORE DRIVERS OF ASSESSMENT

If a single theme unifies the ideas outlined in this chapter, it is that effective assessments are those written with tremendous attention to detail. Schools must take the time to properly and specifically align their tests, to consider in detail the standards they need to approach, and to review the rigor of questions not based on myths and common knowledge but on their own observations. Though often neglected or overlooked, these fundamental principles of assessment creation invariably define the rigor of a school and, in doing so, determine what students will achieve.

In sum, what follows is a rubric that consolidates this information into five key drivers for the principle of assessments.

Assessment: Five Core Drivers

- *Transparent starting point:* Assessments need to be written before the teaching starts, and teachers and schools need to see them in advance: they define the road map.
- *Common and interim:* Assessments should apply to all students in a grade level and should occur every six to eight weeks.
- *Aligned to state tests and college readiness:* Assessments should be aligned to state tests in format, content, and length, and also aligned to the higher bar of college readiness via SAT/AP/ IB exams, research papers, and so on.
- *Aligned to instructional sequence:* Assessments should be aligned to the teachers' sequence of clearly defined grade-level and content expectations, so teachers are teaching what will be assessed.
- *Reassessed:* Interim assessments should continuously reassess previously taught standards.

APPLICATION: FIRST STEPS FOR TEACHERS AND LEADERS

So what is the most effective way to implement these principles of assessment as a classroom teacher, school leader, or multicampus or district office leader? What follows are the first steps that could be taken to put this into action.

Level 1 — Teachers

In some schools, teachers will have a fundamental role in the creation of interim assessments. If that is the case, please view the steps discussed at the district and multicampus level. If, however, you work in a school where you do not have input into the interim assessment, the following steps can help you develop in-class assessment tools that will raise the bar for driving change in your classroom:

- Analyze the interim assessment or end-goal test.
- Build your in-class assessments prior to teaching the unit.
- Plan lessons to meet the rigor of that assessment.
- Where applicable, set your college-ready goal.

Analyze the Interim Assessment or End-Goal Test Acquire the closest version that you can find of your state test, interim assessment, or other year-end assessment by which your students' learning will be measured. (This will vary

from state to state: some states have actual prior-year state tests available, others have one practice test, some just have sample questions. You can also try to acquire the interim assessments from a high-achieving school in your state—assessments that have been proven to work and be aligned to the state test.) Jon Saphier, author of *The Skillful Teacher,* offered me the following precise question to use when analyzing the end-goal assessment: "What are the skills and knowledge needed to master each assessment question?" In the case of a multistep word problem or analytical essay, this list could be quite extensive. This serves as the starting point for determining what to teach your students. Ask yourself, which of these skills and knowledge elements do the students already know, and which ones will I need to teach them?

Build Your In-Class Assessments Prior to Teaching the Unit Before teaching your next unit, design your unit-ending assessment. As you work on it, create questions that mirror the format of the end-goal test that you acquired in the first step. Make sure you have questions that match the rigor, format, and question type. At the same time, include *building-block questions:* questions that are below the rigor of the end assessment but are necessary steps toward proficiency. In math, this could include basic computation skills even as you are pushing for word problem application. In literacy, this could include using a lower-leveled text at first even as you push for students to eventually demonstrate comprehension on grade-level passages.

Plan Lessons to Meet the Rigor of That Assessment Once you've designed your in-class assessment, start planning your lessons. With the end assessment clearly defined, you have a road map for all the skills and knowledge elements—and to what degree of rigor—that you will need to teach to ensure that your students are proficient on the unit-ending assessment. Keep referring back to the actual assessment questions while you plan to make sure that every activity sets up the students to succeed at that level of rigor.

Where Applicable, Set Your College-Ready Goal As stated earlier, proficiency on state assessments is a necessary but insufficient bar for preparing our students for success in college and life beyond. If your students are currently well below grade level, state test proficiency goals could be an appropriate step for the moment. Once you start to achieve that, higher goals can continue to drive needed student college readiness.

