



Time: 14 minutes 26 seconds

Index

- Lesson Overview
- Video Outline
- Your Learning Objectives
- Your Stops
- Suggested Discussion Questions
- Recommended Activities
- Suggested Resources
- Script

Lesson Overview

In this lesson you'll be introduced to:

- scaled scores and performance levels, the two primary metrics used in student score reports;
- how scaled scores and performance levels are created;
- why both scaled scores and performance levels are included in score reports;
- how to interpret and use scaled scores and performance levels;
- recommended activities and suggested resources (which are expanded on in this facilitator guide).

Video Outline

- Introduction (00:00-02:25)
- Scaled Scores (02:26-08:01)
- Performance Levels (08:02-13:49)
- Activities and Resources (13:50-14:26)
 - Recommended Activities
 - Suggested Resources

Your Learning Objectives

Record your objectives and points to focus on.

Your Stops

Make notes on stopping points and content discussion you would like the participants to take part in.

Stopping Point	Content Discussion	Notes

Suggested Discussion Questions

- What are raw scores and scaled scores?
- Why do testing programs report scaled scores?
- What would be the limitation(s) of reporting scaled scores only?
- What are performance levels? What is their relationship to scaled scores?
- Why do testing programs report performance levels?
- What would be the limitation(s) of reporting performance levels only?

Recommended Activities

- Find an individual student's parent report from a statewide or commercial general achievement test that provides the student's scaled scores and performance levels along with other performance data for the class or school. See how the content of this lesson applies to the report.
- Find a friend or colleague to whom you can practice explaining the student's results. Describe to that person how the student performed by explaining the meaning of the student's scaled scores and performance levels in light of the other information available.
- Do the same with a parent of one of your students, using that student's parent report.

Suggested Resources

Tan, X. and Michel, R. Why do standardized testing programs report scaled scores? *R&D Connections*, No. 16, September 2011. Princeton, NH: Educational Testing Service.

https://www.ets.org/Media/Research/pdf/RD_Connections16.pdf

Smarter Balanced Assessment Consortium. Reporting scores. *Smarter Balanced Validity Research*.

<https://validity.smarterbalanced.org/scoring/>

ScriptSlide 1

- In this lesson we'll discuss scores from standardized tests, which districts may purchase from testing companies or which may be required by state departments of education for purposes of school accountability.
- These are the external interim and summative tests we discussed in Lesson 1.

Slide 2

- Throughout the school year, students take different types of tests for different purposes. And the performance of students on those various tests is documented and reported in different ways.
- This graphic is a display from a parent score report for a standardized test. It shows the student's test score and performance level, average scores of groups of students, and information about how student scores are distributed across performance levels.
- In this lesson you will learn why these components are included in the score report, how they are determined, and how to interpret them.

Slide 3

- A good understanding of standardized test scores is important for their accurate interpretation and appropriate use.
- Also, this understanding allows teachers to accurately and effectively explain student test results to parents and students.

Slide 4

- This lesson is organized into two main sections.
- The first section introduces the concept of scaled scores, how they are created, and why they are important.
- The second section introduces the use of performance levels in standardized test reporting, how they are created, and tips for their interpretation.
- We will then provide a lesson summary as well as activities and suggested resources.

Slide 5

- Scaled scores and performance levels are the primary metrics used to communicate and document how students did on standardized tests.
- A scaled score is just a number representing the performance of a student on a particular test. Scaled scores look very different across different tests, for instance, the SAT and ACT, as shown here.
- A performance level describes the performance of students whose scores fall within a particular range, for instance, proficient and advanced.
- Understanding the subtleties underlying these results and how they are produced will make you a better consumer of standardized tests.

Slide 6

<blank>

Slide 7

- Test scores are only meaningful in a relative sense.
 - Relative to the scores of others.
 - Relative to some pre-established standard.
 - Or relative to previous performance.
- In other words, test scores are only meaningful when compared to other test scores.

Slide 8

- The daughter in this picture earned a 36 on a test. But what does this mean?
- The mother would need more information about the test results to know how her daughter did. For instance, the number of possible points on the test and what the average score of her daughter's class was.

Slide 9

- 36 may have been a score on a teacher-made test – the number of points earned or the percentage of possible points earned.
- In the case of standardized tests, the number or percentage of points earned is considered a “raw” score. And, in fact, 36 might be a reasonable raw score on a standardized test.

Slide 10

- Commercial and state testing programs are interested in lots of comparisons. They want to allow a student's score on a test to be compared to average scores of groups of students, for example, students in the same grade in the state or students in the same subgroup within a school. Or perhaps, the student's score needs to be compared to the score the student achieved on a previous test.

- To accomplish this, the testing companies perform “scaling and equating.” Scaling is just transforming raw scores on a test to a different numerical scale.

Slide 11

- We’re all familiar with one particular scale transformation. Lots of thermometers do it for us when they show the temperature in both degrees Fahrenheit and degrees Celsius. They’re just using two different scales to express what the temperature is.
- Whether we report a temperature as 40 degrees Celsius or 104 degrees Fahrenheit, it’s the same temperature. And, of course, there’s a formula we can use to make these conversions ourselves.

Slide 12

- Educators and parents see their kids’ scaled scores all the time – scores from the lowas, STARS [stars], MAPs [maps], and any number of other testing programs.
- High schooler Betina Johnson’s scores show numbers that are very different across the different programs. So what do they mean?
- We know that higher scores are better than lower scores for any particular program, so Betina performed better on the SAT math than reading.
- But we need to know a lot more about SAT scores to understand how Betina actually did.
- And how could we compare her performance on the SAT the first time she took it with her performance the second time if she took a different form of the test?
- Again, scale transformation and equating by the testing company offer the solution.

