TBASE STAAR/SST Scoring Validation Report

2025, Grades 3 to 5

Extended constructed response (ECR) scoring compared with official TEA score points

Copyright and Source Notice

Copyright 2026 Teach BIG. All rights reserved.

TBASE STAAR/SST Scoring Validation Report, 2025, Grades 3 to 5. No part of this report may be reproduced or distributed without written permission from Teach BIG.

STAAR is a registered trademark of the Texas Education Agency. This report is not affiliated with, endorsed by, or sponsored by the Texas Education Agency (TEA). Student sample responses and official score points are quoted from the TEA publications listed below, which are the property of TEA; they are used here only to measure the agreement of an automated scoring tool with the published score points. Student responses are not reproduced in this report.

TEA documents used, titles as published: Grade 3 Reading Language Arts Constructed-Response Scoring Guide (2025); Grade 4 Reading Language Arts Constructed-Response Scoring Guide (2025); Grade 5 Reading Language Arts Constructed-Response Scoring Guide (2025). Source URLs are listed in sources.csv with this report.

TBASE is a Teach BIG product. The results describe one scoring configuration on 39 published sample papers and should not be read as a claim about performance on other papers.

Executive Summary

Exact	Adjacent (1 point off)	Exact plus adjacent
46%	**44%**	**90%**
18 of 39 papers	17 of 39 papers	35 of 39 papers

Teach BIG compared the live TBASE scorer at tbase.teachbig.com with the official TEA score points for the 2025 released Grade 3, 4, and 5 STAAR/SST extended constructed response (ECR) sample papers. TEA scores each ECR from 0 to 5. TBASE reports a STAAR/SST Score from 0 to 5, the sum of Organization and Development of Ideas (0 to 3) and Conventions (0 to 2), the same scale TEA uses. All 39 papers received a TBASE score.

Findings

- Across Grades 3 to 5, TBASE matched the TEA score exactly on 46% of papers (18 of 39), was within one point on 90% (35 of 39), and was 2 points away on 10% (4 papers, listed on each grade page).
- By grade, agreement within one point was 85% in Grade 3, 92% in Grade 4, and 92% in Grade 5. Quadratic weighted kappa was 0.68, 0.87, and 0.86 (0.80 overall).
- On average TBASE is within a tenth of a point of the state in every year, so it does not lean high or low. In this year the mean difference, TBASE minus TEA, was +0.08 points.
- 8 papers received a zero because they did not answer the prompt in the first five sentences. TEA scored 5 of these 0, and gave 2025-G3-ECR-SP1-R1 a 1, 2025-G3-ECR-SP4-R1 a 4, 2025-G5-ECR-SP1-R1 a 1.
- Leakage check: TBASE scores against TEA example papers. For this test every 2025 paper (and every 2026 paper) was removed from those examples, so no paper in this test was used to score itself.

Exact and Adjacent Scores

An exact score means TBASE gave a paper the same score the state gave it. An adjacent score means the two scores are one point apart, higher or lower. Exact plus adjacent, or within one point, is the standard way to measure scoring agreement.

The STAAR/SST scoring system works the same way. Each response is scored two times. The two scores do not have to match exactly: as long as they are the same or adjacent, the two scores are added together. This is why some students receive an odd number score. The state does not simply double one score; two separate scorings can land one point apart, and both count. An adjacent score is part of how the state itself scores papers.

Method

Source

Official TEA 2025 Grade 3, 4, and 5 Reading Language Arts Constructed-Response Scoring Guides (tea.texas.gov), extended constructed response section. Each guide has one ECR prompt with its passage and 13 sample student responses at TEA score points 0 to 5 (39 papers in all). The TEA total score is the sum of Organization and Development of Ideas (0 to 3) and Conventions (0 to 2). Student text was taken from the guides without editing.

Scoring path

Each paper was sent to the live TBASE site through a staff only validation endpoint that runs the same scoring code teachers use, including the first five sentences check, and writes nothing to the database. The TBASE STAAR/SST Score is the number compared with the TEA total. Scoring date: October 17, 2026.

Leakage handling

TBASE uses TEA scored papers from earlier released guides as scoring examples. For this test, every 2025 paper and every 2026 paper is removed from the examples through a staff only setting, so none of the 39 validation papers can appear in a scoring prompt. Papers excluded for leakage: all 39 2025 papers. Papers in the validation set: 39.

Definitions

Exact: TBASE score equals the TEA total. Adjacent: exactly 1 point apart. Within one point: exact plus adjacent. Off by 2 or more: listed by paper on each grade page. Quadratic weighted kappa (QWK): chance corrected agreement that penalizes larger misses more; 1.0 is perfect, 0 is chance, and above 0.80 is commonly read as strong agreement. Mean difference: TBASE minus TEA; negative means TBASE scored lower.

Grade 3: results

Exact	Adjacent (1 point off)	Exact plus adjacent
54%	**31%**	**85%**
7 of 13 papers	4 of 13 papers	11 of 13 papers

13 of 13 papers scored. Mean difference -0.46. QWK 0.68.

