Texas Through-Year Assessment Pilot (TTAP)
Year 1 Technical Report

2022–2023 School Year

Table of Contents

List of Tables

List of Figures

Introduction

The Texas Through-Year Assessment Pilot (TTAP) represents an innovative assessment model designed as a potential alternative to the State of Texas Assessment of Academic Readiness (STAAR®) summative tests. In the context of through-year assessments, this model serves as a progress monitoring system, offering students multiple opportunities throughout the school year to demonstrate their mastery of the curriculum standards. It also contributes to the prediction of their summative performance level reported at the end of the school year.

TTAP was developed through close collaboration with Texas educators, administrators, students, and families. The progress monitoring system incorporates three distinct, short testing opportunities held during the fall, winter, and spring. To ensure that all school districts can maintain their local curriculum, each TTAP progress monitoring opportunity covers the full scope of the curriculum. These opportunities use a multi-stage adaptive design, enabling shorter tests with enhanced accuracy to minimize disruptions to instructional time.
TTAP is a multi-year, fully online pilot program that was initiated in the 2022–2023 school year. The model will continue to undergo testing over several years to assess its benefits while ensuring that its design maintains the rigorous level of validity and reliability that STAAR currently meets. The ultimate goal is to establish a scoring methodology that is comparable and suitable for state accountability. Participation in TTAP is optional and does not negate the obligation to administer STAAR. For additional details about TTAP, refer to the TTAP Assessments webpage[1].

This technical report aims to provide a comprehensive overview of the 2022–2023 TTAP Assessments, focusing on six essential aspects. It delves into TTAP test design and participation, elucidates student growth across opportunities, assesses the reliability and validity of TTAP assessments, and summarizes the special studies conducted to shape TTAP design and reporting decisions.

1) **Test Design and Administration**: This section provides an overview of the TTAP assessment instruments and details involved in assessment administration, such as testing windows and the number of administrations by test title and opportunity.
2) **Test Participation**: This section delves into test participation data at the student, campus, and district levels, and the demographics of students involved.
3) **Scores and Reports**: This section summarizes performance patterns in students' scale scores, performance levels, percentage correct scores by reporting category and item difficulty level, and growth trends across multiple assessment opportunities.
4) **Reliability:** This section discusses TTAP test reliability.
5) **Validity**: This section provides criterion validity evidence for TTAP assessments that are reflected by the correlations between TTAP and STAAR summative scores.
6) **Special Studies**: This section provides a summary of each special study conducted to compare TTAP and STAAR or to inform TTAP design decisions.

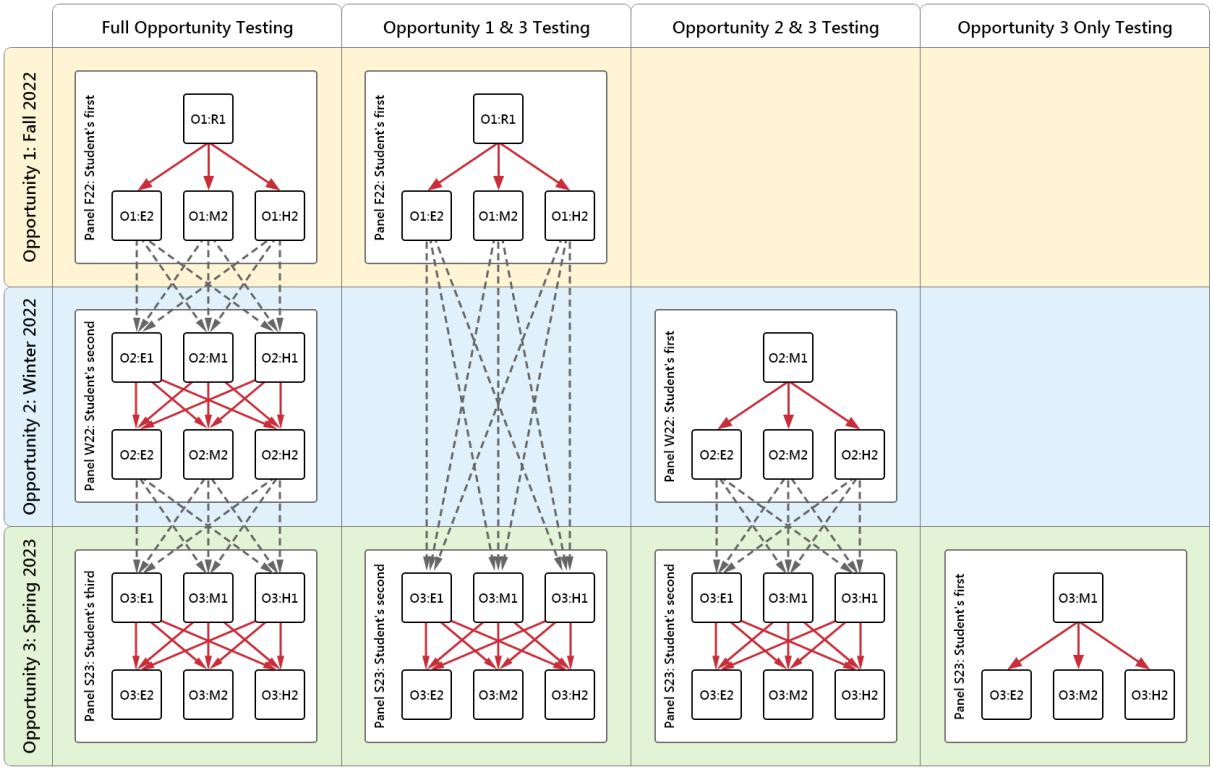
[1] https://tea.texas.gov/student-assessment/assessment-initiatives/texas-through-year-assessment-pilot

