Teacher Effectiveness Evaluation Model 2018-19

This report describes the Teacher Effectiveness Evaluation Model for 2018-19. Measuring teacher effectiveness requires multiple measures, both quantitative and qualitative to capture the range of instructional skills used in teaching and to determine how much students benefit academically from their teachers.

For 2018-19, TUSD has chosen to use a model to evaluate teacher effectiveness. The model is made up of four components including the Danielson Framework, Academic Growth, the Student Survey, and the Teacher Reflection. Each component factors into a teacher's final score, albeit with different weighting. The Danielson Framework comprises the majority of the score determination by making up 56% of the total score. The Academic Growth makes up 33% of the total score per the AZ State Board of Education's Statute #15-203(A)(38). The Student Survey makes up 10% of the total score and the Teacher Reflection is 1% of the total score. Each component is described below and how the points are determined.

Danielson Framework

The Danielson teacher evaluation framework uses 22 criteria nested within four domains. They are: Planning and preparation (N=6); the classroom environment (N=5); instruction (N=5); and professional responsibilities (N=6). Each of the 22 components is scored on a four point rubric:

1 = Unsatisfactory
2 = Basic
3 = Proficient
4 = Distinguished

The maximum number of points possible on the Danielson is 88 points (22 components X 4 pt. rubric).

Academic Growth

Academic growth will be determined by calculating the growth of state standardized scores in English Language Arts (ELA), and Math for grades 3-11 from one year to the next. This approach, however, has some limitations in that the state standardized tests in ELA and Math can measure the academic impact of only about a quarter of our teachers (called 'A' teachers). The non-ELA and non-Math teachers (called 'B' teachers) make up the other three-quarters of the teaching core.

A. What is an 'A' or a 'B' teacher in TUSD?

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

- An 'A' teacher is also any teacher who teaches math or ELA in grades 3 – 11. Elementary teachers in grades 3 – 5 are '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, if you are a 6th grade science teacher, you are considered a 'B' teacher. If you are a 12th grade AP chemistry teacher, you are considered a 'B' teacher. The 'B' teachers will be assigned growth points based on the school or the district average.

B. Who will take the assessment: All students in grades K – 2 will take the DIBELs or EDL assessment. In grades, 3 – 11, students will take the AzMERIT state test in ELA and math.

C. When will the assessment be administered: DIBELS (or another literacy tool) is administered three times a year. The first test in the fall and the last test in the spring will be used. EDL is administered twice a year, once in the fall and once in the spring. AzMERIT is administered in the spring each year.

D. Scoring: AzMERIT scores used are from last year (2017-18). Those scores are compared to AzMERIT scores from 2016-17 with a matched cohort so that students are compared against their own scores to measure growth. If a teacher changes schools in 2018-19, his/her academic growth score is still attached to where s/he taught the year before. For example, if a social studies teacher taught at Gridley in 2017-18 and then changed to teach social studies at Valencia for 2018-19, that teacher would receive still the school average for Gridley as his/her academic growth score in 2018-19. Teachers who teach at multiple schools will be assigned the district academic growth average. Student growth will be assessed on matched students by determining the difference between:

- Grades K – 2: the DIBELS and EDL scores are compared from the beginning of the year 2018-19 to the end of the year 2018-19.
- Grade 3: the AzMERIT 2017-18 scores are compared to the 2016-17 composite SchoolCity BM (a combined score from fall and spring) from 2nd grade.
- Grades 4 – 11: AzMERIT 2017-18 scores are compared to the AzMERIT 2016-17 scores.

E. Point Allocation: Teachers will receive a 1 or 1.5 (below average growth or a total of 11 or 16.5 points), a 2 (average growth or an average of 22 points), or a 2.5 or a 3 (above average growth or an average of 27.5 or 33 points) that will be added to the Teacher Evaluation points total.

