

Teacher Effectiveness Evaluation Model 2015-16 (Draft K)

This report describes the Teacher Effectiveness Evaluation Model for 2015-16. 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. 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

In the past, academic growth has been determined by calculating the growth of state standardized scores in English Language Arts (ELA) and Math for grades 3-10 from one year to the next. The Arizona Department of Education determines labels for each school, ranging from A-F, that is based on student academic performance and growth. This approach, however, has 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. The 'B' teachers have been assigned growth points in the past based on the school or the district label.

This year, TUSD will make all teachers an 'A' teacher. Math and ELA teachers (formally known as 'A' teachers) will use the District's quarterly assessments in math and ELA to show academic growth. These quarterly assessments will be designed to measure growth over time. All other teachers (formally known as 'B' teachers) will administer pre-post assessments to their students that are relevant to their course material. The pre-post assessment strategy will be made up of a multiple choice assessment with a relevant content-specific reading passage that can measure academic growth for all course categories with the exception of math and ELA. The components are listed below:

- A. Courses: TUSD offers a variety of courses at the middle and high school levels including core academic courses, enrichment courses, and technical courses. These courses have been grouped into 41 umbrella categories. Each category encompasses multiple courses. For example, Physical Education is a category that includes body conditioning, yoga, tennis, etc.

- B. Pre-Post Assessment: The pre-assessment will contain one or two short reading passages and up to 10 multiple choice questions that relate to the passage. Each category will have its own passage that is relevant to the content and the standards of the category. These themes of these passages may be similar across grades but will increase in complexity with each subsequent grade. An example of a theme in history/American government, etc might be a passage reflecting on the concept of what constitutes a human 'right' in modern society. The post-assessment will use the same assessment as the pre or the questions may be replaced with parallel questions. Parallel questions are questions of the same difficulty that measure the same concept but do not ask the same question. Parallel questions can be used to measure growth.
- C. Development of the pre-post category assessments: Grades K-2 will use the DIBELS assessment and compare the fall results to the spring results. Grades 3 – 5 and math and ELA teachers in grades 6 – 10 will use the quarterly assessments as their pre-post assessment. The remaining courses in grades 6 – 12 will use category assessments developed by Curriculum and Instruction Department in conjunction with District teachers in the summer 2015. Teams of teachers from all grades and content areas will be asked to participate in the development of these pre-post assessments. All assessments will be standard's based and aligned to the content of the category. Additionally, our psychometric specialist will work with the district's contracted assessment company to ensure that the pre-test and the post-test are parallel in difficulty. Items will be taken from the assessment company's item bank and/or teachers will develop their own questions. All assessments will be completed prior to the start of the 2015-16 school year.
- D. Who will take the assessment: All students in grades K – 2 will take the DIBELS assessment and in grades 3 – 5 will take the quarterly benchmarks. In grades, 6 – 12, pre-post category assessments will be administered by a sampling strategy so that each teacher of record will have a minimum of 30 students participating in the pre-post category assessment. Grades 6 – 10 math and ELA courses will use the quarterly benchmarks.
- E. When will the assessment be administered: The pre-tests will be administered in the early fall and the post-tests will be administered in mid-spring. The quarterly benchmarks are administered at the end of each quarter. For the teacher evaluation, quarters 1 and 3 will be used. DIBELS is administered three times a year. The first test in the fall and the last test in the spring will be used.
- F. Who will score the assessment: The category assessments will be made available on-line through the district's assessment vendor's webpage and will be scored electronically. For schools lacking the technology infrastructure to test on-line, paper tests will be made available that can be scanned into the assessment company's data base. For grades K-2, teachers will score the DIBELS assessments. For grades 3 – 5, the quarterly math and ELA assessments will be available both on-line and with paper tests that can be scanned for electronic scoring.
- G. Scoring and point allocation: Students growth will be assessed by determining the difference between the pre-test and the post-test. Teachers will receive a 1 (below average growth or a total of 11 points), a 2 (average growth or an average of 22 points), or a 3 (above average growth or an average of 33 points) that will be added to the Teacher Evaluation points total. Cut scores will be determined once all teachers have administered pre and post tests and the scores can be evaluated.

