

## Precision Estimates and the AASHTO re:source Proficiency Sample Program: Why Precise Results Aren't Always Satisfactory

By [Tracy Barnhart](#), Quality Manager  
 Posted: April 2014

*Why did I receive low ratings on the samples even though my results meet the precision estimate of the test method? That is bogus! Okay, I made up the "bogus" part, but you get the general idea. Why on Earth does AASHTO re:source disregard the precision estimate information when determining each laboratory's proficiency sample ratings? It seems crazy, doesn't it? Allow me a chance to explain the method to our madness.*



Before we dive in, let's begin with some basic information. If you haven't done so already, you might want to check out this article: [Proficiency Sample Ratings: Being Average Has Never Been So Good](#). If you don't feel like reading it now, or if you just need a refresher, here are some things you should know about the AASHTO re:source Proficiency Sample Program (PSP) final reports:

- AASHTO re:source determines the **grand average** and **standard deviation** (i.e. 1s) for each data set
- The **z-score** indicates how many standard deviations (1s values) a lab's individual result is from the grand average of the results from all participating laboratories (excluding invalid and outlier results)
- The **laboratory rating** is based on the absolute value (+ or -) of the z-score
- High ratings (3's, 4's, and 5's) and low z-scores are considered "good"

Here is an example page from a final report that shows this information:

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	
<b>4</b>	1448	<b>86.0</b>	85.40	0.38	<b>1.56</b>	<b>3</b>	<b>84.1</b>	83.58	0.39	<b>1.34</b>	<b>4</b>	0.24	<b>-0.22</b>	<b>-5</b>	

- = 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

Now let's move on to precision estimates. For those of you who may not be familiar with precision and bias, here are a couple of helpful articles: The Skinny on Precision and Bias: [Aiming Our Sights on Precision](#) and [Aiming Our Sights on Bias](#). If you have ever read a Precision & Bias (P&B) statement in an AASHTO or ASTM test method, you may have noticed that many of the estimates of precision are based on the results from the AASHTO re:source PSP. Oh, what a tangled web we weave...



In a nutshell, precision is a measurement of how close test results are to each other. Precision is measured in terms of repeatability and reproducibility. Repeatability (or single-operator precision) is the precision of test results obtained with the same test method, in the same laboratory, by the same person, with the same equipment. Reproducibility (or multi-laboratory precision) is the precision of test results obtained in different laboratories using the same test method. In other words, repeatability is within-lab precision, while reproducibility is between-lab precision. Together, repeatability and reproducibility establish the upper and lower limits for the precision of a test method.



Whew! Now let's dig in a little deeper. The standard deviation (1s value) in the test methods and the rating system for the AASHTO re:source PSP are determined in a similar manner. That explains the references to the AASHTO re:source PSP results in many Precision & Bias statements. Although the 1s values may look different between the test methods and the AASHTO re:source PSP reports, they are not incompatible. So what's the catch? Well, the information is used for somewhat different purposes and is presented slightly differently.

For those of you that have never seen a P&B statement before, here's one from AASHTO T 11.

## 12. PRECISION AND BIAS

- 12.1. *Precision*—The estimates of precision of this test method listed in Table 1 are based on results from the AASHTO Materials Reference Laboratory Proficiency Sample Program, with testing conducted by this test method and ASTM C 117. The significant differences between the methods at the time the data were acquired is that T 11 required, and ASTM C 117 prohibited, the use of a wetting agent. The data are based on the analyses of more than 100 paired test results from 40 to 100 laboratories.

**Table 1**—Precision

	Standard Deviation (1s), <sup>a</sup> percent	Acceptable Range of Two Results (d2s), <sup>a</sup> percent
Coarse aggregate: <sup>b</sup>		
Single-operator precision	0.10	0.28
Multilaboratory precision	0.22	0.62
Fine aggregate: <sup>c</sup>		
Single-operator precision	0.15	0.43
Multilaboratory precision	0.29	0.82

<sup>a</sup> These numbers represent the (1s) and (d2s) limits as described in ASTM C 670.

<sup>b</sup> Precision estimates are based on aggregates having a nominal maximum size of 19.0 mm (3/4 in.) with less than 1.5 percent finer than the 75- $\mu$ m (No. 200) sieve.

