

Increasing Meaningfulness - Scale Calibration Explanation

End of Year Math

November 2009

One challenge that educators have related to assessments is to establish a level of meaningfulness among staff and students as to why assessments are beneficial. Although there is a point at which we can assess too much, the point at which we reach “too much” occurs when there is no reason for the assessment – no useful information comes of it. The reason we assess, whether its once or multiple times, is to gain useful information.

Information that appears most valued is related to:

1. that which tells us what the student has learned and hasn't learned,
2. that which tells us how we compare among a more broad population (be that the nation, state or district), and
3. that which give us some type of predictive information.

This document will try and provide information on the method we will use to calibrate our local End of Year mathematics tests to the mathematics CST. The intent is to provide administration further background so that they may have another source of information to generate meaningful, constructive, reflective and critical discussion among their staff related to the EOY test and directing instruction.

This approach should help address the two following statements:

“The test does not seem to correlate well with the STAR” – At times students may score Far Below Basic on a local assessment, but had performed at the Advanced level on the state test. This may be based no a variety of factors, and cannot be entirely eradicated; nevertheless, there seems to be a desire to bring a closer alignment to the performances on both. There seems to be a desire to establish some type of predictive validity.

“What do we do with the information?” or, “What’s the purpose of the test?” – All assessments give some type of information, and results from one test can generate different information depending on what is looked at. Classroom assessments that reflect performance on particular standards is considered the most effective source to gain information to guide instruction. Information gained at a more broad level (i.e. the common EOY test) is most useful in attempting to identify trends that can be used in various ways during discussions.

Classroom or school-site level assessments provide important information on achievement at the student level. However, since the number of scores is limited to 20 or 30 for a classroom, or 100 for a grade level, any trend-type of information that can be used for predicting future performance becomes unstable due to the small number in the population. Trend information should only be based on a larger group of students.

The following pages will outline a method used this year to bring a closer alignment between performance on the mathematics EOY test and the mathematics CST.

We will begin to establish this alignment by looking at the students' STAR performance, and using the following steps:

- In the first step we rank order, from the lowest score to the highest score, all students per grade level based on their STAR performance. This may give us a population of nearly 2000 students.
- We can then establish a percentile ranking of all those students.
- We then identify the percentile rank band that most closely represents the different performance bands (Far Below Basic, Below Basic, Basic, Proficient and Advanced). Below is an example of last year's 4th grade math scores (these students are now in grade 5). The percentile rank range, raw score (number of items answered correctly for each performance band), the number of students that fell within each band, and the percent of students that fell into each band is also given.

STAR GRADE 4 MATH (This year's Grade 5 Students)				
	Percentile Band	Raw Score	Students	Percent
Far Below Basic	0% - 3.40%	11 - 19	83	4.11%
Below Basic	4.10% - 22.00%	20 - 34	393	19.47%
Basic	23.50% - 43.00%	35 - 47	440	21.80%
Proficient	45.40% - 66.50%	48 - 56	489	24.23%
Advanced	69.60% - 98.70%	57 - 65	613	30.38%
			2018	

The next step involves following the same process with a local assessment.

Since the students who took the grade 4 CST are now in grade 5, and most recently took the grade 5 End of Year test, we will use an example of the grade 5 End of Year results. Here's the caveat:

We know that the student have not received instruction on much of the material at this point in time. This would imply that we need to look at the data with a critical lens and try to tease out some useful information. When this process is repeated closer to the end of the academic year, we would expect shifts to occur in the information seen. It should be understood that the students as a whole may not have done well on the first administration of the End of Year test when compared to the STAR; however, generally speaking the students who did well on the STAR should still do better on the EOY test than students who performed poorly on the STAR if the construction and administration of the EOY test are sound (this speaks to the reliability of the test).

We can begin our calibration process by beginning in the same manner we had earlier with the STAR, but this time we use the EOY results. Our intent is to identify the similar percentile bands:

- First we rank order all the student grade 5 EOY scores from the lowest score to the highest score.
- We then establish a percentile ranking of all scores.
- If our desire is to calibrate based on these students' scores from the pervious year's STAR (grade 4 STAR), then we use the percentile rank bands from the grade 4 STAR and shift them, as closely as possible, over to the EOY percentile range. In other

words, we break up the percentile ranking of the EOY test to match as closely as we can the percentile ranking of the previous year's STAR.

The following table reflects the percentile bands shifted over. Notice that the bands will not necessarily be exact, since the break-points are dependent upon how many students receive repeated scores (i.e. how many kids get 17 correct). The table reflects the percentile band, the number of items correct on the EOY test to reach that band, the number of students that would fall in that band if this performance band scale is used, and the percent of students that this would represent. This percentage can be compared to the percentage of students that fall within each band for the previous year's STAR.

