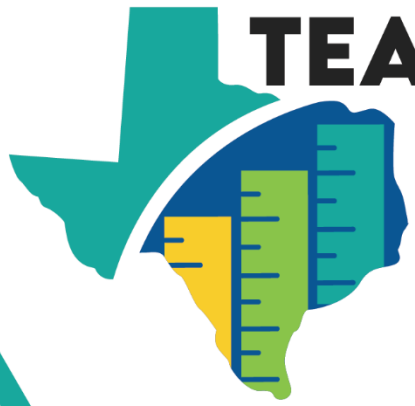

TEACHER INCENTIVE ALLOTMENT

Summary of the 2026-27 Data Validation Process

2026 Fall Data Submission

Table of Contents

Executive Summary

Overview

The Texas Education Agency (TEA) and Texas Tech University (TTU) thank participating districts for their commitment to the Teacher Incentive Allotment (TIA). This resource outlines the Data Validation process used to ensure district designation systems fairly and accurately identify high-performing teachers.

Purpose

This guide explains the structure and purpose of the Data Validation checks conducted by TTU. It is organized into three sections and helps districts prepare for data submission and understand how their designation system will be evaluated. Districts are encouraged to review this summary before engaging with the data submission template and refer back to it while preparing their data.

1. Reader-Friendly Description of Data Validation Checks
    - Describes the statistical procedures used in each check in accessible language
    - Explains what type of evidence each check provides in supporting a district's designation system

2. Validation Rubric
    - Outlines how evidence is evaluated and points are assigned
    - Determines if sufficient evidence exists to support the conclusion that a district's designation system is valid and reliable

3. Description of Statistical Analysis Protocols
    - Provides deeper detail on statistical procedures, test statistics, and decision rules for rubric scoring

Data Validation Updates

Annually, TTU and TEA review the data submission process – including data validation – for the purpose of continuous improvement. **This year, there are three updates to the data validation process**:

- An **Acknowledged** band has been added as the new lowest designation level.
    - Checks impacted by district designation have been updated to reflect this addition
- Check 14 has been added as an un-scored supplemental check in Domain E. The purpose of this check is to compare the correlation between teacher observation scores and student growth when disaggregated by appraisers.
- Check 15 has been added as an un-scored supplemental check in Domain E. The purpose of this check is to compare the correlation between teacher observation scores and student growth when disaggregated by eligible teacher categories.

Please use this guide as a reference when preparing and submitting data.

For questions about the summary or updates, contact TIA@ttu.edu.

Data Validation Report Table

Domain	*Check*	*Possible Points*	*Results*	*Score*	*Weight*	*Score × Weight*
A. Correlation between teacher observation ratings and student growth ratings	C1	0-3	r =		× 10	
B. Relationship between teacher designations and statewide VAM	C2	0-3	τ =		× 7	
	C3	0-3	s =		× 3	
C. Degree of reliability for observation and growth judgements	C4	0-3	sp. ω^2 =		× 3	
	C5	0-3	sp. ω^2 =		× 2	
	C6	0-3	sp. ω^2 =		× 3	
	C7	0-3	sp. ω^2 =		× 2	
D. Comparison of district designation percentage to statewide performance standards	C8	0-3	s = %		× 1	
	C9	0-3	s = %		× 1	
	C10	0-3	w =		× 3	
E. Supplemental checks	C11	0-3	σ =		× 0	–
	C12	0-3	ρ =		× 0	–
	C13	0-3	ρ =		× 0	–
	C14	0-3	I^2 =		× 0	–
	C15	0-3	I^2 =		× 0	–
					Total	/

Reader-Friendly Description of Data Validation Checks

The analyses described below are intended to validate the district designation system by comparing designations with external data and performing internal consistency checks. The purpose is to confirm that the district system functions in a manner that meets certain reliability (consistency) and validity (accuracy) standards, not to confirm or reject the designation of individual teachers. Meeting these standards allows stakeholders to have confidence that the designation system is fair and accurate.

Domain A. Correlation between teacher observation ratings and student performance ratings
Check 1 The correlation coefficient between observation and growth among all eligible teachers is within the range of expected magnitude reported in the research literature.
For this check, analysts calculate the correlation coefficient (Pearson product-moment correlation) between teacher observation scores and student growth scores submitted by the district. This analysis involves looking for a trend or pattern in the relation between teaching proficiency (i.e., observation scores) and the learning gain exhibited by students (i.e., student growth). Based upon findings reported in peer-reviewed research literature, the expectation is that the trend or relation will be at least minimally positive. For example, the analysts will expect to see that teachers who are assigned higher observation ratings by appraisers will also have students that exhibit greater growth. Conversely, teachers who are assigned lower observation ratings by an appraiser would be expected to have students who exhibited less growth. Results from this analysis provide one piece of evidence about the validity of the designation system.

Domain B. Confirm relation between district designations and student growth calculations
Check 2 District designations of Acknowledged, Recognized, Exemplary, and Master (AREM) teachers are found in similar proportion to designations as determined by the statewide VAM.
For this check, analysts calculate a rank correlation coefficient (Kendall's tau) between the designation category assigned to a teacher by the district and the equivalent category derived from state-level value-added scores calculated for teachers in the district. This analysis uses district data from SY 2025-26 restricted to the group of teachers whom the district has designated and for whom a state-level STAAR-based value-added score can be calculated. This analysis looks at the rank of the designation (AREM) and compares it to the rank derived from a value-added score. The expectation is that teachers assigned a Master designation would have a higher-ranking designation as determined by the statewide VAM than those with an Exemplary designation; that teachers assigned an Exemplary designation would have a higher-ranking designation as determined by the statewide VAM than those with a Recognized designation; and that teachers assigned a Recognized designation would have a higher-ranking designation as determined by the statewide VAM than those with an Acknowledged designation. Results from this analysis provide another piece of evidence about the validity of the designation system.
Check 3 District designation decisions for AREM teachers, in tested subjects, are in proximity to designations as determined by the statewide VAM.