<sup>c</sup> Precision estimates are based on fine aggregates having 1.0 to 3.0 percent finer than the 75- $\mu$ m (No. 200) sieve.

To further clarify:

- 1s = standard deviation = how much variation there is from the average value
- d2s = acceptable range of two results = 2.8 x 1s

As explained above, the d2s values in the test method precision statements provide guidance for comparing two different test results. In contrast, the AASHTO re:source PSP is directed toward each laboratory being able to evaluate its own testing performance by comparing its results to the “correct” (or average) value for that material. AASHTO re:source PSP ratings are then determined by how far a laboratory's result is from the “correct” value. The d2s values in the test methods are not appropriate for that purpose. I will explain further later. In the meantime, consider that two test results can be very close to one another in value (i.e. precise), yet be very far from the average value (i.e. not accurate).



Also, it is important to note that the 1s values stated in test method precision estimates usually aren't applicable to any one specific material. Rather, the precision estimates are typically ballpark estimates applicable to a range of materials. Therefore, some judgment must be used when applying the estimates to different materials. For example, the 1s value for sieving a gravel with rounded particles would likely be different than the 1s value for sieving a crushed stone containing angular and elongated particles.

For some sample materials used in the AASHTO re:source PSP, using the 1s values noted in the test methods would have too much variability to provide useful information for laboratories about their own testing. Conversely, these values would be too confining for other types of materials. **The advantage of the AASHTO re:source PSP is that it determines the standard deviation specific to the material that was used for a specific round of testing.** The standard deviation provides a sliding scale that automatically adjusts for any testing bias resulting from the material being tested. For materials that are “easy” to test (such as granite) and show similar results

between laboratories, the 1s value will be smaller. For materials that are more “difficult” to test (such as shale) and show higher variability between laboratories, the 1s value used for comparison will generally be a larger.

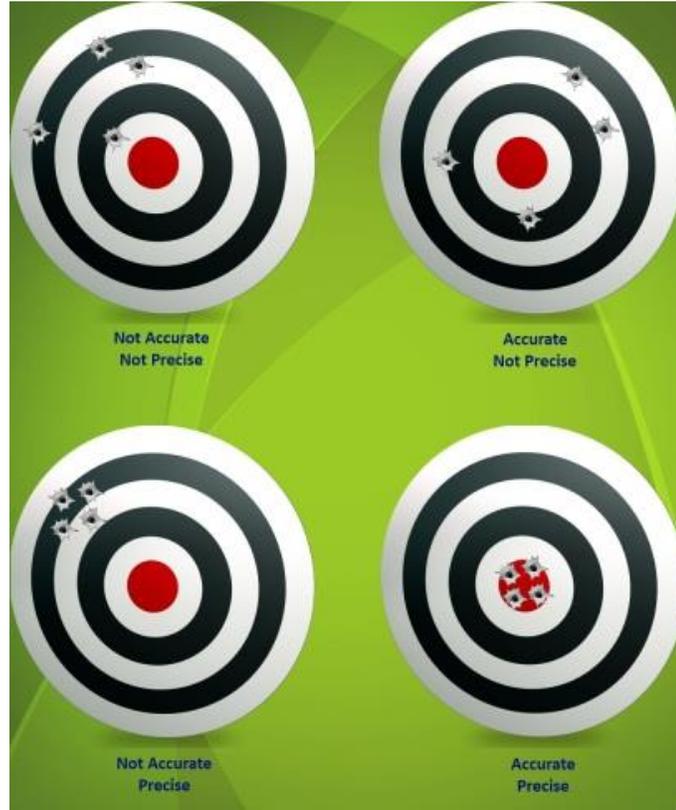


So, what’s the difference between the d2s results in the precision statement of a test method and the way that the AASHTO re:source PSP ratings are determined? Both use the standard deviation of a set of test results from a group of laboratories that are all testing the same material. Both apply the same level of scrutiny or evaluation. Both compare results based on assumptions about how the results would be dispersed for testing that is properly performed. Sounds identical, doesn’t it? Well, once again, the purposes and criteria are a little different. **The**

**AASHTO re:source PSP examines the accuracy of a laboratory’s test result, while the test method precision statement examines the precision of two results compared to one another.** Remember, accuracy is a measurement of how close a test value is from the true (or “correct”) value while precision is a measurement of how close individual results are from each other. Here is a handy graphic that represents the difference between accuracy and precision:

The AASHTO re:source PSP and the AASHTO Accreditation Program (AAP) evaluate participating laboratories based on how far their results are from the “grand average.” The grand average is used as the “correct” (or “expected”) result, and laboratories are judged by how close their results are to the “correct” results for a properly conducted test on that particular material. This is an examination of the accuracy (not precision) of test results.

