

# What's the difference? Criterion-referenced tests vs. norm-referenced tests

[ACSA PARTNER \(HTTPS://CONTENT.ACSA.ORG/AUTHORS/ACSA-PARTNERS\)](https://content.acsa.org/authors/acsa-partners)

2 Votes

By [Renaissance \(https://www.renaissance.com/2018/07/11/blog-criterion-referenced-tests-norm-referenced-tests/\)](https://www.renaissance.com/2018/07/11/blog-criterion-referenced-tests-norm-referenced-tests/), an ACSA [Partner4Purpose \(https://www.acsa.org/Partner4Purpose\)](https://www.acsa.org/Partner4Purpose)

Have you ever been perplexed by a situation like this one?

In the fall, a student named Bruno did well on a district assessment. He scored 55 out of 100, which the district considers “proficient” for his grade level. His percentile rank was 88, which puts him ahead of his peers.

Later that school year, in the spring, Bruno took the same assessment again. This time he scored 60, still “proficient” for his grade, but suddenly his percentile rank has dropped to 38.

What happened? Bruno's spring score of 60 is higher than his fall score of 55, but his percentile rank is lower, dropping from 88 in the fall all the way down to 38 in the spring. How is that even possible?

## Criterion-referenced vs. norm-referenced

To understand what happened, we need to understand the difference between [criterion-referenced tests \(https://www.renaissance.com/edwards/criterion-referenced-test/\)](https://www.renaissance.com/edwards/criterion-referenced-test/) and [norm-referenced tests \(https://www.renaissance.com/edwards/norm-referenced-test/\)](https://www.renaissance.com/edwards/norm-referenced-test/).

The first thing to understand is that even an assessment expert couldn't tell the difference between criterion-referenced test and a norm-referenced test just by looking at them. The difference is actually [in the scores \(https://www.renaissance.com/2016/05/12/giving-meaning-to-test-scores/\)](https://www.renaissance.com/2016/05/12/giving-meaning-to-test-scores/)—and some tests can provide both criterion-referenced results and norm-referenced results!

## How to interpret criterion-referenced tests

Criterion-referenced tests compare a person's knowledge or skills against a predetermined standard, learning goal, performance level, or other criterion. With criterion-referenced tests, each person's performance is compared directly to the standard, without considering how other students perform on the test. Criterion-referenced tests often use "cut scores" to place students into categories such as "basic," "proficient," and "advanced."



Criterion-referenced tests compare a student's knowledge and skills against a predetermined standard, cut score, or other criterion.