For this check, analysts examine the accuracy with which local designation systems designate each eligible teacher in a tested subject based on calculations of the designations if they were determined by the statewide VAM. The table below shows how scores are calculated based on the proximity of district designation to the designation if it was determined by the statewide VAM. Scores reflect a positive value for accurate designation, and a negative value for decisions that are not aligned with designations if they had been determined by the statewide VAM. For example, if a local system designates a teacher as Exemplary, and this designation is consistent with the designation determined by the statewide VAM, then an accuracy score of 1.00 is assigned. On the other hand, if a local system designates a teacher as Master, but the designation as determined by the statewide VAM indicates no designation should be made, then an accuracy score of -1.00 is assigned. The expectation is that local systems will accurately identify teachers, and their levels, for designation. This analysis provides evidence about the concurrent validity of the local designation system.

	Designations if determined by the statewide VAM				
District Designations	*Not Designated*	*Acknowledged*	*Recognized*	*Exemplary*	*Master*
Acknowledged	0.00	**1.00**	0.75	0.50	0.50
Recognized	–0.25	0.75	**1.00**	0.75	0.50
Exemplary	–0.50	0.25	0.75	**1.00**	0.75
Master	–1.00	0.00	0.25	0.75	**1.00**

Domain C. Degree of reliability for observation and growth judgements

Check 4

Across campuses, observation scores are similar for teachers in AREM groups.

For this check, analysts use an analysis of variance (ANOVA) to calculate the extent to which there are similarities in observation scores for AREM teachers across campuses. The expectation is that there will be small, statistically non-significant differences between the same designation levels across campuses within the district. That is, observation scores for teachers designated at the Master level are expected to be comparable regardless of campus. Similar analyses are performed for the observation scores associated with teachers in the other designation groups. If the expected level of consistency is found in the observation data, it provides evidence about the reliability of the district's designation system.

Check 5

Across campuses, percentages of student growth are similar for teachers in AREM groups.

This check is like Check 4 in that ANOVA is used to calculate similarities for AREM teachers across campuses. In this case, however, analysts are interested in comparing student growth scores, or the percentage of students who meet or exceed learning expectations. As above, the expectation is that there will be small, statistically non-significant differences between the same designation levels across campuses within the district. That is, growth scores associated with teachers designated at the Master level are expected to be comparable regardless of campus; and similar analyses are performed for the growth scores associated with teachers in the other designation groups. If the expected level of consistency is found in student-growth data, it provides evidence about the reliability of the district's designation system.

Check 6

Across assignments, observation scores are similar for teachers in AREM groups.

This check is also like Check 4, but instead of making comparisons across campuses, it looks for similarities in observation ratings within AREM groups across teaching assignments. As before, ANOVA is used to calculate similarities among designation groups based on teaching assignments. Teaching assignments and the subsequent comparisons will be defined in one of two ways, based upon the data provided by the district.

- First, assignment may mean looking at similarities in observation scores across eligible teacher categories as identified in the district TIA application; or if districts identify only one category of eligibility, then
- Second, assignment may mean looking at similarities in observation scores across teachers in STAAR-tested vs. non STAAR-tested assignments (e.g., Grade 3 science vs. Grade 5 science).

If the expected level of consistency is found in observation data across assignments, it provides evidence about the reliability of the district's designation system.

Check 7

Across assignments, percentages of student growth are similar for teachers in AREM groups.

This check is like Check 5, but instead of making comparisons in observation, the comparison is of student growth (percentage of students who meet or exceed learning expectations) within the AREM groups across teaching assignments. As before, teaching assignment will be defined as eligible teacher categories or STAAR-tested vs. non-STAAR-tested, depending on the district system and the eligible teacher categories put forward for designation. This check is the last of four checks that are intended to provide evidence about the reliability of the district's designation system.

Domain D. Comparison of district designation percentage to statewide performance standards

Check 8

The percentage of students who meet or exceed expected growth in the district is approximately equal to the statewide performance standards for student growth in each of the teacher-designation levels (AREM).

Check 9

Observation ratings in the district are approximately equal to the statewide performance standards for teaching proficiency in each of the teacher-designation levels (AREM).

Both checks involve simple comparisons between statewide performance standards for each designation level and district-level results. Performance standards were calculated for both student growth and teacher observation ratings for the top 50% (Acknowledged level), top 33% (Recognized level), top 20% (Exemplary level) and top 5% (Master level).

Performance standards for student growth are set for each designation level (i.e., Acknowledged = 50%; Recognized = 55%; Exemplary = 60%; Master = 70%). The district's results for the percentage of students who meet or exceed growth are compared to the performance standards. District results that meet the designation performance standards from the state are considered to be a match, but those that fall below the state standard are considered to be a mismatch.