Slide 13

- Sometimes different forms of a test are used – different forms taken by different students during the same administration or different forms from different times during the school year.
- Different forms in the same subject might vary a little in terms of difficulty, so they have to be equated.
- In essence, equating is the same as adjusting scores on different forms to account for differences in difficulty.
- You only equate tests measuring the same content.
- Scaling and equating is a multi-step process that ultimately leads to scores on the desired reporting scale.
- Take a look at the graphic. It shows what scaling and equating accomplish. Notice raw scores on the harder test were transformed to higher scaled scores than the same raw scores on the easier test.

Slide 14

- In fact, when testing companies scale tests, they actually perform more than one score transformation.
- The first transformation is not a simple one – it takes care of transforming a scale and equating at the same time.

- The second transformation (or sometimes a third) gets the set of scores on whatever scale the testing companies or state testing officials choose to use for reporting.
- In both cases, however, the relative ordering of scores and the appropriate relative “distances” between students’ scores are preserved.

Slide 15

- Scaled scores on tests are just the scores that result from transformations of students’ raw scores.
- If the school average score was 30 on the test our student scored a 36 on, then it would seem that our student did pretty well.
- If those scores were raw scores on a standardized test, they might have been transformed to 351 and 362, which is represented in the parent report shown here
- Here the testing company likely transformed our third grader’s score, and those of all the other third-grade students who took the test, to a scale such that almost all the student scores fell between 300 and 399. Again, the relative ordering of scores and the relative distances between them would remain the same.

Slide 16

<blank>

Slide 17

- This report display, which shows only scaled scores, enables the viewer to make many comparisons. It does not say anything, however, about what our student scoring 362 knows or is able to do in the subject tested.

Slide 18

- The desire and need for information about a student’s knowledge or capabilities has led to greater emphasis on standards-based assessment. “Standards-based” characterizes today’s state and commercial tests in two ways.
- First, tests address curriculum or content standards. These are statements about what students at different grade levels should know or be able to do.
- Second, “standards-based” refers to the use of performance levels (also called achievement levels) in the reporting of test results. This is required of state testing by federal law. Knowing about the capabilities of students who score at different points along the score continuum is necessary for educators and policy makers to determine what scores are “good enough” for various purposes or decisions, such as student promotion and graduation.

Slide 19

- A performance level report for the student we looked at before might look like this. The names of the different levels might vary across states or testing programs. For example some states’

performance level names express the degree to which academic expectations are met by students – exceeds, meets, partially meets, and fails to meet.

- According to this report display, the student with a scaled score of 362 would be considered “proficient,” or at least performing at the proficient level. It’s clear from the graphic that three “cut points” on the score scale have to be determined in order to divide student performance into four levels. How performance-level reporting addresses the need to understand what students with various scores know or are able to do should become clearer when we discuss how cut scores are determined.

Slide 20

- The identification of cut scores is accomplished by a process called “standard setting.” This involves panels of educators and non-educators making judgments about student work on a test or about test item requirements.

Slide 21

- The standard setting judges match the student work or test items to previously developed performance level descriptions, which describe the capabilities of students at the various levels. Those descriptions are more detailed than the general ones in this table. But they would be consistent.
- For example, a descriptor for proficient could claim, among other things, that the student could represent data and mathematical relationships in multiple forms . . . or use a variety of reasoning methods to solve problems.

Slide 22

- The aggregation of the judges’ decisions leads directly to recommended cut scores.
- Cut scores ultimately have to be approved by some policy-making body, for instance, a state board of education.

Slide 23

- Because state tests are equated across years, standard setting only needs to be done in the first year of a testing program. Cut scores would remain the same after that.
- Importantly, this means that the capabilities of a student earning a particular score one year would be the same as a student who earned the same score the next.
- In other words, the capabilities students need to demonstrate to be considered proficient in a subject remain unchanged across years.

Slide 24

- This report display shows school and state percentages of students at various levels.
- Once raw score cut points are determined, the last transformation of scores to the desired scale can take place so that nice numbers (like 340 and 360) can be used for two of the cut points.

- For statewide testing required by federal law, the critical statistic for a school or state is the percentage of students designated as proficient or above – in this case, 34+12 or 46 percent for the school and 30+10 or 40 percent for the state.

Slide 25

- In some ways, information is lost when we move from scaled scores to performance levels. As you can see in this graphic, two students scoring at the opposite extremes of the same performance level, say “Basic,” would have performed very differently on the test, yet would be assigned to the same performance level with the same description of capabilities.

Slide 26

- In the same vein, two students scoring close to the same cut score, but just on opposite sides, would be put in different performance levels with different descriptions of capabilities. This is why testing programs usually report both scaled scores and performance levels in parent reports. It is very appropriate for state accountability testing to focus on the percentage of students in a school who are designated “proficient or above.” But a parent’s statement, “My child is proficient in math, but basic in reading,” does not tell the full story.

Slide 27

- Scaled scores and performance levels.....both important.....both with advantages and disadvantages.....and both important to understand.

Slide 28

<blank>

Slide 29

- These activities give teachers the opportunity to apply what they learned about standardized test reporting by explaining real test results to a colleague for practice, and then to a parent.

Slide 30

- The Kahl paper provides a more in-depth discussion of the points made in the video lesson.