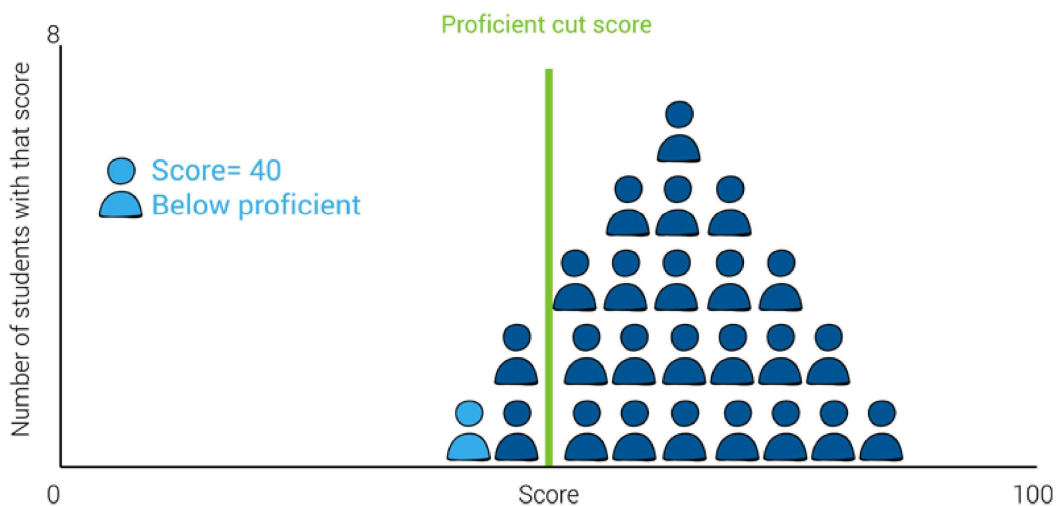
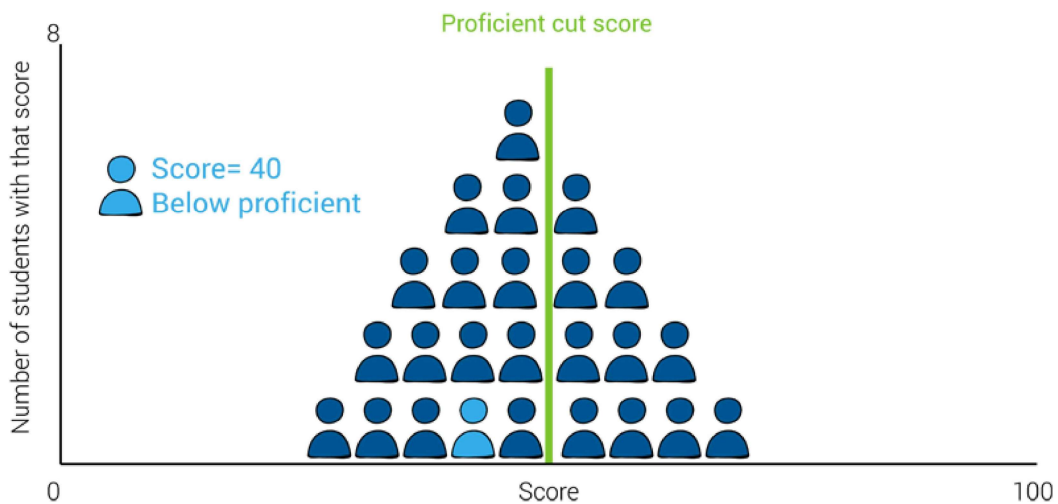
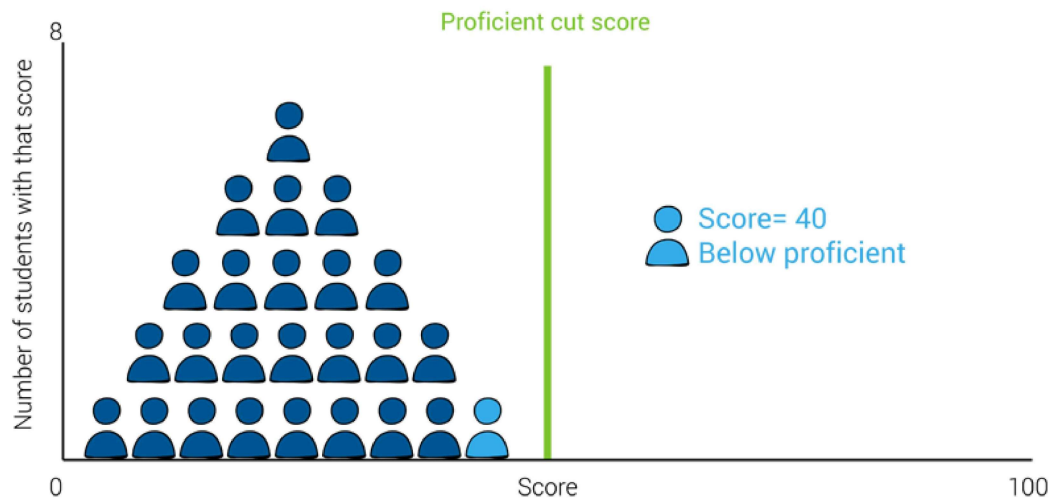
In criterion-referenced tests, the performance of other students does not affect a student's score.

If you've ever been to a carnival or amusement park, think about the signs that read "You must be this tall to ride this ride!" with an arrow pointing to a specific line on a height chart. The line indicated by the arrow functions as the criterion; the ride operator compares each person's height against it before allowing them to get on the ride.

Note that it doesn't matter how many other people are in line or how tall or short they are; whether or not you're allowed to get on the ride is determined solely by your height. Even if you're the tallest person in line, if the top of your head doesn't reach the line on the height chart, you can't ride.

Criterion-referenced assessments work similarly: An individual's score, and how that score is categorized, is not affected by the performance of other students. In the charts below, you can see the student's score and performance category ("below proficient") do not change, regardless of whether they are a top-performing student, in the middle, or a low-performing student.

On a criterion-referenced test, an individual student's score is not affected by the performance of their peers.

