

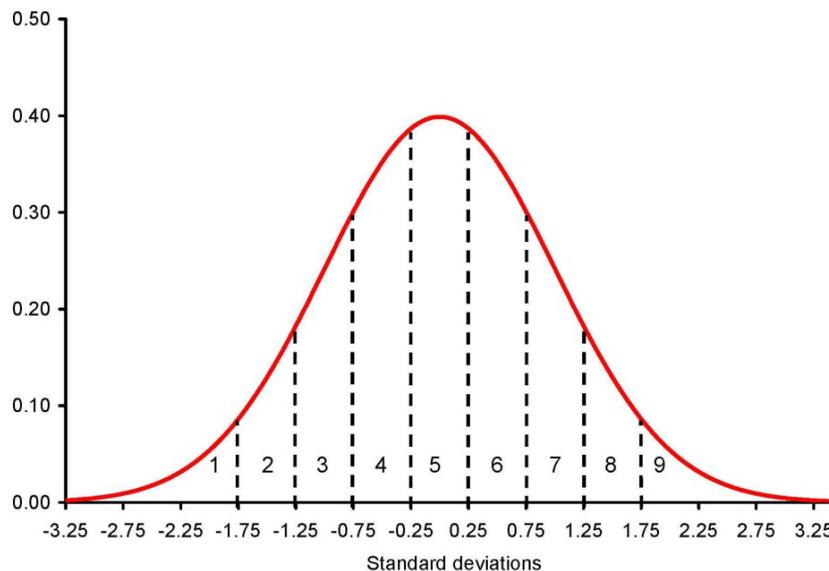
2012 PLTW End of Course Score Interpretive Information

End of Course Test Score

PLTW is utilizing a new test construction process in order to report scores on a norm-referenced scale. This allows for a simpler score reporting procedure and comparison of individuals and classes with their national peers (Angoff, 1971). The current AE, CEA, and POE tests were developed utilizing this new method. One of the goals of the process is to build a test that has a normal distribution of observed scores and a mean percentage score of approximately 50% (Wendler & Walker, 2006). This will cause a shift in the percent correct score distribution for these tests compared to previous years. As PLTW tests are reconstructed over the next couple of testing cycles, the assessments will fall in line with this model. Starting in 2012, PLTW will report norm-referenced stanine scores for all End of Course exams. These stanine scores will be referred to as “End of Course Scores”, and will range from one (1) to nine (9) with one (1) being the lowest possible score and nine (9) being the highest possible score. This score structure will help simplify communication and provide useful interpretation of student performance.

Stanines

The stanine score scale is based on the theory that student achievement is distributed normally in a large enough population. When students take the tests, their scores will reflect their achievement levels and form a normal distribution of test scores. As seen in the figure below, the score distribution is then broken up into nine categories, ranging from 1 to 9, based on how many standard deviations the student’s score was from the mean. A score of 1 indicates student performance at the lowest level and a score of 9 indicates student performance at the highest level. The average student performance across the population is represented by a score of 5. This distribution, therefore, is formed with a mean score of 5 and a standard distribution of 2 (Seashore, 1955).



How PLTW Calculates End of Course Scores

Norm-referenced scores can be used to make more appropriate delineations of student performance compared to raw or percent correct scores. However, in order to calculate the End of Course Score for each student, a conversion from the raw score scale to the norm-referenced End of Course Score scale must be made. This section outlines the process by which PLTW calculates norm-referenced End of Course Scores for test users. The process begins by using the following formula to calculate z-scores for each student:

$$\mathbf{z - score} = \frac{\mathbf{Observed\ Score\ (x) - National\ Mean\ (\mu)}}{\mathbf{National\ Standard\ Deviation\ (\sigma)}} = \frac{\mathbf{x - \mu}}{\mathbf{\sigma}}$$

The resulting z-scores will be set on a scale that has a mean of 0 and a standard deviation of 1, with most z-scores falling between -3 and +3. Once the individual z-scores are calculated, they simply need to be assigned to End of Course Scores based on the categories outlined in the following table.

Z-score to End of Course Score (Stanine) Conversion Table

End of Course Score (Stanine)	Minimum z-score	Maximum z-score	% of population in each stanine	Cumulative % of population
1	$-\infty$	-1.75	4%	100%
2	-1.75	-1.25	7%	96%
3	-1.25	-0.75	12%	89%
4	-0.75	-0.25	17%	77%
5	-0.25	+0.25	20%	60%
6	+0.25	+0.75	17%	40%
7	+0.75	+1.25	12%	23%
8	+1.25	+1.75	7%	11%
9	+1.75	∞	4%	4%

All students with a z-score between negative infinity and -1.75 are assigned an End of Course Score of 1, all students with a z-score between -1.75 and -1.25 are assigned a 2, etc. through a score of 9. Note that z-scores should not be rounded to just two decimal places, as this would lead to confusion in score assignment when rounded to a dividing score (i.e. -0.25). A minimum of 4 decimal places for z-scores is advised.