Performance standards for teacher observation are also set for each designation level based on the average number of points assigned by appraisers for Domain 2 and 3 of T-TESS (i.e., Acknowledged = 3.5 points; Recognized = 3.7 points; Exemplary = 3.9 points; and Master = 4.5 points). The district's results for appraiser ratings are compared to the performance standards. District point values that meet the performance standards are considered to be a match, but those that fall below are considered to be a mismatch. In cases where districts use an observation system other than T-TESS, a crosswalk between the rubrics is performed and equivalent levels are set (i.e., Acknowledged = 70% of possible points; Recognized = 74% of possible points; Exemplary = 78% of possible points; Master = 90% of possible points).

The scoring criteria for these analyses will be based on the number of designation groups with which district data matches the performance standard for growth and observation scores. A greater number of points will be awarded when there is a greater number of groups with which district designations match the levels described above. Results from these analyses provide evidence about the validity of the designation system.

Check 10

The proportion of teachers on district campuses who are designated as Acknowledged, Recognized, Exemplary, or Master is roughly equivalent to other campuses with the same Academic Growth (Domain 2A) rating.

The purpose of this check is to examine patterns in designation groups and compare them to district campuses with the same Academic Growth (Domain 2A) rating. The expectation is that a district's proportion of designated teachers across all campuses with a specific rating will be like other campuses across the state with the same rating. Results from analyses provide evidence about how well districts have calibrated their system to state standards as well as outcomes found among similarly rated peer districts across the state. Results from this

analysis also provide evidence about the validity of the designation system. For the 2026 Data Validation process, TEA will use 2025-26 Domain 2A ratings. Any district submitting a campus receiving a "Not Rated" label will be excluded from this analysis for comparison purposes.

The scoring criteria reflect the size of the difference between the proportion of teachers designated by the district, and the proportion of designated teachers found in statewide averages of districts with the same Domain 2A rating. Smaller differences in proportion (i.e., less than or equal to 0.10 difference) earn more points. Districts with proportions that differ from the statewide average by more than 0.70 receive "0" points on this check.

Note: if needed, TEA may use an alternative method to calculate Check 10.

Domain E. Supplemental system checks (not scored)

Check 11
The variability in observation ratings among all eligible teachers is within the range of historical magnitude.

The purpose of this check is to display the spread of teacher observation scores. For this check, analysts calculate standard deviation of (max-scaled) observation scores among all eligible teachers.

The expectation is that observation scores for the district will be distributed in a manner that gives some evidence about the ability of the local system to differentiate between ineffective and effective instruction.

Check 12
The ranking of teachers based on observation scores closely aligns with their ranking on statewide performance standards for teaching proficiency.

Check 13
The ranking of teachers based on percentages of student growth closely aligns with their ranking on statewide performance standards for teaching proficiency.

For these checks, analysts calculate a rank correlation coefficient (Spearman's ρ) between observation scores (Check 12) or student growth scores (Check 13) and state-level value-added scores calculated for teachers in the district. This analysis uses district data from SY 2025-26 restricted to the group of teachers for whom a state-level STAAR-based value-added score can be calculated. This analysis looks at the rank of the observation or growth score and compares it to the rank derived from the statewide value-added score. The expectation is that teachers with a higher-ranking on the observation or growth score would have a higher-ranking as determined by the statewide VAM. Results from this analysis provide another piece of evidence about the validity of the designation system.

Check 14
The correlation coefficient between observation and growth is consistent across appraisers among all eligible teachers.

Check 15
The correlation coefficient between observation and growth is consistent across eligible teacher categories among all eligible teachers.

For these checks, analysts examine whether the relationship between teacher observation scores and student academic growth is consistent across two different groupings: by appraiser (Check 14) and by eligible teacher category (Check 15). A correlation coefficient is calculated separately for each appraiser and each teacher category. The expectation is that this relationship should be roughly the same regardless of who conducted the evaluation or what type of teacher is being assessed. In other words, the connection between a teacher's observation score and their students' academic growth should not depend on which appraiser evaluated the teacher or which category the teacher falls into. If large differences are found across appraisers, it may suggest that observation standards are being applied inconsistently. Similarly, large differences across teacher categories may indicate that observation and growth measures do not capture teaching effectiveness in the same way for all types of teachers. Consistent results across both groupings provide evidence that the district's teacher evaluation system operates in a fair, uniform, and reliable manner.

Validation Rubric

None or almost no evidence supports judgements	Limited evidence supports the accuracy of judgements	Some evidence supports the accuracy of judgements	Most evidence supports the accuracy of judgements
Score of 0	**Score of 1**	**Score of 2**	**Score of 3**

Domain A. Correlation between teacher observation ratings and student performance ratings

This check is intended to confirm that teachers' appraisal scores are related to student growth scores.

1. The correlation coefficient between observation and growth among all *eligible* teachers is within the range of expected magnitude reported in research literature.

Earned points x 10 = weighted score for this check

0 points	0-1 points	1 point	1-2 points	2 points	2-3 points	3 points
$r \leq 0$	$\frac{r - 0}{0.08 - 0}$	$r = 0.08$	$\frac{r - 0.08}{0.16 - 0.08} + 1$	$r = 0.16$	$\frac{r - 0.16}{0.24 - 0.16} + 2$	$r \geq 0.24$

Domain B. Confirm the relation between district designations and statewide VAM

These checks are intended to confirm that district designations are aligned with state-level student-growth calculations. For the current year, this analysis compares district designations to SY 2025-26 statewide VAM data.

2. District designations of Acknowledged, Recognized, Exemplary, and Master (AREM) teachers are found in similar proportion to designations as determined by the statewide VAM.