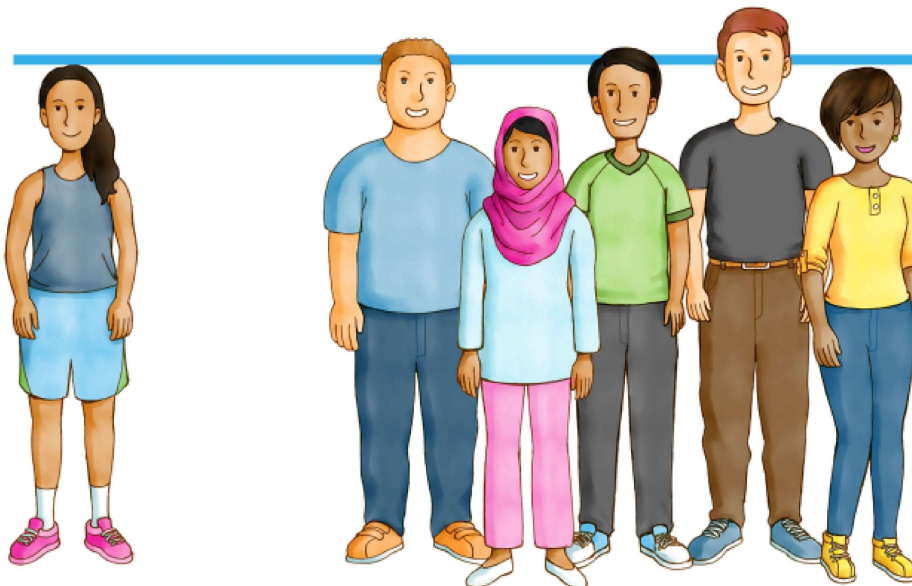


This means knowing a student's score for a criterion-referenced test will only tell you how that specific student compared in relation to the criterion, but not whether they performed below-average, above-average, or average when compared to their peers.

### How to interpret norm-referenced tests

Norm-referenced measures compare a person's knowledge or skills to the knowledge or skills of the norm group. The composition of the norm group depends on the assessment. For student assessments, the norm group is often a nationally representative sample of several thousand students in the same grade (and sometimes, at the same point in the school year). Norm groups may also be further narrowed by age, English Language Learner (ELL) status, socioeconomic level, race/ethnicity, or many other characteristics.

Norm-referenced tests compare a student's performance against the performance of their peers.



One norm-referenced measure that many families are familiar with is the baby weight growth charts in the pediatrician's office, which show which percentile a child's weight falls in. A child in the 50th percentile has an average weight; a child in the 75th percentile weighs more than 75% of the babies in the norm group and the same as or less than the heaviest 25% of babies in the norm group; and a child in the 25th percentile weighs more than 25% of the babies in the norm group and the same as or less than 75% of them. It's important to note that these norm-referenced measures do not say whether a baby's birth weight is "healthy" or "unhealthy," only how it compares with the norm group.