1. Test Design and Administration

1.1. Test Design

In the 2022–2023 school year, four TTAP assessments were administered, which included grade 6 mathematics, grade 7 mathematics, grade 5 science, and grade 8 social studies. Each assessment features three distinct opportunities administered in the fall, winter, and spring. Each opportunity is a multi-stage assessment with two panels or stages. The multi-stage adaptive test is depicted in Figure 1.

Figure 1: TTAP Design



O1: Opportunity I, O2: Opportunity II, O3: Opportunity III
R1: Router Segment 1, E1: Easy Segment 1, M1: Medium Segment 1, H1: Hard Segment 1
E2: Easy Segment 2, M2: Medium Segment 2, H2: Hard Segment 2

The process unfolds as follows: in Stage 1, the system scores the item responses, comparing the scores against routing cut scores established during test construction. Depending on their performance in Stage 1, students are directed to the Stage 2 panel (i.e., easy, medium, or hard) that best aligns with their demonstrated performance in Stage 1.

During Opportunity I in Stage 1, students take a router form. In Stage 2, they are directed to a form at the appropriate difficulty level based on their performance during Stage 1. In Opportunity II or III, if the opportunity is the first for the student, he or she take the medium form as the router form

For a student who has tested in a prior opportunity, Opportunities 2 and 3 start the student on the easy, medium, or hard segment, based on their ability estimated from the most recent opportunity completed and the routing rule for a specific form.

1.2. Blueprints

TTAP test forms are constructed by the vendor and approved by TEA staff based on criteria detailed in the test construction specifications and test blueprints, which represent proportionally shortened versions of STAAR summative assessments. TTAP assessment blueprints are closely aligned with the STAAR summative assessment blstudeueprints. Table 1 compares the number of items on STAAR and TTAP assessments by reporting category (RC), and Table 2 lists the names of the RCs.

Table 1. Summary of Reporting Categories

Assessment	RC 1	RC 2	RC 3	RC 4
Grade 5 Science	Matter and Energy	Force, Motion, and Energy	Earth and Space	Organisms and Environments
Grade 6 Mathematics	Numerical Representations and Relationships	Computations and Algebraic Relationships	Geometry and Measurement	Data Analysis and Personal Financial Literacy
Grade 7 Mathematics	Probability and Numerical Representations	Computations and Algebraic Relationships	Geometry and Measurement	Data Analysis and Personal Financial Literacy
Grade 8 Social Studies	History	Geography and Culture	Government and Citizenship	Economics, Science, Technology, and Society