Earned points x 7 = weighted score for this check

0 points	0-1 points	1 point	1-2 points	2 points	2-3 points	3 points
$\tau \leq 0$	$\frac{\tau - 0}{0.10 - 0}$	$\tau = 0.10$	$\frac{\tau - 0.10}{0.30 - 0.10} + 1$	$\tau = 0.30$	$\frac{\tau - 0.30}{0.50 - 0.30} + 2$	$\tau \geq 0.50$

3. District designations for AREM teachers, in tested subjects, are in proximity to designations as determined by the statewide VAM.

Earned points x 3 = weighted score for this check

0 points	0-1 points	1 point	1-2 points	2 points	2-3 points	3 points
$s \leq 0$	$\frac{s - 0}{0.10 - 0}$	$s = 0.10$	$\frac{s - 0.10}{0.30 - 0.10} + 1$	$s = 0.30$	$\frac{s - 0.30}{0.70 - 0.30} + 2$	$s \geq 0.70$

Domain C. Degree of reliability for observation and growth judgements

These checks are intended to confirm that observation ratings and student performance are determined in a consistent manner across campuses and teaching assignments. [1]

4. Across campuses, observation scores are similar for teachers in AREM groups.
Earned points x 3 = weighted score for this check

0 points	0-1 points	1 point	1-2 points	2 points	2-3 points	3 points
$sp.\omega^2$ ≥ 0.25	$\frac{0.25 - sp.\omega^2}{0.25 - 0.14}$	$sp.\omega^2$ $= 0.14$	$\frac{0.14 - sp.\omega^2}{0.14 - 0.06} + 1$	$sp.\omega^2$ $= 0.06$	$\frac{0.06 - sp.\omega^2}{0.06 - 0.01} + 2$	$sp.\omega^2$ ≤ 0.01

5. Across campuses, percentages of student growth are similar for teachers in AREM groups.
Earned points x 2 = weighted score for this check

0 points	0-1 points	1 point	1-2 points	2 points	2-3 points	3 points
$sp.\omega^2$ ≥ 0.25	$\frac{0.25 - sp.\omega^2}{0.25 - 0.14}$	$sp.\omega^2$ $= 0.14$	$\frac{0.14 - sp.\omega^2}{0.14 - 0.06} + 1$	$sp.\omega^2$ $= 0.06$	$\frac{0.06 - sp.\omega^2}{0.06 - 0.01} + 2$	$sp.\omega^2$ ≤ 0.01

6. Across assignments, observation scores are similar for teachers in AREM groups.
Earned points x 3 = weighted score for this check

0 points	0-1 points	1 point	1-2 points	2 points	2-3 points	3 points
$sp.\omega^2$ ≥ 0.25	$\frac{0.25 - sp.\omega^2}{0.25 - 0.14}$	$sp.\omega^2$ $= 0.14$	$\frac{0.14 - sp.\omega^2}{0.14 - 0.06} + 1$	$sp.\omega^2$ $= 0.06$	$\frac{0.06 - sp.\omega^2}{0.06 - 0.01} + 2$	$sp.\omega^2$ ≤ 0.01

7. Across assignments, percentages of student growth are similar for teachers in AREM groups.
Earned points x 2 = weighted score for this check

0 points	0-1 points	1 point	1-2 points	2 points	2-3 points	3 points
$sp.\omega^2$ ≥ 0.25	$\frac{0.25 - sp.\omega^2}{0.25 - 0.14}$	$sp.\omega^2$ $= 0.14$	$\frac{0.14 - sp.\omega^2}{0.14 - 0.06} + 1$	$sp.\omega^2$ $= 0.06$	$\frac{0.06 - sp.\omega^2}{0.06 - 0.01} + 2$	$sp.\omega^2$ ≤ 0.01

[1] Observation and growth should be equal when compared across campuses and assignments. A smaller effect-size indicates small differences, thus a greater level of agreement. A larger effect-size indicates larger differences, thus a smaller level of agreement.

Domain D. Comparison of district designation percentage to statewide performance standards

These checks are intended to confirm that designation rates in each district are aligned with statewide projections of the proportion of designated teachers in each district.

8. *The percentage of students who meet or exceed expected growth* in the district is approximately equal to the statewide performance standards for student growth in each of the teacher-designation levels (AREM).

Earned points x 1 = weighted score for this check

0 points	**0-1 points**	**1 point**	**1-2 points**	**2 points**	**2-3 points**	**3 points**
$s \leq 0.55$	$\frac{s - 0.55}{0.60 - 0.55}$	$s = 0.60$	$\frac{s - 0.60}{0.65 - 0.60} + 1$	$s = 0.65$	$\frac{s - 0.65}{0.70 - 0.65} + 2$	$s \geq 0.70$

9. *Observation ratings* in the district are approximately equal to the statewide performance standards for teaching proficiency in each of the AREM levels.

Earned points x 1 = weighted score for this check

0 points	**0-1 points**	**1 point**	**1-2 points**	**2 points**	**2-3 points**	**3 points**
$s \leq 0.50$	$\frac{s - 0.50}{0.60 - 0.50}$	$s = 0.60$	$\frac{s - 0.60}{0.70 - 0.60} + 1$	$s = 0.70$	$\frac{s - 0.70}{0.80 - 0.70} + 2$	$s \geq 0.80$

10. The proportion of teachers on district campuses who are designated as *Acknowledged*, *Recognized*, *Exemplary*, or *Master* is roughly equivalent to other campuses with the same Academic Growth (Domain 2A) rating.[2]

Earned points x 3 = weighted score for this check

0 points	**0-1 points**	**1 point**	**1-2 points**	**2 points**	**2-3 points**	**3 points**
$w \geq 0.70$	$\frac{0.70 - w}{0.70 - 0.50}$	$w = 0.50$	$\frac{0.50 - w}{0.50 - 0.30} + 1$	$w = 0.30$	$\frac{0.30 - w}{0.30 - 0.10} + 2$	$w \leq 0.10$

[2] Note: if needed, TEA may use an alternative method to calculate Check 10.