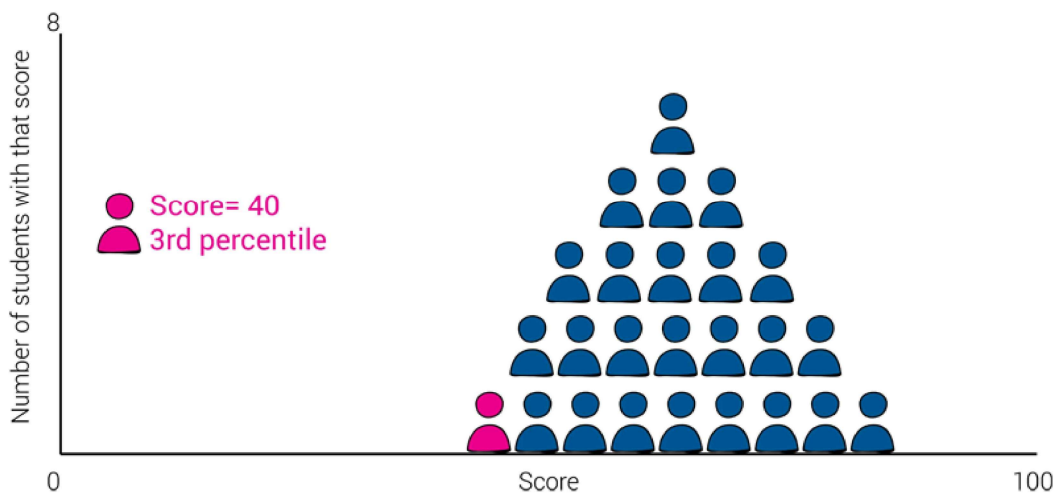
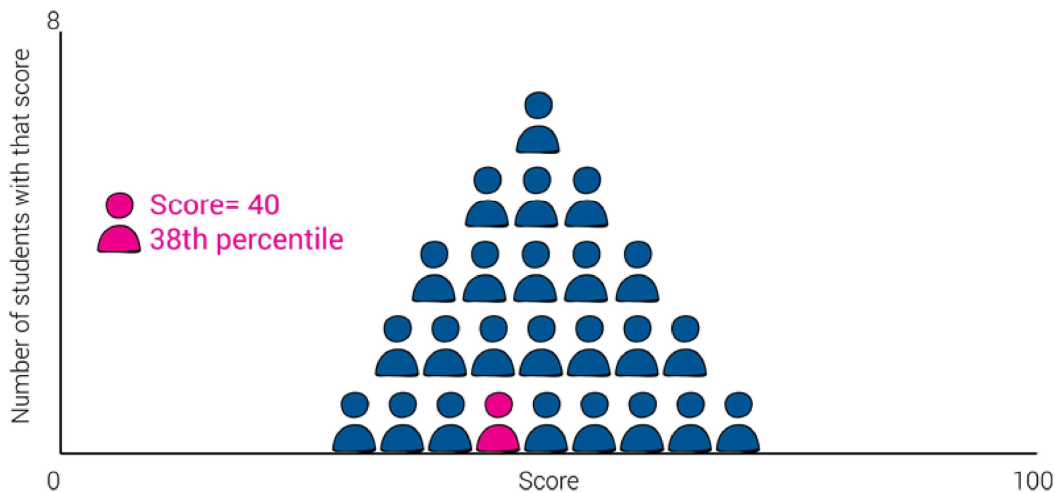
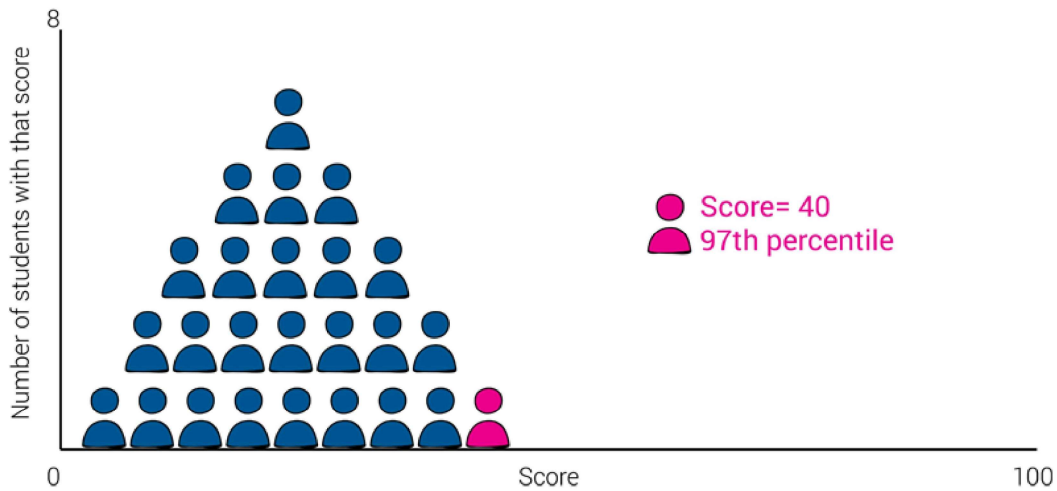
For example, a baby who weighed 2,600 grams at birth would be in the 7th percentile, weighing the same as or less than 93% of the babies in the norm group. However, despite the very low percentile, 2,600 grams is classified as a normal or healthy weight for babies born in the United States—a birth weight of 2,500 grams is the cut-off, or criterion, for a child to be considered low weight or at risk. (For the curious, 2,600 grams is about 5 pounds and 12 ounces.) Thus, knowing a baby's percentile rank for weight can tell you how they compare with their peers, but not if the baby's weight is "healthy" or "unhealthy."

Norm-referenced assessments work similarly: An individual student's percentile rank describes their performance in comparison to the performance of students in the norm group, but does not indicate whether or not they met or exceed a specific standard or criterion.