Table 2. Comparison Between STAAR and TTAP Blueprints

Assessment	Test	RC 1	RC 2	RC 3	RC 4	Total Items
Grade 5 Science	STAAR	4–6	6–8	8–10	10–12	32
	Through Year OP1	3	4	4	6	17
	Through Year OP2	3	4	4	6	17
	Through Year OP3	4	6	8	10	28
	Through Year Total	10	14	16	22	62
Grade 6 Mathematics	STAAR	8–10	13–15	5–7	6–8	36
	Through Year OP1	5	7	3	5	20
	Through Year OP2	5	7	3	5	20
	Through Year OP3	7	11	6	6	30
	Through Year Total	17	25	12	16	70
Grade 7 Mathematics	STAAR	4–6	14–16	11–13	5–7	38
	Through Year OP1	3	7	6	4	20
	Through Year OP2	3	7	6	4	20
	Through Year OP3	4	13	10	5	32
	Through Year Total	10	27	22	13	72
Grade 8 Social Studies	STAAR	15–17	8–10	8–10	5–7	40
	Through Year OP1	8	4	5	3	20
	Through Year OP2	8	4	5	3	20
	Through Year OP3	13	8	8	5	34
	Through Year Total	29	16	18	11	74

Note: English and Spanish grade 5 science have the same blueprint and reporting categories.

1.3. 2022–2023 TTAP Administration

The 2022–2023 TTAP assessments include three testing opportunities. Table 3 represents the TTAP scopes and administration schedules. The grade 5 science assessment offered both English and Spanish test forms with identical content differing only in language. All assessments were exclusively conducted online with no paper-based options. While it was expected for participating students to complete all three assessments, some students may have missed certain opportunities. The TTAP assessments were delivered through the Test Delivery System (TDS) and include the same accommodations that are available for STAAR summative assessments.

Table 3. 2022–2023 TTAP Administration Schedule

Assessment	Opportunity I	Opportunity II	Opportunity III
Grade 5 Science	November 14–18, 2022	January 30–February 3, 2023	April 3–7, 2023
Grade 5 Science Spanish			
Grade 6 Mathematics			
Grade 7 Mathematics			
Grade 8 Social Studies			

In the 2022–2023 school year, more than 60,000 TTAP assessments were administered. Table 4 provides insight into the number of students who participated in each opportunity for each TTAP test. Additionally, the two rightmost columns present the count of students who completed all three opportunities of a TTAP test and those who took at least one opportunity of a TTAP test. The numbers in Table 4 reflect sample sizes following the application of exclusion rules, which help exclude test cases like off-grade test takers and students who did not meet attemptedness rules. A comprehensive list of these exclusion rules can be found in Appendix A. It is worth noting that the Spanish grade 5 science test has a relatively small sample size, which could potentially limit the interpretability of results. In contrast, the other four tests all have sample sizes exceeding 7,000 ensuring that meaningful results can be derived from the data.

Table 4. TTAP Administered in the 2022–2023 School Year

Assessment	Opp. I (N)	Opp. II (N)	Opp. III (N)	Took All 3 Opps (N)	Took at Least 1 Opp. (N)
Grade 5 Science	16,614	16,968	16,741	15,250	17,864
Grade 5 Science Spanish	233	260	299	209	313
Grade 6 Mathematics	10,084	10,196	9,920	8,984	10,854
Grade 7 Mathematics	8,669	8,735	8,659	7,561	9,544
Grade 8 Social Studies	24,317	24,585	24,348	21,980	26,163
Total	59,917	60,744	59,967	53,984	64,738

1.4. Test Participation

Table 5 provides additional insight into the counts of students that engaged in at least one TTAP assessment during the 2022–2023 school year. A total of 121 school districts and 64,738 students participated in TTAP administrations.

Table 5. 2022–2023 TTAP Student Participation

Assessment	Number of Unique Students
Grade 5 Science	17,864
Grade 5 Science Spanish	313
Grade 6 Mathematics	10,854
Grade 7 Mathematics	9,544
Grade 8 Social Studies	26,163
Total	64,738

In addition, the demographic characteristics of 2022–2023 TTAP assessment participants has been compared with the state's student population in the same year to evaluate the sample

representativeness of TTAP participants. Summarized demographic data for all students who took the STAAR summative tests in spring 2023 and those who participated in at least one TTAP assessment are presented in Table 6 through Table 10. The variable names and mappings can be found in Appendices B and C.