Data-Driven Success Story

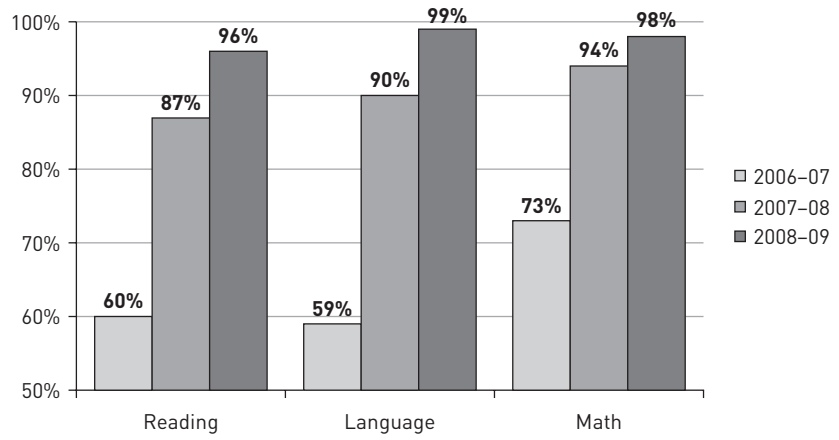
South Bronx Classical: Excellence in Real Time

The Results

TerraNova: Percent at or Above Grade Level

Grade	Kindergarten			First Grade		
Year	**Reading**	**Language**	**Math**	**Reading**	**Language**	**Math**
2006–07	50%	53%	79%	70%	65%	67%
2007–08	92%	85%	93%	82%	94%	95%
2008-09	**98%**	**97%**	**97%**	**93%**	**100%**	**98%**
Gains	+48	+44	+19	+23	+35	+31

Figure 1.5 TerraNova Average at or Above Grade Level: Kindergarten and First-Grade Students at South Bronx Classical.



The Story

Located in a gritty New York neighborhood, South Bronx Classical was founded in 2006 to serve a student body where 100 percent are minority students and 90 percent qualify for free or reduced lunch. When Scott Hudnor took over as principal in 2007–08, he was determined to put data-driven instruction in place. Having attended the Data-Driven

Instruction Comprehensive Leadership Workshop (see Chapter Twelve), Hudnor first focused on creating effective interim assessments.

In the summer of 2007, Hudnor assembled a team of teachers who worked to identify precisely what level of rigor the New York State exam and the national TerraNova exam required, and what specific skills were needed for students to reach them. Once they knew exactly what their end goal required, Hudnor and his team set about creating a series of interim assessments in math and in English and language arts. Rather than simply implement regular interim assessments, however, South Bronx Classical created an aggressive follow-up system in which students took daily math assessments and daily English assessments to track their performance in real time. This up-to-the-minute awareness of student strengths and weaknesses was coupled with formal tests every two weeks that served as miniature interim assessments within the larger structure of quarterly tests. Critically, Hudnor made sure that each layer of this assessment structure was built into the academic calendar before the year began.

Once South Bronx Classical's multilayered, real-time assessment structure was in place, the school was able to implement effective and rigorous follow-up to aggressively target standards that posed problems for students. Every teacher at the school spent at least two of the ten weekly "specials" periods (such as gym class) tutoring small groups of students on key re-teaching standards. In 2008, the school added daily twenty-minute re-teach blocks devoted to these standards. Daily assessments were constantly updated to reflect student achievement, with areas of weaknesses reinforced and spiraled back into teaching as soon as they were identified. And as the facts have shown, this single-minded focus on knowing exactly what students have learned has paid off in a tremendous way!

Key Drivers from Implementation Rubric

- *Aligned assessments:* In creating interim assessments, South Bronx Classical staff made sure that they worked backwards from the original text of their end-goal exam to ensure aligned rigor.
- *Ongoing assessment:* By tracking student understanding through interim assessments, bi-weekly tests, and short daily quizzes, South Bronx Classical staff knew exactly what students were learning and what standards needed work, and they knew it in real time as teaching was occurring.
- *Teacher action plans:* Hudnor and his staff were willing to think outside the box to find time for critical re-teaching. By making targeted re-teaching a key priority and adapting the nonacademic schedule to fit it in, South Bronx Classical was able to greatly increase the amount of effective teaching time in the school day.
- *Deep analysis:* South Bronx Classical looked closely at student answer choices to determine precisely where the learning gap was occurring.

Level 2—School-Based Leaders

Coaches, assistant principals, principals, and other school-based leaders all have different levels of authority and interaction with teachers. The degree to which each of the following steps is implemented will depend on your role. Here are the critical first steps you can do in the area of assessment.

Make Sure Your Interim Assessments Are Aligned and Rigorous The first thing to do is analyze the quality of the interim assessment vis-à-vis your state test. Acquire the closest version that you can find of your state test. (Some creative means of doing so are listed in the Level 1 Teacher section.) Line up actual test items from both assessments that are assessing the same standard, and determine if the interim assessment is meeting or exceeding the rigor of the state assessment. Exhibit 1.1 models a guiding worksheet that could help with doing that analysis. If interim assessments are not aligned or rigorous, see Chapter Five for creative workaround solutions.