In the charts below, you can see that, while the student's score doesn't change, their percentile rank does change depending on how well the students in the norm group performed. When the individual is a top-performing student, they have a high percentile rank; when they are a low-

performing student, they have a low percentile rank. What we can't tell from these charts is whether or not the student should be categorized as proficient or below proficient.

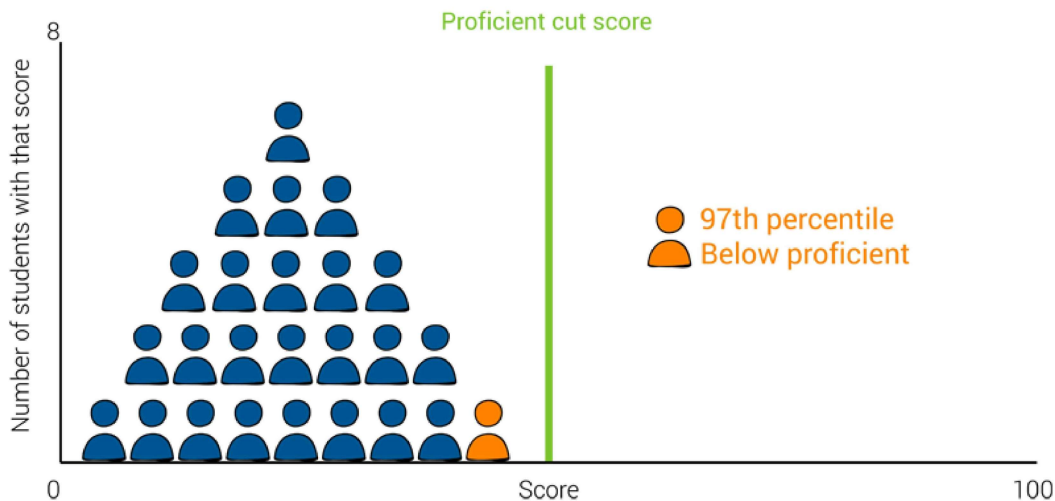
On a norm-referenced test, an individual student's percentile rank is calculated according to the performance of their peers.



This means knowing a student's percentile rank on a norm-referenced test will tell you how well that specific student performed compared to the performance of the norm group, but will not tell you whether the student met, exceeded, or fell short of proficiency or any other criterion.

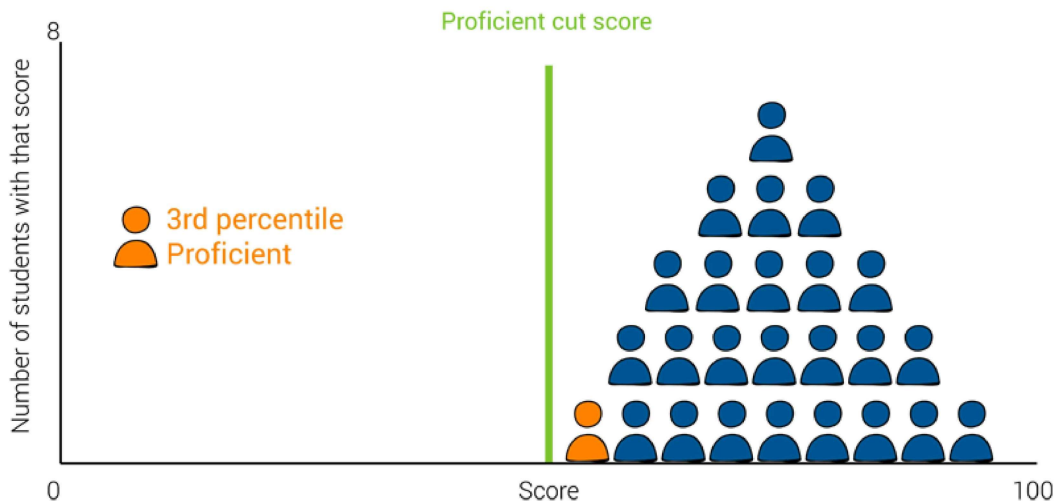
Some assessments provide both criterion-referenced and norm-referenced results, which can often be a source of confusion. For example, you might have a student who has a high percentile rank, but doesn't meet the criterion for proficiency. Is that student doing well, because they are outperforming their peers, or are they doing poorly, because they haven't achieved proficiency?