Notably, in nearly all demographic comparisons, the percentages within each category exhibit similarities, with differences consistently below 5 percent. However, there are a few exceptions to this trend. All percentage differences exceeding 5 percent are marked in bold within the tables. It is important to acknowledge that the TTAP sample size for Spanish grade 5 science is relatively small, limiting the meaningfulness of direct comparisons. When analyzing the other tests in comparison to their respective state student populations, the following trends are noticed:

1) There is a slightly higher representation of white students (grade 5, 6) and Title I students (grade 5) in the TTAP assessments.
2) Conversely, there are slightly lower percentages of Hispanic/Latino students (grade 5), limited English proficiency students (grade 5, 6, 7), bilingual students (grade 5), and English as a Second Language (ESL) students (grade 7) among TTAP participants.

These observations provide insights into the demographic composition of TTAP assessment participants in relation to the broader student population even though the differences are generally small. While a few variations can be observed, these differences are generally minor and should not be overemphasized.

Table 6. TTAP Participating Student Demographic Characteristics (Grade 5 Science)

Demographic	STAAR Spring 2023	TTAP 2022–2023	Difference in Percentage
Number of Students	378,742	17,864	NA
Male	50.8	51.4	0.6
Female	48.9	48.5	0.4
Hispanic/Latino	50.9	45.8	5.1
American Indian or Alaska Native	0.3	0.2	0.1
Asian	5.7	4.7	1.0
Black or African American	12.8	13.5	0.7
Native Hawaiian or Pacific Islander	0.2	0.1	0.1
White	26.6	31.8	5.2
Two or More Races	3.1	3.0	0.1
Economically Disadvantaged	60.2	60.3	0.1
Title I, Part A Participants	70.8	77.9	7.1
Migrant	0.3	0.6	0.3
Current Limited English Proficient	24.7	16.7	8.0
Bilingual	12.5	6.9	5.6
ESL Participants	7.3	6.2	1.1
Special Education	14.4	15.3	0.9
Gifted/Talented Participants	11.4	10.9	0.5
At-Risk	50.6	46.2	4.4

Table 7. TTAP Participating Student Demographic Characteristics (Spanish Grade 5 Science)

Demographic	STAAR Spring 2023	TTAP 2022–2023	Difference in Percentage
Number of Students	9,775	313	NA
Male	50.0	47.9	-2.1
Female	49.8	51.4	+1.6
Hispanic/Latino	97.9	93.0	-4.9
American Indian or Alaska Native	0.3	0.6	+0.3
Asian	0.0	NA	NA
Black or African American	0.0	NA	NA
Native Hawaiian or Pacific Islander	0.0	NA	NA
White	1.0	1.3	+0.3
Two or More Races	0.1	NA	NA
Economically Disadvantaged	82.4	73.5	-8.9
Title I, Part A Participants	91.7	87.2	-4.5
Migrant	0.6	0.6	0.0
Current Limited English Proficient	98.1	99.0	+0.9
Bilingual	79.5	72.8	-6.7
ESL Participants	4.5	15.3	+10.8
Special Education	7.1	4.8	-2.3
Gifted/Talented Participants	3.8	4.8	+1.0
At-Risk	91.9	85.3	-6.6

Table 8. TTAP Participating Student Demographic Characteristics (Grade 6 Mathematics)

Demographic	STAAR Spring 2023	TTAP 2022–2023	Difference in Percentage
Number of Students	384,766	10,854	NA
Male	50.6	50.6	0.0
Female	49.1	49.3	0.2
Hispanic/Latino	52.8	50.2	2.6
American Indian or Alaska Native	0.3	0.3	0.0
Asian	5.0	5.4	0.4
Black or African American	12.8	9.9	2.9
Native Hawaiian or Pacific Islander	0.2	0.1	0.1
White	25.5	30.9	**5.4**
Two or More Races	3.0	2.9	0.1
Economically Disadvantaged	60.9	60.2	0.7
Title I, Part A Participants	62.1	60.8	1.3
Migrant	0.3	0.8	0.5
Current Limited English Proficient	25.8	17.3	**8.5**
Bilingual	3.0	0.2	2.8
ESL Participants	17.5	13.1	4.4
Special Education	13.3	13.9	0.6
Gifted/Talented Participants	10.7	6.9	3.8
At-Risk	54.7	50.6	4.1