Domain E. Supplemental Checks

These checks are intended to provide additional, non-scored evidence to districts about the validity of their local designation system. Check 11 shows the variance in a district's teacher observation scores as an indicator of the extent to which observers differentiate between more effective and less effective instruction. Checks 12 and 13 indicate the level of agreement between the rankings of teachers within the district on observation/growth and statewide VAM scores. Checks 14 and 15 indicate the level of correlation consistency across appraisers and eligible teacher categories. For the current year, these checks are supplemental and are not factored into data validation scores or system validation decisions.

11. The variability in observation ratings among all eligible teachers is within the range of expected magnitude.
No points assigned for supplemental check

0 points	**0-1 points**	**1 point**	**1-2 points**	**2 points**	**2-3 points**	**3 points**
$\sigma \leq 0.06$	$\frac{\sigma - 0.06}{0.08 - 0.06}$	$\sigma = 0.08$	$\frac{\sigma - 0.08}{0.10 - 0.08} + 1$	$\sigma = 0.10$	$\frac{\sigma - 0.10}{0.12 - 0.10} + 2$	$\sigma \geq 0.12$

12. The ranking of teachers based on observation scores closely aligns with their ranking on statewide performance standards for teaching proficiency.
No points assigned for supplemental check

0 points	**0-1 points**	**1 point**	**1-2 points**	**2 points**	**2-3 points**	**3 points**
$\rho \leq 0$	$\frac{\rho - 0}{0.10 - 0}$	$\rho = 0.10$	$\frac{\rho - 0.10}{0.37 - 0.10} + 1$	$\rho = 0.37$	$\frac{\rho - 0.37}{0.64 - 0.37} + 2$	$\rho \geq 0.64$

13. The ranking of teachers based on percentages of student growth closely aligns with their ranking on statewide performance standards for teaching proficiency.
No points assigned for supplemental check

0 points	**0-1 points**	**1 point**	**1-2 points**	**2 points**	**2-3 points**	**3 points**
$\rho \leq 0$	$\frac{\rho - 0}{0.10 - 0}$	$\rho = 0.10$	$\frac{\rho - 0.10}{0.25 - 0.10} + 1$	$\rho = 0.25$	$\frac{\rho - 0.25}{0.50 - 0.25} + 2$	$\rho \geq 0.50$

14. The correlation coefficient between observation and growth is consistent across appraisers among all eligible teachers.
No points assigned for supplemental check

0 points	**0-1 points**	**1 point**	**1-2 points**	**2 points**	**2-3 points**	**3 points**
$I^2 \geq 0.75$	$\frac{0.75 - I^2}{0.75 - 0.50}$	$I^2 = 0.50$	$\frac{0.50 - I^2}{0.50 - 0.25} + 1$	$I^2 = 0.25$	$\frac{0.25 - I^2}{0.25 - 0.10} + 2$	$I^2 \leq 0.10$

15. The correlation coefficient between observation and growth is consistent across eligible teacher categories among all eligible teachers.
No points assigned for supplemental check

0 points	**0-1 points**	**1 point**	**1-2 points**	**2 points**	**2-3 points**	**3 points**
$I^2 \geq 0.75$	$\frac{0.75 - I^2}{0.75 - 0.50}$	$I^2 = 0.50$	$\frac{0.50 - I^2}{0.50 - 0.25} + 1$	$I^2 = 0.25$	$\frac{0.25 - I^2}{0.25 - 0.10} + 2$	$I^2 \leq 0.10$

Description of Statistical Analysis Protocols

Check 1. The correlation coefficient between observation and growth among all eligible teachers is within the range of expected magnitude reported in research literature.

Pearson product-moment correlation coefficient (r) is calculated between the teacher observation and growth scores of all eligible teachers. Pearson's coefficient is a measure of the strength and direction of linear association between two variables, which can be written as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}},$$

where n is the sample size; x_i and y_i are the person i's values on x and y (e.g., x = observation score, y = growth score); and $\bar{x}$ and $\bar{y}$ are the sample means of x and y.

Correlation coefficient has a value between –1 (a perfect negative correlation) and +1 (a perfect positive correlation). A positive correlation indicates a positive relationship while a negative correlation signifies a negative relationship. For example, when teachers with higher observation scores show higher growth scores, the correlation will be positive; in contrast, when teachers with higher observation scores show lower growth scores, the correlation will be negative. Two correlations with the same numerical value have the same strength whether the correlation is positive or negative. A zero correlation indicates no relationship between the variables. The following guidelines are useful when determining the strength of a correlation: ±0.1 (small), ±0.3 (moderate), and ±0.5 (large) (Cohen, 1988, 1992).

Check 2. District designations of Acknowledged, Recognized, Exemplary, and Master (AREM) teachers are found in similar proportion to designations as determined by the statewide VAM.

