APPENDIX IV – 47

Explanation of Methodological Improvements for 'B' Teachers in the Teacher Evaluation 2017-18

Growth Component of the Teacher Evaluation

The State requires that 33% of a teacher's overall annual evaluation be based on student growth. Growth is measured as the change in academic performance of a teacher's students across two points in time. Change in academic performance is measured by comparing the results for individual students on the AzMERIT ELA and Math tests.

Teachers are designated into one of two groups: 'A' teachers or 'B' teachers. The groups are defined as:

- An 'A' teacher is any K-2 teacher with fall and spring DIBELS or DRA scores. Grades K-2 will use the DIBELs, DRA or some other assessment to compare the literacy fall results to the spring results.

- An 'A' teacher is also any teacher who teaches Math or ELA in grades 3 – 11. All elementary teachers in grades 3 – 5 are considered 'A' teachers because they teach both Math and ELA. All Math and ELA teachers in grades 6 – 11 are 'A' teachers.
  - ELA: Grades 3 – 8 and ELA 9 - 11 are used because they are assessed by AzMERIT
  - Math: Grades 3 – 8 and Algebra 1, Geometry, and Algebra II are used because they are assessed by AzMERIT

- A 'B' teacher is any teacher who is not an 'A' teacher. For example, a 6th grade science teacher or a 12th grade AP chemistry teacher are considered 'B' teachers. The 'B' teachers will be assigned growth points based on the school or the district average

- Any teacher will fewer than 15 students with matched scores will receive the school average. 'A' teachers will receive the school average for their subject area and 'B' teachers will receive the school average for Math and ELA.

The conceptual shortcoming with the calculation of growth scores for 'B' teachers is twofold: teachers are not evaluated on their own students but rather on the entire school population and teachers are not evaluated in their own subject area but rather in subjects they do not teach (Math and ELA). This issue is compounded because an effective 'B' teacher who teaches in an ineffective school can receive an ineffective growth score. Conversely, an ineffective 'B' teacher can receive an effective growth score because s/he teaches in an effective school. Addressing this lack of validity in the 'B' teacher's growth scores is paramount because more than 75% of our teachers are designated as 'B' teachers. The growth score for 'B' teachers therefore is not a true indicator of their teaching ability. Nonetheless, TUSD is obliged to comply with the model that the State of Arizona has provided.

In light of the conceptual shortcomings, the growth score for 'B' teachers should be as neutral as possible. By neutralizing the growth score for 'B' teachers, the other components of the Teacher Evaluation (Danielson Observation by principals, Student Survey of Teachers, Self-Reflection) take on more prominence in the determination of a teacher's level of effectiveness.

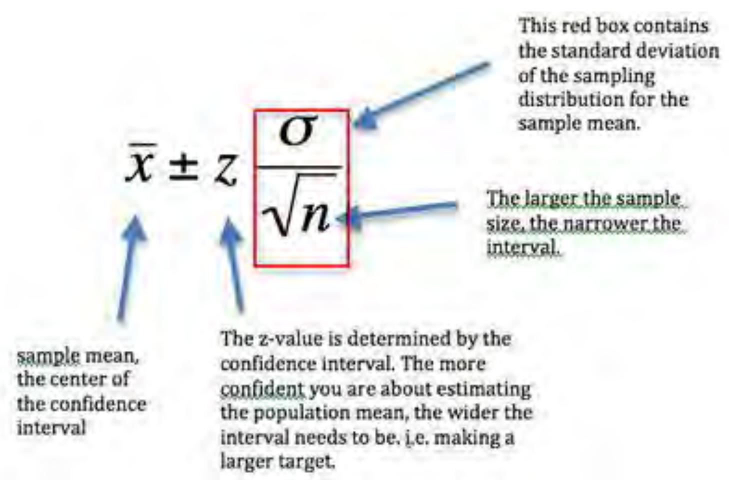
Explanation of Methodological Improvements for 'B' Teachers in the Teacher Evaluation 2017-18

Current Statistical Procedures

Assignment of growth scores for **'A' Teachers** (Math and ELA subjects) has been based on the creation of 95% confidence intervals around teacher mean student gains. This method is equivalent to a t-test of the null hypothesis where the question is whether the teacher's mean is significantly different from zero (the overall district mean). The confidence limits were established by using the **standard error of the mean (SEM)** formula below.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The SEM is the **standard deviation (SD)** of student gains in the teacher's classroom divided by the square root of the number of students in the teacher's classroom. If the number of students increases and the SD remains equal, the SEM will, in turn, decrease which results in a narrower confidence band around the mean.