Table 9. TTAP Participating Student Demographic Characteristics (Grade 7 Mathematics)

Demographic	STAAR Spring 2023	TTAP 2022–2023	Difference in Percentage
Number of Students	331,698	9,544	NA
Male	50.8	51.1	0.3
Female	49.0	48.8	0.2
Hispanic/Latino	54.6	55.5	0.9
American Indian or Alaska Native	0.3	0.3	0.0
Asian	4.3	1.6	2.7
Black or African American	13.1	11.5	1.6
Native Hawaiian or Pacific Islander	0.2	0.1	0.1
White	24.4	28.0	3.6
Two or More Races	2.7	2.5	0.2
Economically Disadvantaged	63.2	67.9	4.7
Title I, Part A Participants	61.6	63.3	1.7
Migrant	0.3	0.8	0.5
Current Limited English Proficient	25.6	18.2	**7.4**
Bilingual	1.0	0.1	0.9
ESL Participants	19.2	12.7	**6.5**
Special Education	13.5	15.3	1.8
Gifted/Talented Participants	7.4	4.7	2.7
At-Risk	56.9	58.0	1.1

Table 10. TTAP Participating Student Demographic Characteristics (Grade 8 Social Studies)

Demographic	STAAR Spring 2023	TTAP 2022–2023	Difference in Percentage
Number of Students	414,692	26,163	NA
Male	50.9	50.7	0.2
Female	48.8	49.3	0.5
Hispanic/Latino	53.3	53.4	0.1
American Indian or Alaska Native	0.3	0.3	0.0
Asian	5.0	4.8	0.2
Black or African American	12.7	11.8	0.9
Native Hawaiian or Pacific Islander	0.2	0.1	0.1
White	25.3	26.5	1.2
Two or More Races	2.8	2.6	0.2
Economically Disadvantaged	59.8	60.1	0.3
Title I, Part A Participants	59.2	63.2	4.0
Migrant	0.3	0.4	0.1
Current Limited English Proficient	22.9	19.5	3.4
Bilingual	0.7	0.6	0.1
ESL Participants	17.9	13.6	4.3
Special Education	11.0	11.0	0.0
Gifted/Talented Participants	10.7	9.8	0.9
At-Risk	53.6	51.3	2.3

2. Scores and Reports

At the individual student level, the reported scores included item scores (i.e., whether a student answered each item correctly), scale scores, score gain/loss/no change between opportunities, percentage of correct responses categorized by reporting category and item difficulty level, and current performance levels.

In this section, a detailed overview of the results from each of these reported scores is provided. Additionally, a comprehensive comparison of these reported scores across multiple opportunities is available to identify insights into the trends and patterns of student growth as they progress through the school year.

2.1. Scale Score Gain or Loss Between Opportunities

One of the reported scores is the scale score, which allows comparisons across different opportunities and test forms. Students' growth in terms of their scale scores across three opportunities is analyzed. Descriptive statistics of scale scores from each opportunity are presented in Table 11. In general, students' average scale scores exhibit an increase across opportunities, except for Spanish grade 5 science where the observed anomaly may be attributed to the relatively small sample size. Notably, the change in scale scores for social studies appears relatively modest when compared to the other subjects.

Table 11. Student Score Growth Across TTAP Opportunities

Assessment	Opportunity	N	Mean	SD	Min	25th P	50th P	75th P	Max
Grade 5 Science	Opp. I	16,614	832.383	48.707	562	801	832	868	1092
	Opp. II	16,968	843.621	52.638	562	811	848	881	1092
	Opp. III	16,741	853.453	57.019	660	818	860	890	1092
Grade 5 Science Spanish	Opp. I	233	810.678	46.231	706	774	812	838	934
	Opp. II	260	813.685	50.595	685	780	810	843	960
	Opp. III	299	811.632	53.137	702	772	811	852	1000
Grade 6 Mathematics	Opp. I	10,084	625.833	117.675	64	545	632	703	1191
	Opp. II	10,196	659.356	142.820	159	564	648	759	1354
	Opp. III	9,920	683.654	154.826	199	579	673	791	1315
Grade 7 Mathematics	Opp. I	8,669	672.314	122.541	119	587	671	741	1412
	Opp. II	8,735	701.761	129.207	119	617	698	777	1412
	Opp. III	8,659	704.410	132.251	270	612	708	788	1289
Grade 8 Social Studies	Opp. I	24,317	900.641	49.205	639	864	901	937	1176
	Opp. II	24,585	906.696	49.785	639	871	908	941	1176
	Opp. III	24,348	913.803	51.389	733	874	915	948	1150

Note: The notations 25th P, 50th P, and 75th P correspond to the 25th, 50th, and 75th percentiles, respectively.