Kendall's rank correlation coefficient (τ) is calculated between the designation level that the district has made for their teachers (Master, Exemplary, Recognized, or Acknowledged) and the same teachers' designation level that is determined by their statewide value-added (VAM) score (Master, Exemplary, Recognized, Acknowledged, or Not Designated). Kendall's coefficient is a measure of the strength and direction of ordinal association between two variables, which can be written as:

$$\tau_{xy} = \frac{n_c-n_d}{\sqrt{(n_0-n_1)(n_0-n_2)}},$$

where n is the sample size; $n_0 = \frac{n(n-1)}{2}$; $n_1 = \sum_i \frac{t_i(t_i-1)}{2}$; $n_2 = \sum_j \frac{u_j(u_j-1)}{2}$; n_c is the number of concordant pairs; n_d is the number of discordant pairs; t_i is the number of tied values in the i^{th} group of ties for the first quantity; and u_j is the number of tied values in the j^{th} group of ties for the second quantity. Any pair of observations (x_i, y_i) and (x_j, y_j), where $i < j$, are said to be concordant if the sort of (x_i, y_i) and (x_j, y_j) agrees—that is, if either both $x_i > x_j$ and $y_i > y_j$ holds or both $x_i < x_j$ and $y_i < y_j$. Otherwise, they are said to be discordant.

For example, the correlation will be +1 (a perfect positive correlation) when the agreement between the district's designation and designations if determined by the statewide VAM is perfect (i.e., the two rankings are the same). The correlation will be positive when the two designations are similar. The correlation will be −1 (a perfect negative correlation) when the disagreement between the two designations is perfect (i.e., one ranking is the reverse of the other). When the two designations are independent, then the correlation will be approximately zero.

Check 3. District designation decisions for AREM teachers, in tested subjects, are in proximity to designations as determined by the statewide VAM.

For teachers of tested subjects who earned a designation in the district (Master, Exemplary, Recognized, or Acknowledged), it is determined whether the district designation is in the same, higher, or lower than the designation if it were determined by the statewide VAM. An "accuracy" score ranging from –1.00 to +1.00 is

assigned based on the proximity between the district designation and the designation if it were determined by the statewide VAM. The table below shows how values are assigned based on proximity:

	Designations if determined by the statewide VAM				
District designations	*Not Designated*	*Acknowledged*	*Recognized*	*Exemplary*	*Master*
Acknowledged	0.00	**1.00**	0.75	0.50	0.50
Recognized	–0.25	0.75	**1.00**	0.75	0.50
Exemplary	–0.50	0.25	0.75	**1.00**	0.75
Master	–1.00	0.00	0.25	0.75	**1.00**

More points are given when the district designation is closer to the designations if determined by the statewide VAM model. After a score has been assigned to each teacher, these scores are averaged to produce an overall score for the district.

Check 4. Across campuses, observation scores are similar for teachers in AREM groups.

Check 5. Across campuses, percentages of student growth are similar for teachers in AREM groups.

Analysis of variance (ANOVA) is performed to compare teachers' observation score (Check 4) or growth score (Check 5) across different campuses. The analysis model includes the main effects of campus and teacher designation (Master, Exemplary, Recognized, or Acknowledged) as well as their interaction effect. Then, semi-partial omega-squared (ω^2) for the campus effect is calculated. Semi-partial omega-squared is a measure of standardized group difference (effect size)—the proportion of the variance in a dependent variable (e.g., observation or growth score) that is accounted for by the independent variable (e.g., campus), with other effects (terms) in the model parsed out of the independent variable. It can be written as:

$$\text{semi-partial } \omega^2 = \frac{df_{\text{effect}}(MS_{\text{effect}} - MS_{\text{error}})}{df_{\text{effect}} MS_{\text{effect}} + (N - df_{\text{effect}}) MS_{\text{error}}},$$

where N is the sample size; df is the degrees of freedom; MS_{effect} is the mean sum of squares for the independent variable; and MS_{error} is the mean sum of squares for the error. (Semi-partial) omega-squared is widely viewed as a lesser biased alternative to (semi-partial) eta-squared, especially when sample sizes are small.

Semi-partial omega-squared can have a value between –1 and +1. The following guidelines are useful when determining the strength of a semi-partial omega-squared: 0.01 (small), 0.06 (moderate), and 0.14 (large) (Cohen, 1988, 1992). A zero or negative value indicates no effect of the independent variable when controlling for the other effects included in the model.

Check 6. Across assignments, observation scores are similar for teachers in AREM groups.

Check 7. Across assignments, percentages of student growth are similar for teachers in AREM groups.

ANOVA is performed to compare teachers' observation score (Check 6) or growth score (Check 7) across different teaching assignments. Teaching assignment is defined as two or more eligible teacher groups; or defined as tested subjects, non-tested subjects, or both subjects when there is only one eligible teacher group. The analysis model includes the main effects of teaching assignment and teacher designation (Master, Exemplary, Recognized, or Acknowledged) as well as their interaction effect. Then, semi-partial omega-squared (ω^2) for the teaching assignment effect is calculated.

Check 8. The percentage of students who meet or exceed expected growth in the district is approximately equal to the statewide performance standards for student growth in each of the teacher-designation levels (AREM).

Check 9. Observation ratings in the district are approximately equal to the statewide performance standards for teaching proficiency in each of the AREM levels.

