

Proficiency Sample Ratings: Being Average Has Never Been So Good

By [Brian Johnson](#), AASHTO Accreditation Program Manager
Posted: October 2010

So you opened up your email notification to see that the latest AASHTO re:source proficiency sample ratings were just posted, you log into the website to view your ratings (Figure 1), and you see ratings of **, -5, -3, 5,4. You think to yourself, "I know that 4 and 5 are good, but what about the negative numbers? Those are below 3, so they must be bad... and what are the stars for? I doubt they're like the stars that my elementary school teacher used to give me... and what is this repeatability rating?"

Sieve Analysis

Total Material Passing the 2.36-mm (No. 8) Sieve (percent) - T27/C136

[View Youden Diagram](#) | [View Performance Chart](#)

	Sample 167						Sample 168					Repeatability (within-lab)		
	Total Labs	Lab Data	Avg	1S	Z-Score	Lab Rating	Lab Data	Avg	1S	Z-Score	Lab Rating	1S	Z-Score	Lab Rating
4	1448	86.0	85.40	0.38	1.56	3	84.1	83.58	0.39	1.34	4	0.24	-0.22	-5

-  = the line of data on the report
-  = the total labs that submitted data just for that line of data
-  = the results of sample 167 showing the lab's data, the average of all labs' data, 1 standard deviation, the z-score, and the rating
-  = the results of sample 168 showing the same information as shown for sample 167
-  = the repeatability, which shows you how close your z-scores were to each other

Figure 1: A Typical Line of Proficiency Sample Data (Color-Coded)

Calculating Averages and Standard Deviations

The first thing that you should understand is that laboratory ratings are based on the average of the results, although the reported averages are determined only after removing invalid and outlier results. It is important to eliminate them from the rating determination equations so that the ratings are not affected based on what some might consider to be "bad data." We determine a standard deviation for each data set (displayed as "1S" in Figure 1 above) and then begin the process of calculating ratings.

Calculating Z-Scores and Ratings

Each laboratory is rated with two values: a z-score and a lab rating. In statistics, the z-score, also known as the standard score, indicates how many standard deviations a result is from the average. The z-score is determined by the following calculation:

$$\text{Z-Score} = \frac{\text{(Laboratory Test Result - Average Value)}}{\text{(Standard Deviation)}}$$

The laboratory rating calculation is based on the absolute value of the z-score:

- If Z-Score <= 1 Then Rating = 5
- If Z-Score > 1 And <= 1.5 Then Rating = 4
- If Z-Score > 1.5 And <= 2 Then Rating = 3
- If Z-Score > 2 And <= 2.5 Then Rating = 2
- If Z-Score > 2.5 And <= 3 Then Rating = 1
- If Z-Score > 3 Then Rating = 0

Which Way Is Up?

If you're confused by all of this, check out Figure 2 below for a graphical representation of z-scores and ratings. Here are a few quick points to remember:

- Low z-scores are good.
- High ratings are good.
- A negative sign on a z-score or laboratory rating merely indicates that the laboratory's result was below the average, while a positive z-score or rating indicates that the laboratory's result was above the average.

Simply put, the closer your result is to the average, the better your rating. In the competitive world we live in, being average conjures up words like commonplace, mediocre, or ordinary; but in the world of proficiency testing, being average is the definition of excellence!

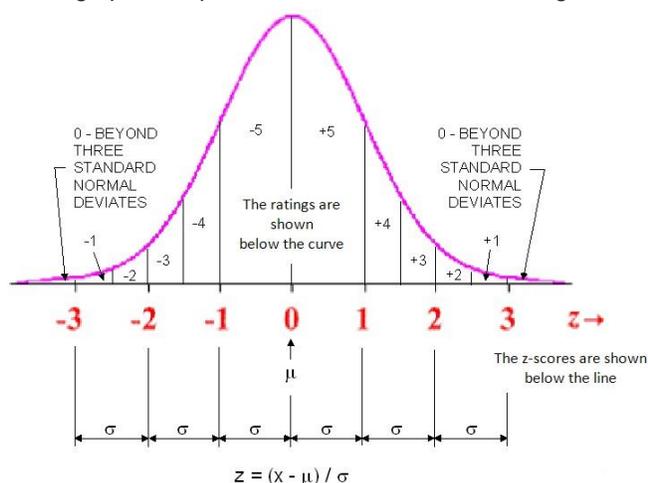


Figure 2: The Normal Distribution of AASHTO re:source Proficiency Sample Data

Low Ratings

Any rating less than a 3 (z-score > 2) is considered a low rating according to the AASHTO Accreditation

Program, but don't let that bother you unless you consistently receive low ratings. (See Figure 3 and the section below on Performance Charts.) Yes, low ratings are worth investigating, and you might even uncover an equipment problem or procedural mistake. Sometimes, however, your investigation of low ratings will lead you nowhere, and that's okay. The laws of statistics govern that some laboratories have to get low ratings - every lab will be on the low side of the ratings every once in a while. When an AASHTO-accredited laboratory receives low ratings for a given test, they are required to perform a root cause analysis and implement corrective action. If the laboratory receives low ratings again for that test, it might be a sign that either the corrective action was not effective or that the laboratory did not actually apply any corrective action. Now that you understand the concept of ratings, let's discuss a couple of other items that cause confusion.

The ** Rating

The ** rating indicates that the test results have been suppressed. Ratings may be suppressed for several reasons, but usually this is an indication of one of three things: 1) The data collected was for informational purposes only and is not a measure of the laboratory's competency, 2) data received is unusual and does not fit a normal distribution, or 3) there were not enough data points to provide an accurate analysis.

Repeatability (Within-Lab)

Ratings Repeatability is an estimate of the variation in results that you might expect if you repeated the same test over and over in your laboratory. The within-lab rating is based on the difference between the two individual lab results, but also any actual differences between the two sample materials.

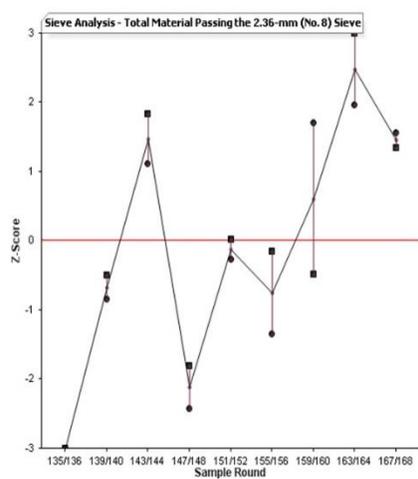


Figure 3: A Sample Performance Chart

Performance Charts

Performance charts provide an easy way to gauge your laboratory's proficiency testing performance over time (see Figure 3). As stated above, too much emphasis should not be placed on an occasional low rating. However, patterns in performance charts should be analyzed carefully, as they are usually good indicators of testing problems. The ideal scenario is to have all points over the center line - results right on the average time after time. Generally speaking, however, points scattered within the bands of +2 and -2 are indicative of good testing performance. Points drifting away from the centerline and points consistently on one side of the centerline are indicative of performance problems.

Now What?

I'm glad you asked. You've just learned all that you need to know about the proficiency sample program and how the results are reported. Now you have to take that knowledge and use it to get the most out of the program. You'll be reviewing your results, repeatability ratings, performance charts, and taking meaningful corrective actions so that you can score 5's and -5's - and you'll be more excited than ever to be average!