To evaluate the magnitude of scale score growth across opportunities, the effect size of scale score gain between opportunities is calculated and presented in Table 12. The effect size is determined using Cohen's *d*, a widely used statistical measure that quantifies the effect size of the difference between two groups or conditions and assesses the magnitude of an effect. In addition to scale scores, students receive a gain, loss, or no change score that refect their scale score changes across opportuniteis. The table also presents the percentage of students who experienced gains, losses, or no changes in their scale scores.

For reference, Cohen's *d* values are typically interpreted as follows: approximately 0.2 signifies a small effect size; 0.5 represents a medium effect size; and values around 0.8 or higher indicate a large effect size. The effect sizes in Table 8 are around 0.2 or lower implying that the observed growths in scale scores are relatively small. For Spanish grade 5 science, the effect size is close to 0 across all opportunities. For grade 7 mathematics, the effect size is close to 0 for growth between Opportunities II and III indicating minimal change between winter and spring. Overall, the effect

sizes for Opportunity II versus Opportunity I tend to be larger than those for Opportunity III versus Opportunity II suggesting that students showed more progress from fall to winter than they did from winter to spring. The effect sizes reflecting annual growth, specifically between Opportunity III and Opportunity I range from small to medium.

These trends are similarly reflected in the percentages of students who gained, lost, or experienced no change in their scale scores from Opportunity II to Opportunity I, Opportunity III to Opportunity II, and Opportunity III to Opportunity I presented in Table 13.

Table 12. Effect Size of Student Scale Score Growth Across TTAP Opportunities

Assessment	Opp. II vs. Opp. I	Opp. III vs. Opp. II	Opp. III vs. Opp. I
Grade 5 Science	0.222	0.179	0.397
Grade 5 Science Spanish	0.062	-0.039	0.019
Grade 6 Mathematics	0.256	0.163	0.420
Grade 7 Mathematics	0.234	0.020	0.252
Grade 8 Social Studies	0.122	0.140	0.262

Table 13. Percentage of Students with Gain, Loss, or No Change in Scale Scores Across TTAP Opportunities

Assessment	Opp. II vs. Opp. I Percentage of Gain/Loss/No Change			Opp. III vs. Opp. II Percentage of Gain/Loss/No Change			Opp. III vs. Opp. I Percentage of Gain/Loss/No Change		
	Loss	Gain	No Change	Loss	Gain	No Change	Loss	Gain	No Change
Grade 5 Science	36.2	62.0	1.7	38.3	60.7	1.1	26.7	72.3	1.0
Grade 5 Science Spanish	42.1	57.0	0.9	50.8	47.6	1.6	41.4	58.1	0.5
Grade 6 Mathematics	37.4	62.6	0.0	38.8	61.0	0.2	27.2	72.5	0.3
Grade 7 Mathematics	35.9	63.6	0.5	47.2	52.1	0.7	34.3	65.3	0.4
Grade 8 Social Studies	43.6	55.7	0.7	40.7	58.7	0.6	33.5	65.3	1.3

2.2. Performance Level

Student performance on TTAP is categorized into four performance levels: 1) *Currently Does Not Meet Grade Level*, 2) *Currently Approaches Grade Level*, 3) *Currently Meets Grade Level*, and 4) *Currently Masters Grade Level*. The distribution of students among these performance levels is summarized in Table 17 for each TTAP opportunity, as well as the distribution of performance levels for STAAR. Overall, students exhibited a trend of advancing to higher achievement levels across the opportunities. When comparing the distribution of students' performance levels between TTAP Opportunity III and STAAR, it is notable that STAAR reports higher percentages of

students at the *Masters* and *Meets* levels than TTAP. In general, the percentages at each performance level between TTAP Opportunity III and STAAR are showing similar trends.