For teachers who earned a designation in the district (Master, Exemplary, Recognized, or Acknowledged), it is determined how close their growth score (Check 8) or observation score (Check 9) is to the published cut-point that corresponds to their designation category. A closeness score based on the proximity of the growth score or observation score to the corresponding performance standard at each designation level is established on a 0-100% scale. The score value is calculated using an exponential equation that assigns a score based on the proximity of each teacher's score to the corresponding performance standard. More points are given when the score is closer to the performance standard. After a score has been assigned to each teacher, these scores are averaged. The state published cut-points used are shown below:

Growth standard group	**% of students meeting or exceeding growth targets**
Acknowledged	50%
Recognized	55%
Exemplary	60%
Master	70%

Observation standard group	**Based on T-TESS**	**Based on another rubric**
Acknowledged	3.5	70% of points
Recognized	3.7	74% of points
Exemplary	3.9	78% of points
Master	4.5	90% of points

The exponential equations used are shown below:

In Check 8
For Master teachers,

$$s_i = f(x_i) + g(x_i)\left(\frac{x_i-0.5}{0.7-0.5}\right)^2,$$

$$f(x_i) = \begin{cases} 1 & 0.7 \le x_i \\ 0 & \text{otherwise} \end{cases}, g(x_i) = \begin{cases} 1 & 0.5 \le x_i < 0.7 \\ 0 & \text{otherwise} \end{cases};$$

For Exemplary teachers,

$$s_i = f(x_i)\left(\frac{x_i-0.5}{0.6-0.5}\right)^2 + g(x_i) + h(x_i)\left(1-\frac{x_i-0.7}{1-0.7}\right)^2,$$

$$f(x_i) = \begin{cases} 1 & 0.5 \le x_i < 0.6 \\ 0 & \text{otherwise} \end{cases}, g(x_i) = \begin{cases} 1 & 0.6 \le x_i < 0.7 \\ 0 & \text{otherwise} \end{cases}, h(x_i) = \begin{cases} 1 & 0.7 \le x_i \\ 0 & \text{otherwise} \end{cases};$$

For Recognized teachers,

$$s_i = f(x_i)\left(\frac{x_i-0.5}{0.55-0.5}\right)^2 + g(x_i) + h(x_i)\left(1-\frac{x_i-0.6}{1-0.6}\right)^2,$$

$$f(x_i) = \begin{cases} 1 & 0.5 \le x_i < 0.55 \\ 0 & \text{otherwise} \end{cases}, g(x_i) = \begin{cases} 1 & 0.55 \le x_i < 0.6 \\ 0 & \text{otherwise} \end{cases}, h(x_i) = \begin{cases} 1 & 0.6 \le x_i \\ 0 & \text{otherwise} \end{cases};$$

For Acknowledged teachers,

$$s_i = f(x_i)\left(\frac{x_i-0.45}{0.5-0.45}\right)^2 + g(x_i) + h(x_i)\left(1-\frac{x_i-0.55}{1-0.55}\right)^2,$$

$$f(x_i) = \begin{cases} 1 & 0.45 \le x_i < 0.5 \\ 0 & \text{otherwise} \end{cases}, g(x_i) = \begin{cases} 1 & 0.5 \le x_i < 0.55 \\ 0 & \text{otherwise} \end{cases}, h(x_i) = \begin{cases} 1 & 0.55 \le x_i \\ 0 & \text{otherwise} \end{cases};$$

where s_i and x_i are the person *i*'s values on closeness score and growth score, respectively.

In Check 9
For Master teachers,

$$s_i = f(x_i) + g(x_i)\left(\frac{x_i - 0.7}{0.9 - 0.7}\right)^2,$$

$$f(x_i) = \begin{cases} 1 & 0.9 \le x_i \\ 0 & \text{otherwise} \end{cases}, g(x_i) = \begin{cases} 1 & 0.7 \le x_i < 0.9 \\ 0 & \text{otherwise} \end{cases};$$

For Exemplary teachers,

$$s_i = f(x_i)\left(\frac{x_i - 0.7}{0.78 - 0.7}\right)^2 + g(x_i) + h(x_i)\left(1 - \frac{x_i - 0.9}{1 - 0.9}\right)^2,$$

$$f(x_i) = \begin{cases} 1 & 0.7 \le x_i < 0.78 \\ 0 & \text{otherwise} \end{cases}, g(x_i) = \begin{cases} 1 & 0.78 \le x_i < 0.9 \\ 0 & \text{otherwise} \end{cases}, h(x_i) = \begin{cases} 1 & 0.9 \le x_i \\ 0 & \text{otherwise} \end{cases};$$

For Recognized teachers,

$$s_i = f(x_i)\left(\frac{x_i - 0.7}{0.74 - 0.7}\right)^2 + g(x_i) + h(x_i)\left(1 - \frac{x_i - 0.78}{1 - 0.78}\right)^2,$$

$$f(x_i) = \begin{cases} 1 & 0.7 \le x_i < 0.74 \\ 0 & \text{otherwise} \end{cases}, g(x_i) = \begin{cases} 1 & 0.74 \le x_i < 0.78 \\ 0 & \text{otherwise} \end{cases}, h(x_i) = \begin{cases} 1 & 0.78 \le x_i \\ 0 & \text{otherwise} \end{cases};$$

For Acknowledged teachers,

$$s_i = f(x_i)\left(\frac{x_i - 0.66}{0.7 - 0.66}\right)^2 + g(x_i) + h(x_i)\left(1 - \frac{x_i - 0.74}{1 - 0.74}\right)^2,$$

$$f(x_i) = \begin{cases} 1 & 0.66 \le x_i < 0.7 \\ 0 & \text{otherwise} \end{cases}, g(x_i) = \begin{cases} 1 & 0.7 \le x_i < 0.74 \\ 0 & \text{otherwise} \end{cases}, h(x_i) = \begin{cases} 1 & 0.74 \le x_i \\ 0 & \text{otherwise} \end{cases};$$

where s_i and x_i are the person *i*'s values on closeness score and observation score, respectively.