A student can have a high percentile rank but not achieve proficiency.



The opposite is also possible. A student could have a very low percentile rank, but still meet the criterion for proficiency. Is this student doing poorly, because they aren't performing as well as their peers, or are they doing well, because they've achieved proficiency?

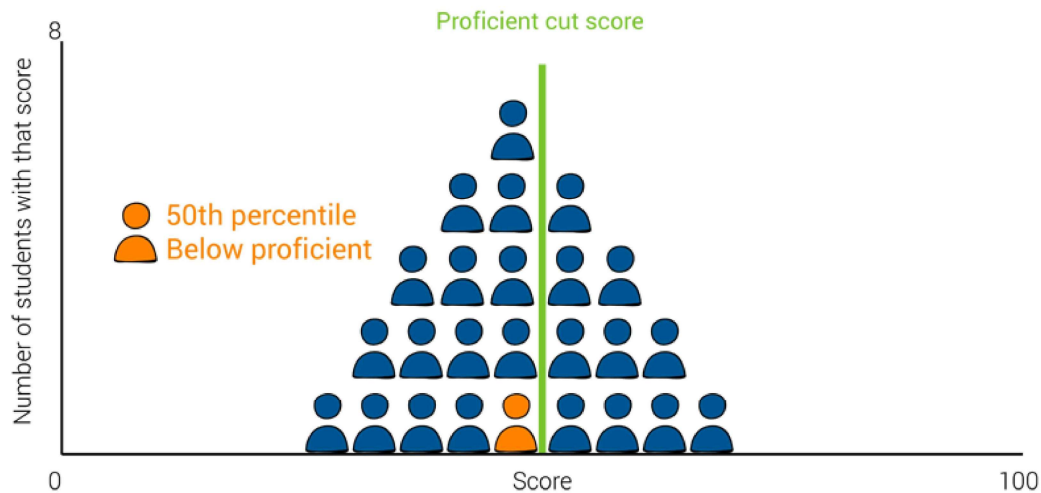
A student can have a low percentile rank but still achieve proficiency.



However, these are fairly extreme and rather unlikely cases. Perhaps more common is a “typical” or “average” student who does not achieve proficiency because the majority of students are not achieving proficiency. In fact, this is the pattern we see with [National Assessment of Educational Progress \(NAEP\) scores](https://www.nationsreportcard.gov/reading_2017/#/nation/scores?grade=4) ([https://www.nationsreportcard.gov/reading\\_2017/#/nation/scores?grade=4](https://www.nationsreportcard.gov/reading_2017/#/nation/scores?grade=4)), where the “typical” fourth-grade student (50th percentile) has a score of 226 and the “average” fourth-grade student (average of all student scores) has a score of 222, but proficiency requires a score of 238 or higher.



A student may be “typical” according to their norm-referenced measures but not achieve proficiency according to criterion-referenced measures.



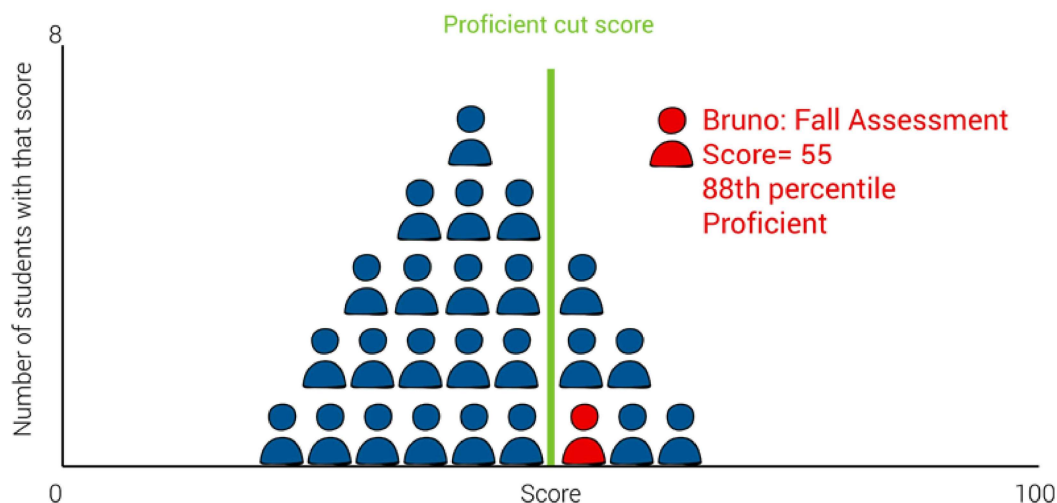
In all of these cases, educators must use their professional judgement, knowledge of the student, familiarity with standards and expectations, understanding of available resources, and subject-area expertise to determine the best course of action for each individual student. The assessments—and the data they produce—merely provide information that the educator can use to help inform decisions.