Table 14. Student Performance-Level Distribution Across TTAP Opportunities

Assessment	PL	Opp. I (N)	Opp. II (N)	Opp. III (N)	STAAR (N)	Opp. I (%)	Opp. II (%)	Opp. III (%)	STAAR (%)
Grade 5 Science	1	10,657	9,027	7,387	6,768	64.1	53.2	44.1	37.9
	2	4,633	5,229	5,303	5,231	27.9	30.8	31.7	29.3
	3	917	1,762	2,435	3,291	5.5	10.4	14.5	18.4
	4	407	950	1,616	2,574	2.4	5.6	9.7	14.4
	Total	16,614	16,968	16,741	17,864	100.0	100.0	100.0	100.0
Grade 5 Science Spanish	1	187	207	220	199	80.3	79.6	73.6	63.6
	2	35	36	59	79	15.0	13.8	19.7	25.2
	3	10	7	18	21	4.3	2.7	6.0	6.7
	4	1	10	2	14	0.4	3.8	0.7	4.5
	Total	233	260	299	313	100.0	100.0	100.0	100.0
Grade 6 Mathematics	1	4,572	4,006	3,290	2,639	45.3	39.3	33.2	24.3
	2	3,957	3,542	3,405	4,066	39.2	34.7	34.3	37.5
	3	1,400	2,066	2,258	2,545	13.9	20.3	22.8	23.4
	4	155	582	967	1,604	1.5	5.7	9.7	14.8
	Total	10,084	10,196	9,920	10,854	100.0	100.0	100.0	100.0
Grade 7 Mathematics	1	5,529	4,655	4,137	3,942	63.8	53.3	47.8	41.3
	2	1,931	1,959	2,385	2,640	22.3	22.4	27.5	27.7
	3	1,064	1,882	1,904	2,290	12.3	21.5	22.0	24.0
	4	145	239	233	672	1.7	2.7	2.7	7.0
	Total	8,669	8,735	8,659	9,544	100.0	100.0	100.0	100.0
Grade 8 Social Studies	1	13,542	13,043	11,526	10,374	55.7	53.1	47.3	39.7
	2	6,980	6,749	6,810	7,493	28.7	27.5	28.0	28.6
	3	2,403	3,037	3,487	4,084	9.9	12.4	14.3	15.6
	4	1,392	1,756	2,525	4,212	5.7	7.1	10.4	16.1
	Total	24,317	24,585	24,348	26,163	100.0	100.0	100.0	100.0

Note: PL stands for Performance Level. In the PL column, 1 = Currently Does Meet Grade Level, 2 = Currently Approaches Grade Level, 3 = Currently Meets Grade Level, and 4 = Currently Masters Grade Level.

3. Reliability

A marginal reliability coefficient (Samejima, 1977, 1994) is used to evaluate the internal test reliability. This measure evaluates how well the items on a test that reflect the same construct yield similar results. Marginal reliability is the result of combining measurement errors estimated at different points on the achievement scale into a single index. The formula used to calculate marginal reliability is:

$$\rho_\theta = \frac{\sigma_\theta^2 - M_{S_\theta^2}}{\sigma_\theta^2}$$

where σ_θ^2 is the observed variance of the ability estimates, θ, and $M_{S_{\hat{\theta}}^2}$ is the observed mean of the score's conditional error variances at each value of θ. In the field of educational measurement, assessments are considered reliable when their reliability coefficients are 0.80 and above. Typically, high-stakes assessments achieve higher levels of reliability in the range of upper 0.80s to lower 0.90s (Dorans et al., 2007; Phillips & Camara, 2006).

Table 18 provides a comparison of the marginal reliability coefficients for TTAP and STAAR during the 2022–2023 school year. The table also includes reliabilities at the subgroup level for gender and ethnicity but only for subgroups with sample sizes equal to or larger than 200. Reliabilities for smaller subgroups are omitted to prevent potentially misleading conclusions based on limited data.

When assessing the three opportunities within TTAP, it is evident that Opportunity I exhibits relatively lower reliabilities while Opportunity III demonstrates the highest reliabilities. The longer test length of Opportunity III contributes to the expected increase in reliability. Comparing the reliability of TTAP Opportunity III with STAAR, Opportunity III demonstrates higher reliabilities for English and Spanish grade 5 science, as well as grade 6 mathematics. However, STAAR reports higher reliabilities for grade 7 mathematics and grade 8 social studies both at the overall level and in certain subgroup analyses.