The precision estimate in a test method uses the “acceptable range of two results” (d2s) to evaluate how close the test results are to one another, regardless of what the “correct” result should be. This is an examination of the precision (not accuracy) of test results. That being said, keep in mind that it is possible for a pair of test results to be precise but not accurate. For example, if a balance isn’t providing correct readings and that balance is used repeatedly for a set of tests, the test results might be very close to the same value, (i.e. “precise”) but that measurement might be very inaccurate.



*So, why can’t the test method precision statement be used to compare a test result to the “average” test result? If a result lies within the “acceptable range of two results” when compared to the “average” result, why isn’t that good enough?* The reason is that the test result in question satisfies the **precision** requirement, but not the **accuracy** requirement. The reverse is also true. It is possible for two results to be **accurate** enough for the AASHTO re:source PSP, but not **precise** enough to meet the precision of the test method.

Here’s an example:

Test Method: ..... AASHTO T 11  
 Material Type: ..... Coarse Aggregate  
 Average T 11 Result from PSP Report ..... 1.570  
 1s value from PSP report: ..... 0.209  
 d2S value from T 11 P&B statement  
 (multi-laboratory precision): ..... 0.62

**PSP Results That Are Accurate but Not Precise**

**(i.e. satisfactory PSP ratings, but results do not meet d2S value)**

Laboratory	Lab Data (%)	Z-Score	Rating
A	1.98	+1.96	+3
B	1.16	-1.96	-3
<b>Difference between results: 1.98 – 1.16 = 0.82</b>			
<b>0.82 (difference between results) &gt; 0.62 (d2S value)</b>			

The reverse can also occur...

**PSP Results That Are Precise but Not Accurate**

**(i.e. results meet d2s value, but unsatisfactory PSP ratings)**

Laboratory	Lab Data (%)	Z-Score	Rating
A	2.53	+4.59	0
B	2.21	+3.06	0
<b>Difference between results: 2.53 – 2.21 = 0.32</b>			
<b>0.32 (difference between results) &lt; 0.62 (d2S value)</b>			

**Why?**

The ratings methodology currently in use by the AASHTO re:source is considered to be fair and useful for the purpose of the AASHTO re:source Proficiency Sample Program and the AASHTO Accreditation Program. These programs have a common goal which is to judge a laboratory's testing results by comparing them to the average of a large body of results for a specific material. Since the precision estimates stated in the test methods are not based on an average result for a specific material, applying the test method precision estimates to the AASHTO re:source PSP results would not be appropriate. If you have any further questions, please contact John Malusky at [jmalusky@ashtoresource.org](mailto:jmalusky@ashtoresource.org) or 240-436-4830.

**References**

- AASHTO, "AASHTO T 11, Standard Method of Test for Materials Finer Than 75-µm (No. 200) Sieve in Mineral Aggregates by Washing," *Standard Specifications for Transportation Materials and Methods of Sampling and Testing*, Part 2A: Tests, 2013.
- Johnson, Brian. "Proficiency Sample Ratings: Being Average Has Never Been So Good." AASHTO re:source *In-Focus Newsletter*. AASHTO re:source, Oct. 2010. Web. 07 Apr. 2014.
- Norris, Rob. "The Skinny on Precision and Bias: Aiming Our Sights on Precision." AASHTO re:source *In-Focus Newsletter*. AASHTO re:source, Apr. 2012. Web. 07 Apr. 2014.
- Price, Evan. "The Skinny on Precision and Bias: Aiming Our Sights on Bias." AASHTO re:source *In-Focus Newsletter*. AASHTO re:source, Oct. 2012. Web. 07 Apr. 2014.