Transforming Scores to Grades and Calculating Class Averages

There is no set protocol for assigning letter grades or cut scores to specific End of Course Scores. Most importantly, the transformation needs to follow the school or teacher's grading policy. A common way the scale is broken up is by separating it into three categories of three scores each: 1-3 is designated as below average, 4-6 as average, and 7-9 as above average.



It is not recommended to assign letter grades based on the percent correct scores for PLTW assessments. In the POE example below, student H scored a 24, the highest observed score in the class. If a teacher simply calculated the percent-correct scale ($24/40=60\%$) and used that value along with the standard high school score scale (90%-100%=A, 80%-89%=B...) the student would improperly be assigned a D while being in the above average category.

An administrator or teacher can also look at the aggregated data to compare how well a class or school did as a whole. Teachers can determine how their class scored compared to every other POE class in the country, for example, by summing the individual End of Course Scores from their class and dividing by the number of students. For a precise measure, a z-score can be calculated for the class to determine how many standard deviations from the mean the class scored. In order to do these calculations, the national mean and standard deviation should be set at the stanine scale values of 5.0 and 2.0, respectively. This is shown in the example below.

End of Course Score Example

A POE classroom has 9 students and the national mean for the POE Part A test was 19.0 with a standard deviation of 5.6095. A table showing the conversion from POE Part A raw scores to End of Course Scores is as follows:

Sample POE Part A Raw Score to End of Course Score Table

Student ID	Observed Score	z-score calculation	z-score	EoC Score
A	19	$(19-19)/5.6095$	0	5
B	21	$(21-19)/5.6095$	+0.3565	6
C	22	$(22-19)/5.6095$	+0.5348	6
D	19	$(19-19)/5.6095$	0	5
E	15	$(15-19)/5.6095$	-0.7131	4
F	23	$(23-19)/5.6095$	+0.7131	6
G	20	$(20-19)/5.6095$	+0.1782	5
H	24	$(24-19)/5.6095$	+0.8913	7
I	14	$(14-19)/5.6095$	-0.8914	3

Using the three verbal scales mentioned above, Student I scored below average, students A, B, C, D, E, F, and G scored about average, and student H scored above average. Note that since the raw score scale is being compressed from a 41 point (0-40) scale to a 9 point (1-9) scale, students do not have to answer the same exact number of items correctly to get the same End of Course Score.



Again, as indicated before, a teacher can also look at the aggregated data to compare how well the class did as a whole. Teachers can determine how their class scored compared to every other POE class in the country by summing the individual End of Course Scores and dividing by the number of students. For example, $(5+6+6+5+4+6+5+7+3)=47$, and $47/9=5.2222$. Therefore, with an aggregated score of 5.2222, this particular class scored about average. To be more precise, the z-score can be calculated by the following calculation: $(5.2222-5.0)/2=.1111$, indicating that the class scored .1111 standard deviations above the national mean.

PLTW Welcomes Your Feedback

PLTW is committed to providing a high-quality and rigorous assessment program for our courses, as we believe that student learning and achievement can be strengthened through the use of reliable assessments that allow for valid score interpretations. As we continually improve our processes we plan to increase the number of resources that help students, teachers, and administrators understand the purpose and use of the assessment tools. In order to provide materials that are useful to the PLTW community, we encourage feedback from all participants and will work to provide assessment resources to help our students and teachers succeed.

References:

Angoff, W.H. (1971). Scales, Norms, and Equivalent Scores. In R. L. Thorndike, (Ed.), *Educational Measurement* (2nd ed.) (pp.508-600). Washington, D.C.: American Council on Education.

Downing, S.M., & Haladyna, T.M. (Eds.). (2006) *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

O'Connor, D.P. "Stanine Scale." Image. *PEP 6305 Measurement in Health & Physical Education*, 1 June 2012.

<http://grants.hhp.uh.edu/doconnor/pep6305/topic%20005%20normal%20distribution.htm>

Seashore, H.G. (1955). Methods of Expressing Test Scores. *Test Service Bulletin*, n48, 7-10.