- 'A' teachers with 15 students or more with 2 years of AzMERIT data will receive their own score.
    i. Grades K - 2: Scores are for ELA only
    ii. Grades 3 – 5: Scores are the average of the ELA and Math scores per teacher
    iii. Grades 6 – 11: Scores are from the subject (ELA or Math) specific to that teacher

- 'A' teachers in grades 3 – 11 with fewer than 15 students with 2 years of AzMERIT data (or SC Bench Mark for 2nd grade) will receive the school subject mean in which they teach.
- 'A' teachers in grades K – 2 with fewer than 15 students with fall and spring DIBELS or EDL scores will receive the DIBELS/EDL ELA mean.
- 'B' teachers who support math (math interventionist, AP calculus teacher, etc.) will receive the school math mean
- 'B' teachers who support ELA (literacy specialist, AP English lit, etc.) will receive the school ELA mean
- 'B' teachers who do not support ELA or math (PE teacher, art teacher, science teacher, etc.) will get the school mean which is a combination of the math and ELA mean.

Student Survey

The three Student Surveys are: Grades K-2, Grades 3 – 5, and Grades 6 – 12. Using the Tripod Study from Harvard University as the conceptual foundation, these surveys measure 7 classroom climate constructs including: Care, Challenge, Control, Clarify, Captivate, Confer, and Consolidate. Each survey has a different number of total questions. The K-2 Survey has 10 questions, the 3-5 Survey has 20 questions and the 6-12 Survey has 25 questions. Each of these 3 surveys is scored on a 4-point Likert scale:

1 = Strongly Disagree
2 = Disagree
3 = Agree
4 = Strongly Agree

Responses on the Likert scale are averaged and result in an overall score that ranges from 1 to 4. So, regardless of the grade level and/or number of questions, the score will be the averaged number from the responses.

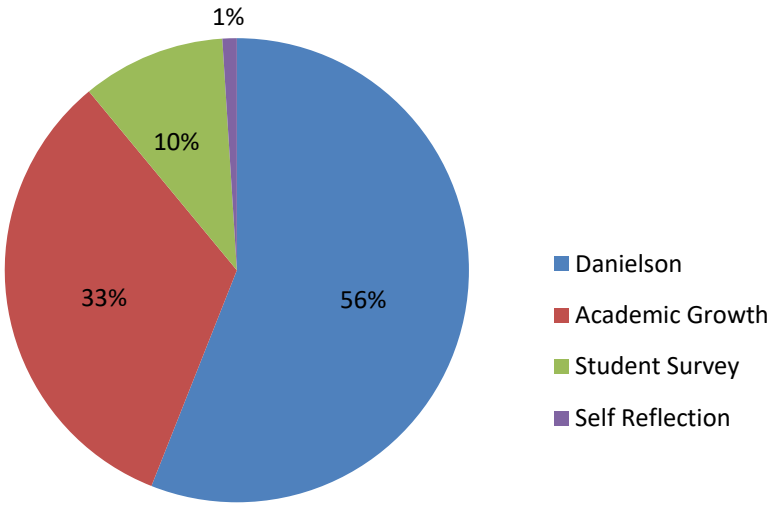
Teacher Self Reflection

The Teacher Self Reflection is completed by the teacher and is scored either 1 or zero depending on whether it was completed or not.

Teacher Effectiveness Evaluation Model 2018-19

Converting Raw Scores into Weighted Scores

Figure 4. Breakdown by Percent of the Four Components in the Teacher Model



Each component of this model carries a different weight as represented in the pie chart above. For example, the results of the Danielson observations are weighted the most heavily because they represent 56% of the total model. The results from the Danielson observations, therefore, will have the greatest impact on a teacher's overall score. Secondly, the academic growth represents 33% of the total model so that it can impact a teacher's overall score, but not necessarily determine the outcome. Finally, the results of the Student Survey (10%) and the Self Reflection Survey (1%) will have a smaller impact on a teacher's overall score.

To get the ration of the current maximum raw points to desired maximum points, we must divide the desired maximum points by the current raw maximum points. Calculating the ration using scaling factors will produce properly weighted components.