Manage and Support Teachers to Use Effective Assessments Look at teacher materials and in-class assessments when observing. For example, observe the quality and the rigor of the actual activities and assignments going on in the classroom. Do the in-class assessments meet or exceed the rigor of the interim assessments and year-end assessments? (See the criteria established in the Level 1 section.) Do the teacher's activities and plans match the rigor of these assessments? Where are there gaps?

Facilitate teachers' creating high-quality, in-class assessments and planning backwards from them. Teachers can always use more planning time to focus on doing the activities listed in Level 1. Use individual meetings with teachers, grade-level meetings, and professional development time to give teachers the opportunities to do this sort of planning. Once the unit assessment has been properly aligned to the interim assessment, the planning process can mirror the work of Understanding by Design or Kim Marshall's curriculum units highlighted in *Re-Thinking Teacher Supervision and Evaluation.*[4] What makes the process so valuable is that the interim assessment has already clearly defined the bar for rigor, so the planning is double the value!

Level 3—District-Level or Multicampus Leaders

The biggest impact that district-level leadership can have is in the creation or selection of rigorous, high-quality interim assessments. One of the single most

Exhibit 1.1 Interim Assessment Evaluation Worksheet.

INTERIM ASSESSMENT ITEM	STATE TEST ITEM	THE RIGHT CONTENT Addresses the same standards, and addresses the standards as rigorously as the state test.	THE RIGHT FORMAT Reflects format and type of questions from state exam; (If applicable) reflects format of and types of questions from exam. Rubrics are used, if applicable. Wrong answers illuminate misunderstanding.	THE RIGHT COLLEGE-READY EXPECTATIONS Rigor and content seem appropriate for developing college-bound students. Content is "State test plus" in areas where state test is not college-preparatory. More complex than state tests (require additional critical thinking and application). More standards covered within the test and within the same question.	COMMENTS Comments and suggestions to improve question.
1		Yes/no	Yes/no	Yes/no	
2		Yes/no	Yes/no	Yes/no	
3		Yes/no	Yes/no	Yes/no	
4		Yes/no	Yes/no	Yes/no	
5		Yes/no	Yes/no	Yes/no	
6		Yes/no	Yes/no	Yes/no	
7		Yes/no	Yes/no	Yes/no	
8		Yes/no	Yes/no	Yes/no	
9		Yes/no	Yes/no	Yes/no	

limiting factors in schools' achievement growth are poor interim assessments mandated by their districts. The five core drivers of assessment are listed here. Using this checklist, does your district or network have quality interim assessments?

If your district's interim assessments don't meet all the criteria listed here for each subject, then your critical task is to redesign them. Here are some key points that are worth reiterating.

Interim Assessments are instructional tools first, validity tools second: Many companies that sell interim assessments do not allow schools to see their product—either before or after administration—because they want to keep the results "valid." It cannot be said more strongly: if standards are meaningless until you define how to assess them, then curriculum scope and sequences lack a road map for rigor without a transparent assessment. Transparent assessments

Core Ideas: Assessment Checklist

1. Common interim assessments: four to six times a year. __/4
2. Transparent starting point: teachers see the assessments at the beginning of each cycle; they use the assessments to define the road map for teaching. __/4
3. Aligned to state tests and college readiness. __/4
4. Aligned to instructional sequence of clearly defined grade-level and content expectations. __/4
5. Reassess previously taught standards. __/4

Rating Key
4 = Exemplary Implementation; 3 = Proficient Implementation; 2 = Beginning Implementation; 1 = No Implementation.

allow teachers to plan more effectively and increase rigor across schools. The goal is not to compare schools (that's the purpose of summative state tests!)—it is to guide instruction at the classroom level. This is not possible without transparent assessments.

Don't take anyone's word for it—check out the test itself: Since assessments define standards, then it is insufficient to align an interim assessment to a scope and sequence alone. You must compare it with the end assessment to make sure it assesses standards at a similar or higher bar of rigor (see the Level 2 section for more details). Ask the assessment creator to prove alignment by showing actual tests in comparison to your state tests.

Involve teachers and leaders in the interim assessment selection or creation process: Don't underestimate the talent of your highest-achieving teachers and leaders: they can be an invaluable resource in building a quality interim assessment program.

Chapter One: Reflection and Planning

Take this opportunity to reflect upon assessment at your own school or district. Answer the following questions:

- After reading this chapter, what are the key action steps around assessment that you are going to implement in your school (and that you can realistically do)?

- Who are the key people in your school with whom you need to communicate this plan and have on board?

- How are you going to get them on board? What are you going to do when someone says no? (What's Plan B?)

- Set the key dates for each action step, write them here, and then put them in your personal agenda or calendar to hold yourself accountable for implementing these steps.