### What happened to Bruno?

So what happened to Bruno in the scenario described at the beginning of this post?

In the fall, Bruno scored 55 out of 100 on his district's assessment. The district had set the cut-score for proficiency at 50, meaning that Bruno counts as “proficient.” The district's assessment provider compared Bruno's score of 55 to the fall scores of their norm group, and found that Bruno scored higher than 88% percent of his peers in the norming group. This gives him a percentile rank of 88.

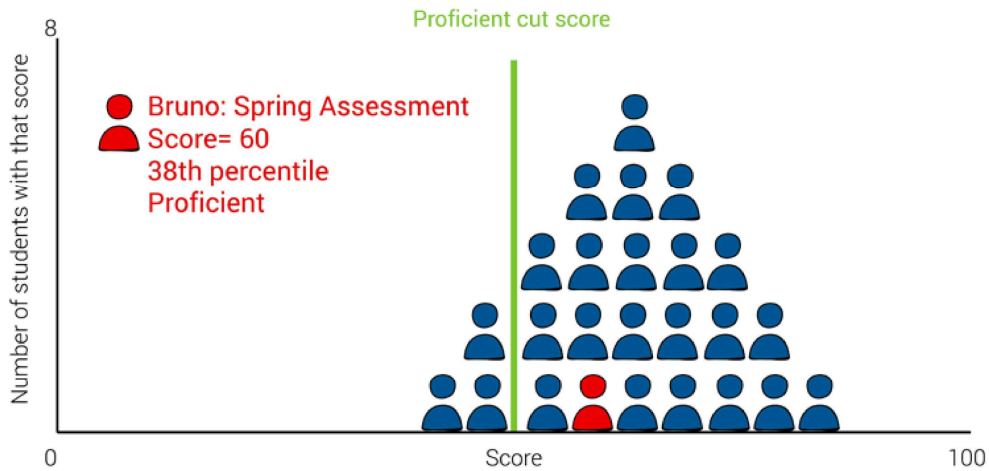
In the fall, Bruno scored 55 on the district assessment, putting him in the 88th percentile. He is categorized as proficient.



In the spring, Bruno takes the same test again. This time he scores 60, higher than this fall score. Since the district's criterion for proficiency hasn't changed, he is still categorized as proficient.

Just like Bruno, students in the norm group took the assessment twice—once in the fall and once in the spring. This time, the district’s assessment provider compares Bruno’s spring score to the spring scores of the norm group. In this case, the students in the norm group had notable gains and scored much higher in the spring than they did in the fall. Because students in the norm group generally had much larger gains from fall to spring than Bruno did, Bruno’s spring score now puts him at the 38th percentile.

In the spring, Bruno scored 60 on the district assessment, putting him in the 38th percentile. He is categorized as proficient.



For Bruno’s teacher, this is a sign of concern. Although Bruno is still categorized as proficient, he’s not keeping up with his peers and may be at risk of falling behind in future years. In addition, if the district or state raises the criterion for proficiency—which can happen when standards or assessments change—he might fall short of that new criterion and struggle to make enough gains in one year to meet more rigorous expectations.

This is one reason why it’s important for educators to monitor growth in addition to gains.

### The importance of Student Growth Percentiles (SGP)

Gains are calculated by taking a student’s current score and simply subtracting their previous score. Gains indicate if a student has increased their knowledge or skill level, but do not indicate if a student is keeping up with their peers, surging ahead, or falling behind. For that, a growth measure is needed.