In Tables 1 - 3, the raw maximum points are converted into weighted or desired maximum points using a scaling factor. The scaling factor is derived by dividing the Desired Maximum Points (the weighted percent of each component that adds up to 100) by the Current Maximum Raw Points. The scaling factor, therefore, changes the raw points into the weighted points for each component.

Because the Desired Maximum Points always add up to 100, it does not matter how many raw maximum points are allocated on the Student Survey or the other components. The scaling factor will always change in response to a change in the maximum raw points of each component so that the weight (Desired Maximum Points) remains constant.

Teacher Effectiveness Evaluation Model 2018-19

Table 1. Grades K-12 Distribution of Points			
Component	Maximum Raw Points	Scaling Factor*	Desired Max Points
Danielson	88	.636	56
Academic Growth	3	11	33
Student Survey	4	2.5	10
Teacher Self Reflection	1	1	1
Total	96		100

** Scaling Factors are derived by dividing the Desired Points by the Maximum Points.*

The following examples show three different Grade 4 teachers with three different raw points. Their points were converted using the Scaling Factor Conversion to give the weighted points.

Teacher A – Grade 4

Table 5. Calculation of Points of a Teacher Scoring about Half of the Possible Points (Developing Teacher Status)			
Component	Raw Points	Scale Conversion	**Weighted Points**
Danielson	44	44 x .636	**28**
Academic Growth	2	2 x 11	**22**
Student Survey	2	2 x 2.5	**5**
Teacher Self Reflection	1	1 x 1	**1**
Total	49		**56**

Teacher B – Grade 4

Table 6. Calculation of Points of a Teacher Scoring about Average of the Possible Points (Effective Teacher Status)			
Component	Raw Points	Scale Conversion	**Weighted Points**
Danielson	73	73 x .636	**46**
Academic Growth	2	2 x 11	**22**
Student Survey	3.2	3.2 x 2.5	**8**
Teacher Self Reflection	1	1 x 1	**1**
Total	79.2		**77**

Teacher C – Grade 4

Table 4. Calculation of Points of a Teacher Scoring Most Points (High Effective Status)			
Component	Raw Points	Scale Conversion	**Weighted Points**
Danielson	77	77 x .636	**49**
Academic Growth	2	2 x 11	**22**
Student Survey	3.2	3.2 x 2.5	**8**
Teacher Self Reflection	1	1 x 1	**1**
Total	83.2		**80**

Cut Scores for 2018-19

The cut scores for 2018-19 are:

Ineffective	0 – 46 total points
Developing	47 – 60 total points
Effective	61 – 78 total points
Highly Effective	79 - 100 total points

Based on these cut scores, Teacher A above would be considered "Developing", Teacher B would be considered "Effective", Teacher C would be considered "Highly Effective". To be considered "Ineffective", a teacher would have to score low on the Danielson Framework, on the Academic Growth and/or Student Survey. The Teacher Self Reflection will have only a small impact on the overall score.

Teacher D – Grade 4

Table 7. Grades 3-5 Calculation of Points of a Teacher Scoring Some of the Possible Points			
Component	Raw Points	Scale Conversion	**Weighted Points**
Danielson	40	40 x .636	**25**
Academic Growth	1	1 x 11	**11**
Student Survey	2.75	2.75 x 2.5	**7**
Teacher Self Reflection	1	1 x 1	**1**
Total	44.75 or **45**		**44**

A methodological improvement was implemented in 2017-18 and will be continued. It provides greater equity of growth scores to 'B' teachers by standardizing the standard deviation and N size of each school to produce an equivalent statistical power across schools. This methodological change results in a reduction of the number of Ineffective (1) and Highly Effective (3) 'B' teachers' growth scores, while not changing the model for 'B' teachers. With a greater number of 'B' teachers receiving the neutral growth score of 2, the final determination of the evaluation relies more heavily on the other components (Danielson Observation by principals, Student Survey of Teachers, Self-Reflection).