In summary, in order for formally 'B' teachers now to be considered 'A' teachers, each subject needs a valid and reliable pre-post assessment that is specific to that subject for grades 6 - 12. Currently TUSD does not have consistent district-developed pre-post assessments for each subject. These assessments will be developed in the summer of 2015 to roll out for the 2015-16 school year. However, if the academic growth model is to be continued in future years, TUSD will be prepared to refine these assessments and the process in collaboration with teachers who specialize in each subject.

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 the a 4-point Likert scale:

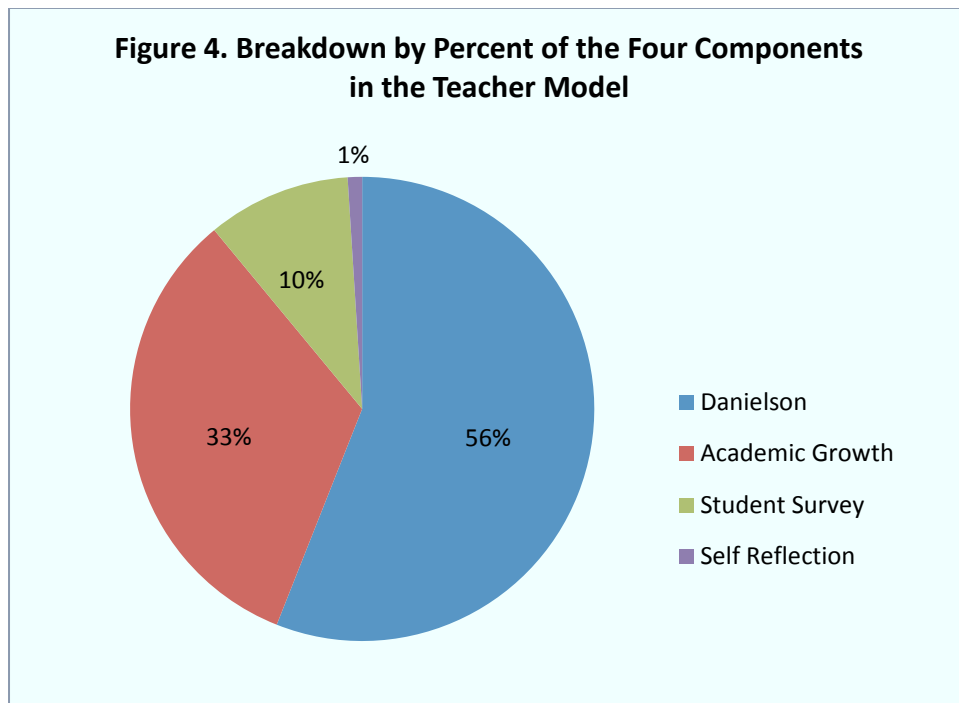
- 1 = Strongly Disagree
- 2 = Disagree
- 3 = Agree
- 4 = Strongly Agree

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.

Converting Raw Scores into Weighted Scores



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. The amount of impact from the academic growth is dependent upon how the cut scores are determined. Finally, the results of the Student Survey (10%) and the Self Reflection Survey (1%) each only will have a negligible impact on a teacher's overall score.

To get the ratio 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 ratio 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.

Table 1. Grades K-12 Distribution of Points			
Component	Current Max Raw Points	Desired Max Points	Scaling Factor*
Danielson	88	56	.636
Academic Growth	3	33	11
Student Survey	4	10	2.50
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 3 different Grade 4 teachers with three different raw points. Their points are converted using the Scaling Factor Conversion to give the weighted points.