In this scenario, the null hypothesis, denoted as H_0 states that the "True" mean is zero (district average). For a teacher to receive a growth score of either a 1 or a 3, the null hypothesis must be rejected. The possible results are:

- If we *cannot* reject the H_0 the teacher receives a value of 2, indicating an average (Effective) teacher.

- If we *can* reject H_0 we can assert that the teacher is not an average teacher:
  - If the teacher's average is significantly greater than zero, we can say with 95% confidence that the teacher is above average, resulting in a score of 3 (Highly Effective).
  - If the teacher's average is significantly less than zero, we can say with 95% confidence that the teacher is below average, resulting in a score of 1 (Ineffective).

Explanation of Methodological Improvements for 'B' Teachers in the Teacher Evaluation 2017-18

Discovering a "Truly" significant difference (i.e., rejecting H_0) depends largely upon the **power** or sensitivity of the statistical test. Power is defined as (1 – Beta) where beta is the probability of making a Type II error or falsely accepting H_0 when a significant difference truly exists. Therefore, (1 – beta) is the probability of correctly rejecting H_0 when a significant difference truly exists.

Power is largely influenced by the following factors: alpha level chosen, (.05), N-size, and variance. All things being equal, a larger N-size will result in a more powerful test (i.e., ability to detect a true difference).

The calculation of growth scores for 'B' teachers has been analyzed in the same manner as that for 'A' teachers in TUSD. The only difference in the two calculations has been that for 'B' teachers, the growth was based on the average student gain for the teacher's entire school. All things being equal, a higher probability exists of rejecting the null hypothesis for schools with a larger N-sizes because of the added **power**. In other words, 'B' teachers in the medium and larger schools within the district are more likely to receive a score of 1 or 3.

Decreasing the Statistical Power for "B" Teachers

To make the calculation of growth scores more equitable for 'B' teachers, an improved model has been designed that equates the statistical test in terms of its power. To produce greater equity in scores, the SEM was recalculated using the school's standard deviation and an average class size of 30. An impact analysis was conducted using the current year's growth data. The results for ELA and Math are presented below.

ELA *Growth based on School N*

Growth Score	# Schools Count	Percent Schools
1	10	11.76
2	57	67.06
3	18	21.18

ELA *Growth based on an N of 30*

Growth Score	# Schools Count	Percent Schools
1	0	0.00
2	80	94.12
3	5	5.88

Math *Growth based on School N*

Growth Score	# Schools Count	Percent Schools
1	23	27.06
2	39	45.88

3	22	25.88

Math *Growth based on an N of 30*

Growth Score	# Schools Count	Percent Schools
1	10	11.90
2	69	82.14
3	5	5.95

The results of decreasing the statistical power for 'B' teachers, the desired effect to reduce the number of schools being assigned a 1 or 3 was accomplished, and the number of schools receiving a 1 was reduced to zero. It should also be noted that the number of "Highly Effective" 'B' teachers was also reduced from 21% to 6% in ELA and 26% to 6% in Math.

In summary, the methodological improvement to produce greater equity of growth scores to 'B' teachers was accomplished by standardizing the standard deviation and N size of each school to produce an equivalent power across schools. It is predicted that this methodological change will result in a reduction of the number of Ineffective (1) and Highly Effective (3) 'B' teachers' growth scores, while not fundamentally changing the model for 'B' teachers. With a neutral growth score of 2, the final determination of the evaluation will rely more heavily on the other components (Danielson Observation by principals, Student Survey of Teachers, Self-Reflection).