Paper id	TEA score	TBASE score	Difference
2025-G3-ECR-SP0-R1	0	0	+0
2025-G3-ECR-SP0-R2	0	0	+0
2025-G3-ECR-SP1-R1	1	0	-1
2025-G3-ECR-SP1-R2	1	3	+2
2025-G3-ECR-SP2-R1	2	2	+0
2025-G3-ECR-SP2-R2	2	2	+0
2025-G3-ECR-SP3-R1	3	2	-1
2025-G3-ECR-SP3-R2	3	2	-1
2025-G3-ECR-SP3-R3	3	3	+0
2025-G3-ECR-SP4-R1	4	0	-4
2025-G3-ECR-SP4-R2	4	3	-1
2025-G3-ECR-SP5-R1	5	5	+0
2025-G3-ECR-SP5-R2	5	5	+0

Confusion matrix (rows TEA score, columns TBASE score; counts)

TEA \ TBASE	0	1	2	3	4	5
0	2	.	.	.	.	.
1	1	.	.	1	.	.
2	.	.	2	.	.	.
3	.	.	2	1	.	.
4	1	.	.	1	.	.
5	.	.	.	.	.	2

Off by 2 or more points (2): 2025-G3-ECR-SP1-R2 (TEA 1, TBASE 3), 2025-G3-ECR-SP4-R1 (TEA 4, TBASE 0).
First five sentences check failed (score forced to 0): 2025-G3-ECR-SP0-R1, 2025-G3-ECR-SP0-R2, 2025-G3-ECR-SP1-R1, 2025-G3-ECR-SP4-R1.

Grade 4: results

Exact	Adjacent (1 point off)	Exact plus adjacent
46%	**46%**	**92%**
6 of 13 papers	6 of 13 papers	12 of 13 papers

13 of 13 papers scored. Mean difference +0.31. QWK 0.87.

Paper id	TEA score	TBASE score	Difference
2025-G4-ECR-SP0-R1	0	0	+0
2025-G4-ECR-SP0-R2	0	0	+0
2025-G4-ECR-SP1-R1	1	2	+1
2025-G4-ECR-SP1-R2	1	2	+1
2025-G4-ECR-SP2-R1	2	3	+1
2025-G4-ECR-SP2-R2	2	2	+0
2025-G4-ECR-SP3-R1	3	2	-1
2025-G4-ECR-SP3-R2	3	3	+0
2025-G4-ECR-SP3-R3	3	5	+2
2025-G4-ECR-SP4-R1	4	5	+1
2025-G4-ECR-SP4-R2	4	3	-1
2025-G4-ECR-SP5-R1	5	5	+0
2025-G4-ECR-SP5-R2	5	5	+0

Confusion matrix (rows TEA score, columns TBASE score; counts)

TEA \ TBASE	0	1	2	3	4	5
0	2	.	.	.	.	.
1	.	.	2	.	.	.
2	.	.	1	1	.	.
3	.	.	1	1	.	1
4	.	.	.	1	.	1
5	.	.	.	.	.	2

Off by 2 or more points (1): 2025-G4-ECR-SP3-R3 (TEA 3, TBASE 5).
First five sentences check failed (score forced to 0): 2025-G4-ECR-SP0-R1, 2025-G4-ECR-SP0-R2.

Grade 5: results

Exact	Adjacent (1 point off)	Exact plus adjacent
38%	**54%**	**92%**
5 of 13 papers	7 of 13 papers	12 of 13 papers

13 of 13 papers scored. Mean difference +0.38. QWK 0.86.

Paper id	TEA score	TBASE score	Difference
2025-G5-ECR-SP0-R1	0	0	+0
2025-G5-ECR-SP0-R2	0	2	+2
2025-G5-ECR-SP1-R1	1	0	-1
2025-G5-ECR-SP1-R2	1	2	+1
2025-G5-ECR-SP2-R1	2	2	+0
2025-G5-ECR-SP2-R2	2	3	+1
2025-G5-ECR-SP3-R1	3	2	-1
2025-G5-ECR-SP3-R2	3	3	+0
2025-G5-ECR-SP3-R3	3	4	+1
2025-G5-ECR-SP4-R1	4	5	+1
2025-G5-ECR-SP4-R2	4	5	+1
2025-G5-ECR-SP5-R1	5	5	+0
2025-G5-ECR-SP5-R2	5	5	+0

Confusion matrix (rows TEA score, columns TBASE score; counts)

TEA \ TBASE	0	1	2	3	4	5
0	1	.	1	.	.	.
1	1	.	1	.	.	.
2	.	.	1	1	.	.
3	.	.	1	1	1	.
4	.	.	.	.	.	2
5	.	.	.	.	.	2

Off by 2 or more points (1): 2025-G5-ECR-SP0-R2 (TEA 0, TBASE 2).
First five sentences check failed (score forced to 0): 2025-G5-ECR-SP0-R1, 2025-G5-ECR-SP1-R1.

Limitations

- Small sample: one prompt per grade and 13 TEA sample papers per grade (39 in all). Each paper moves a grade percentage by about 8 points. TEA chose these papers to illustrate score points, so they are not a random sample of student work.
- Automated scoring varies slightly from run to run, mainly by one point on Conventions. The figures in this report come from one complete live run.
- Three papers returned a temporary server error on the first attempt and were scored on a second attempt.
- TEA papers appear in the guides as typed text. Line breaks inside a response were collapsed to spaces.
- This report covers 2025 Grades 3 to 5. The 2025 and other released years have their own reports, and Grades 6 to 9 will follow.

This report describes measured agreement with TEA published score points. It does not show that TBASE matches TEA scoring on every unpublished paper.