Teacher A – Grade 4

Table 4. Grades 3-5 Calculation of Points of a Teacher Scoring Full Points			
Component	Max Raw Points	Scale Conversion	Weighted Points
Danielson	88	$88 \times .636$	56
Academic Growth	3	3×11	33
Student Survey	4	4×2.5	10
Teacher Self Reflection	1	1×1	1
Total	96		100

Teacher B – Grade 4

Table 5. Grades 3-5 Calculation of Points of a Teacher Scoring about Half of the Possible Points			
Component	Max Raw Points	Scale Conversion	Weighted Points
Danielson	44	$44 \times .636$	28
Academic Growth	1.5	1.5×11	16.5
Student Survey	2	2×2.5	5
Teacher Self Reflection	1	1×1	1
Total	48.5 or 49		50

Teacher C – Grade 4

Table 6. Grades 3-5 Calculation of Points of a Teacher Scoring about Average of the Possible Points			
Component	Max Raw Points	Scale Conversion	Weighted Points
Danielson	73	$73 \times .636$	46
Academic Growth	2	2×11	22
Student Survey	3.5	3.5×2.5	9
Teacher Self Reflection	1	1×1	1
Total	79.5 or 80		78

Cut Scores from 2013-14

The cut scores established for last year’s teacher evaluation were:

- Ineffective 0 – 39 total points
- Developing 40 – 55 total points
- Effective 56 – 73 total points
- Highly Effective 74 - 100 total points

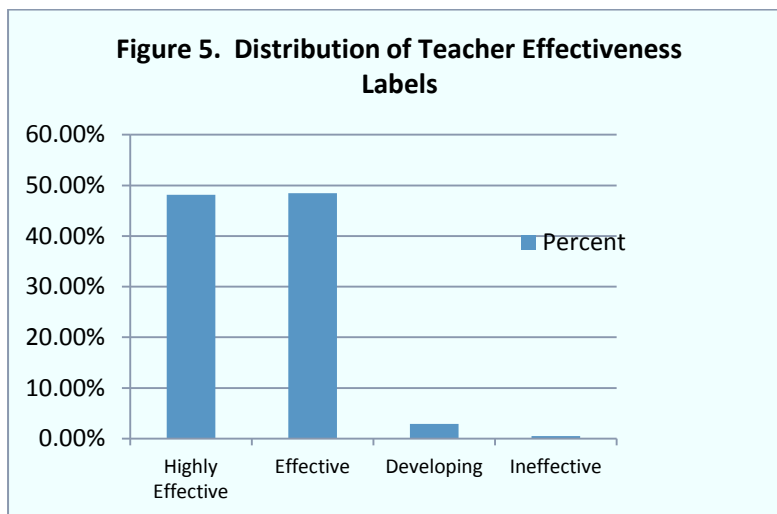
Based on last year’s cuts Teacher A above would be considered “Highly Effective”, Teacher B would be considered “Developing”, and Teacher C would also be considered “Highly Effective”.

To be considered “Ineffective”, a teacher would have to score very low on the Danielson Framework. The weighted percent of the Academic Growth, Student Survey, and the Teacher Self Reflection will have only a modest impact on the overall score. The only way a teacher can score ‘ineffective’ with the cut scores is to score about 32 points (out of a possible 88) on the Danielson observation. No teacher scored below 39 on the Danielson observation last year (2013-14).

Teacher D – Grade 4

Table 7. Grades 3-5 Calculation of Points of a Teacher Scoring Some of the Possible Points			
Component	Max Raw Points	Scale Conversion	Weighted Points
Danielson	32	32 x .636	20
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	36.75 or 37		39

An analysis was conducted of the distribution of the teacher effectiveness labels for 2013-14. The graph below reveals that the results were very skewed because the cut scores for effectiveness was low. It is recommended that new cuts are established to provide a more realistic distribution of teacher effectiveness.



Summary

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 2015-16, TUSD has chosen to use a simple model to evaluate teacher effectiveness. The majority of the points (56%) will derive from the Danielson observation that is conducted and scored by principals. The Danielson model calls for multiple observations over the course of the year and can be time intensive. The student growth piece has changed in design for next year and now stipulates that all teachers will be designated as 'A' teachers. Measuring student growth for each teacher for each subject, grades 6 – 12 will be developed in collaboration with teacher teams to be implemented in the fall of 2015. Also, 10% of the teacher evaluation is accounted for by the on-line student survey. This assessment will provide student feedback on the instructional qualities of their teachers. Finally, a reflection survey (1%) is to be filled out by teachers.