<u>Check 10</u>. The proportion of teachers on district campuses who are designated as Acknowledged, Recognized, Exemplary, or Master is roughly equivalent to other campuses in the same Academic Growth (Domain 2A) rating.

The campus cumulative percentage of (1) Master designation, (2) Exemplary or higher designations, (3) Recognized or higher designations, or (4) Acknowledged or higher designations are each compared to a State average of campuses within each of the Domain 2A categories. In other words, the district and state percentages are obtained for teachers within Domain 2A A-rated campuses, Domain 2A B-rated campus, etc. Then, Cohen's *w* is calculated from each possible comparison in the Domain 2A categories, and a mean value is calculated over the (1), (2), (3), and (4) designation levels.

Cohen's *w* is a measure of association between two nominal variables. With a binary outcome (e.g., designated vs. not designated), it can be written as follows with directionality considered:

$$w = \text{sign}(p_1 - p_0)\sqrt{\frac{(p_1 - p_0)^2}{p_0} + \frac{(p_1 - p_0)^2}{(1 - p_0)}},$$

where p_1 is the district percentage and p_0 is the statewide expected percentage. The value will be 0 when the district percentage is equal to the statewide percentage for a Domain 2A category. In contrast, the value will be positive

when the district percentage is larger than the statewide percentage; or it will be set to zero when the district percentage is smaller than the statewide percentage.

Note: if needed, TEA may use an alternative method to calculate check 10.

<u>Check 11.</u> The variability in observation ratings among all eligible teachers is within the range of expected magnitude.

Standard deviation (σ) is calculated for the (max-scaled) observation score of all eligible teachers. Standard deviation is a measure of variation or dispersion of a variable, which can be written as:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}},$$

where n is the sample size; x_i is the person i's values on x (e.g., observation score); and $\bar{x}$ is the sample mean of x. A low standard deviation indicates that teachers' observation scores are close to each other and to the mean, while a high standard deviation indicates that scores are spread out over a wider range.

<u>Check 12</u>. The ranking of teachers based on observation scores closely aligns with their ranking on statewide performance standards for teaching proficiency.

<u>Check 13</u>. The ranking of teachers based on percentages of student growth closely aligns with their ranking on statewide performance standards for teaching proficiency.

Spearman rank correlation coefficient (ρ) is calculated between observation scores (Check 12) or growth scores (Check 13) and statewide VAM scores among teachers of tested subjects. Spearman's coefficient is a measure of the strength and direction of monotonic association between the rankings on two variables, which can be written as:

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(Rx_i - R\bar{x})(Ry_i - R\bar{y})}{\sqrt{\sum_{i=1}^{n}(Rx_i - R\bar{x})^2 \sum_{i=1}^{n}(Ry_i - R\bar{y})^2}},$$

where n is the sample size; Rx_i and Ry_i are the person i's ranks on x and y (e.g., x = observation or growth score, y = VAM score); and $R\bar{x}$ and $R\bar{y}$ are the sample means of Rx and Ry.

For example, the correlation will be +1 (a perfect positive correlation) when there is a perfect agreement between the rankings of teachers within the district on the observation and statewide VAM scores (i.e., the two rankings are identical). The correlation will be positive when the two rankings are similar. The correlation will be −1 (a perfect negative correlation) when there is a perfect disagreement between the two rankings (i.e., one ranking is the exact opposite of the other). When the two rankings are independent, then the correlation will be close to zero.

<u>Check 14</u>. The correlation coefficient between observation and growth is consistent across appraisers among all eligible teachers.

<u>Check 15</u>. The correlation coefficient between observation and growth is consistent across eligible teacher categories among all eligible teachers.

Pearson product-moment correlation coefficient (r) is computed between the teacher observation and growth scores for all eligible teachers, calculated separately by teacher group—appraiser (Check 14) or eligible teacher category (Check 15). Cochran's Q test is then applied to determine whether the observed variability in correlations across teacher groups exceeds what would be expected due to sampling error alone. Cochran's Q can be written as:

$$Q = \sum_{i=1}^{k} w_i(z_i - \bar{z}_w)^2,$$

where k is the number of teacher groups; z_i is the Fisher's z-transformed correlation for group i; $\bar{z}_w$ is the weighted mean of z_i across all groups; and w_i is the weight assigned to group i, defined as the inverse of the sampling variance ($w_i = n_i - 3$, where n_i is the number of teachers in group i). Subsequently, I^2 is computed as a measure of heterogeneity (effect size)—the proportion of total variance in group-specific correlations attributable to true between-group differences rather than sampling error. I^2 can be written as:

$$I^2 = \frac{Q-df}{Q}, \text{ where}$$

Q is Cochran's Q statistic and $df = k - 1$. I^2 ranges from 0 to 1 (0-100%). Higher I^2 values indicate greater heterogeneity, suggesting that, for example, teachers across eligible teacher categories differ systematically in the relationship between their observation scores and student growth outcomes. Conversely, lower I^2 values indicate greater consistency, implying that teaching effectiveness is assessed in a more uniform and reliable manner across teacher groups. The following thresholds are commonly used to characterize the magnitude of heterogeneity: 0.25 (low), 0.5 (moderate), and 0.75 (high) (Higgins et al., 2003).