EOC GRADE 5 MATH (Scaled to previous Grade 4 STAR)				
	Percentile Band	Raw Score	Students	Percent
Far Below Basic	0% - 3.80%	0 - 7	63	4.22%
Below Basic	4.20% - 20.50%	8 - 15	311	20.83%
Basic	25.00% - 43.90%	16 - 19	369	24.72%
Proficient	49.70% - 66.60%	20 - 23	323	21.63%
Advanced	71.40% - 100%	24 - 56	427	28.60%
			1493	

Of course the preceding two tables do not suggest that the 19.47% of the students who were below basic on the grade 4 STAR were the same as the 20.83% of students who were below basic on the End of Year test. Regardless of the approach used in attempting to calibrate the performance levels of the two tests, teachers will notice that there will always be students that do not match performances within the bands of the two tests. Some factors that contribute to this mismatch may include the following:

- The time at which the local assessment was given with respect to the amount of instruction taking place.
- The performance level of the student at the time of earlier administered assessments. One would expect more stability at the higher ability level ranges when compared to the lower score ranges.
- The motivation level of the student taking the test. If there was an elevated focus in the environment on performing on the state test, this may cause some students to perform better on that test than on the local test.

The best that can be done is to try and establish the bands, and construct the test, so that more students match bands between two tests when comparing performance at the same grade level. Notice that in this example, we are comparing performance at two different grade levels, and this means students were measured using two distinct sets of standards.

Nevertheless, the following table provides an indication of how individual student performances match between performance bands of the grade 5 End of Year (labeled EOC on the table) and the grade 4 STAR test. When considering factors such as the academic readiness level of the student at the point in time when the assessment was given, one would anticipate fewer students matching bands.

In an ideal situation we would want the diagonal reflected in yellow to maintain the largest number in each row. This would indicate that the largest number of kids within the re-established End of Year test matched in their performance on the STAR which measured standards of the previous year. In our example below, the Far Below Basic band showed the most inconsistency

since only two students were FBB on both the EOY/EOC test and the STAR. The most stable match occurred in the advanced range where 295 students who were advanced on the EOY test were also advanced on the STAR the previous year.

MATH EOC Grade 5 vs. STAR Grade 4 Crosstabulation

Count		STAR Grade 4					Total
		1FBB	2BB	3B	4P	5A	
EOC	1FBB	2	11	14	12	15	54
Grade 5	2BB	22	101	72	54	35	284
	3B	17	92	102	90	43	344
	4P	6	40	81	110	72	309
	5A		4	33	87	295	419
Total		47	248	302	353	460	1410

This approach is retrospective and does not provide any predictive information since we look at the current EOY test and compare band performance to the previous grade's STAR band performance. To establish predictive information, we would need to look at the current grade EOY band performance (i.e. grade 5 EOY) and compare it with the same grade STAR band performance (i.e. grade 5 STAR). This can only be done in one of two ways. The most useful information will be gained when we keep the results from the current year (preferably the last assessment of the year, or one immediately preceding the STAR administration) and compare the performance to the performance of those same kids when they take the STAR that year. This would give matching scores comparing STAR with the EOY test after instruction has taken place. Obviously, we will need to wait until next year to do this.

The other option will still generate a degree of alignment, but can only be based on a prediction since the scores cannot be matched student-to-student. We can take the current grade performance on the EOY (lets say grade 5) and compare it to the performance of last year's grade 5 students on the STAR. Since last year's grade 5 students are this year's grade 6 students, the comparison involves different population of students, but would be a comparison within the same grade. In essence, this would entail calibrating the current EOY performance bands for grade 5 in this example, to the performance bands of last year's grade 5 STAR performance. The following table is how the band breakdown would look if we took the previous year's grade 5 STAR performance and forced it onto this year's grade 5 EOY test that was administered at the beginning of this year.

EOC GRADE 5 MATH (Scaled to previous Grade 5 STAR)

	Percentile Band	Raw Score	Students	Percent
Far Below Basic	0% - 8.20%	0 - 11	145	9.71%
Below Basic	9.70% - 31.30%	12 - 17	406	27.19%
Basic	36.90% - 49.70%	18 - 20	284	19.02%
Proficient	55.90% - 80.50%	21 - 26	409	27.39%
Advanced	83.30% - 100%	27 - 56	249	16.68%
			1493	

For the sake of comparison, the following table reflects the percentage of students that represented each band on last year's grade 5 STAR test:

STAR GRADE 5 MATH (This year's Grade 6 Students)				
	Percentile Band	Raw Score	Students	Percent
Far Below Basic	0% - 7.90%	8 - 21	185	9.57%
Below Basic	9.50% - 30.40%	22 - 31	436	22.54%
Basic	32.10% - 52.10%	32 - 41	425	21.98%
Proficient	54.10% - 79.20%	42 - 54	527	27.25%
Advanced	81.30% - 99.10%	55 - 65	361	18.67%
			1934	

Since the students in the previous two tables reflect distinct populations, we cannot obtain actual matched scores to determine how many scores fell in the same band when comparing the EOY and the STAR.

Calibrating with the intent to provide some level of predictability between local assessments (i.e. EOY) and the STAR test requires monitoring the data over a broad scale of students. A primary reason that common assessments are recommended includes providing a foundation on which we might be able to compare classroom and school-level performance. One desired value in formative-type assessments is to provide information that teachers can use to try and guide instruction. Formative assessments are typically at the classroom level. District or state level assessments provide information for the sake of comparison. The value of the information that is gained from formative assessments is increased when they can be compared to a general, broader, perspective.

We will provide calibrated district-level information after administration of the End of Year test in mathematics in attempts to provide another option to view information with the purpose of increasing the meaning of the assessments by generating constructive dialogue among staff.