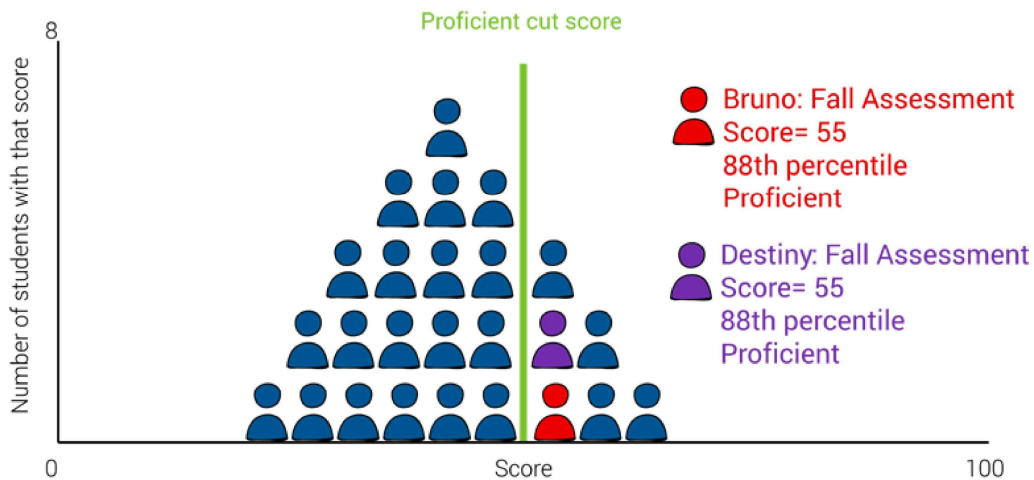
Growth—specifically a Student Growth Percentile or SGP (<https://www.renaissance.com/resources/student-growth-percentile/>)—is a norm-referenced measure that compares a student’s gains from one period to another with the gains of their academic peers nationwide during a similar time span. Academic peers are defined as students in the same grade with a similar score history, which means low-performing students are compared to other low-performing students and high-performing students are compared to other high-performing students.

As a result, SGP helps educators quickly see if a student is making typical growth, or if they are growing much more quickly or much more slowly than their academic peers. SGP also allows teachers to see if two students with the same score are truly academically similar or if they actually

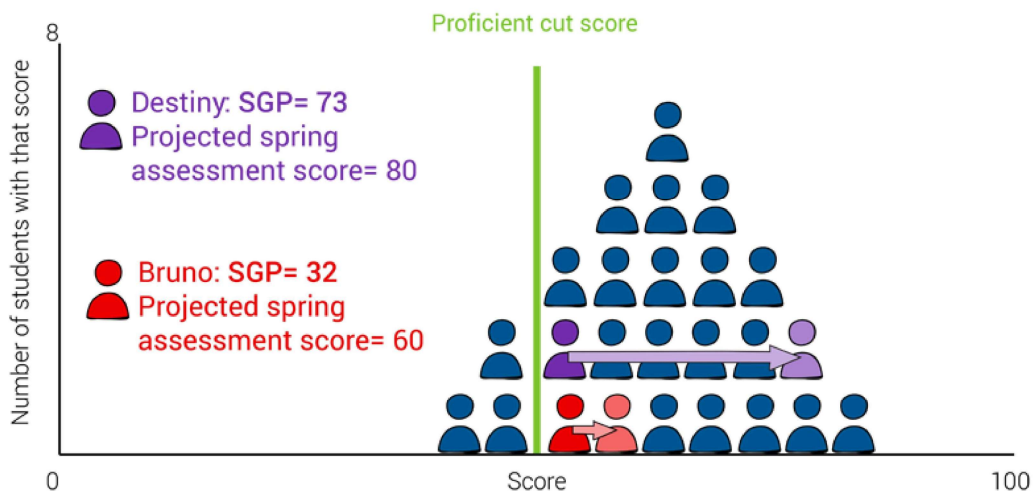


have very different learning needs.

Looking only at scores and percentile ranks, Bruno and Destiny appear to be very academically similar students.



However, Bruno and Destiny have very different SGPs. Their SGPs reveal that Bruno may need additional instruction or support.



In Bruno's case, knowing his SGP in the fall would have allowed his teacher to see that he has been making slower-than-expected growth. At this point, she could have proactively worked to boost his growth—perhaps by giving him additional practice opportunities, assigning him to a different instructional group, providing more targeted supports or scaffolding during lessons, or pairing him with a higher-performing student for peer tutoring. She might have also decided to assess him more frequently, perhaps every two or three months throughout the school year, to monitor his gains and growth more closely.

These efforts may have helped Bruno to end the school year the same way he started it—as a top-performing student—and be better prepared him for the challenges of the next grade.