Upon examining reliabilities at the subgroup level, there is a general pattern of comparability across subgroups, with a few exceptions for Black or African American and Hispanic or Latino students. In certain instances, the reliabilities for these subgroups fall below 0.7 (indicated by bolded values in Table 16), which are relatively lower when compared to other ethnicity subgroups.

Table 15. Test Reliabilities for TTAP and STAAR

Assessment	Group	N	Opp. I	Opp. II	Opp. III	STAAR
Grade 5 Science	All	17,864	0.750	0.787	0.880	0.855
	Ethnic: A	844	0.706	0.747	0.856	0.825
	Ethnic: B	2,409	0.737	0.786	0.874	0.810
	Ethnic: H	8,178	0.725	0.771	0.869	0.830
	Ethnic: T	542	0.741	0.780	0.867	0.847
	Ethnic: W	5,678	0.716	0.744	0.850	0.846
	Sex: F	8,663	0.737	0.763	0.873	0.847
	Sex: M	9,181	0.760	0.803	0.886	0.859
Grade 5 Science Spanish	All	313	0.734	0.778	0.867	0.779
	Ethnic: H	291	0.744	0.779	0.870	0.783
Grade 6 Mathematics	All	10,854	0.736	0.824	0.896	0.889
	Ethnic: A	589	0.768	0.829	0.895	0.848
	Ethnic: B	1,071	**0.698**	0.781	0.866	0.839
	Ethnic: H	5,448	**0.695**	0.796	0.874	0.862
	Ethnic: T	315	0.728	0.828	0.892	0.887
	Ethnic: W	3,354	0.724	0.813	0.886	0.886
	Sex: F	5,349	0.723	0.819	0.892	0.887
	Sex: M	5,496	0.747	0.829	0.899	0.891
Grade 7 Mathematics	All	9,544	0.736	0.780	0.869	0.872
	Ethnic: B	1,097	**0.684**	0.746	0.844	0.828
	Ethnic: H	5,296	0.721	0.770	0.865	0.864
	Ethnic: T	243	0.737	0.775	0.873	0.858
	Ethnic: W	2,674	0.744	0.772	0.862	0.874
	Sex: F	4,660	0.714	0.770	0.861	0.868
	Sex: M	4,873	0.753	0.788	0.877	0.876
Grade 8 Social Studies	All	26,163	0.765	0.797	0.882	0.894
	Ethnic: A	1,261	0.757	0.792	0.871	0.862
	Ethnic: B	3,092	0.740	0.788	0.868	0.875
	Ethnic: H	13,971	0.736	0.771	0.866	0.874
	Ethnic: T	687	0.773	0.800	0.880	0.895
	Ethnic: W	6,927	0.754	0.785	0.878	0.892
	Sex: F	12,886	0.740	0.778	0.871	0.887
	Sex: M	13,256	0.783	0.812	0.890	0.899

Note: Reliability is only reported for subgroups with sample sizes equal to or greater than 200.
Sex: F – Female, Sex: M – Male
Ethnic: A – Asian, B – Black or African American, H – Hispanic/Latino, T – Two races, W – White

4. Validity

Correlations between TTAP and STAAR scale scores are calculated as criterion validity evidence of the TTAP scores. Correlation is a statistical measure that quantifies the strength and direction of the linear relationship between two continuous variables. It provides a value between -1 and +1, where -1 indicates a perfect negative linear relationship, +1 indicates a perfect positive linear relationship, and 0 suggests no linear relationship between the variables. Table 19 shows the correlations between TTAP and STAAR scores by opportunity, subject, and grade.

Table 19 also showcases patterns of associations across different TTAP opportunities and STAAR. Across the various values in the table, the correlations between Opportunity I, Opportunity II, and Opportunity III are moderately strong, generally ranging between 0.677 and 0.799. This suggests a consistent positive relationship in scores across these opportunities. The correlations with STAAR are also moderately strong, falling between 0.673 and 0.846. It is notable that the correlation values between Opportunity III and STAAR tend to be higher than those between Opportunity I or Opportunity II and STAAR, implying that Opportunity III might be a better predictor of STAAR scores. Overall, the results indicate moderate to strong positive relationships between the various TTAP opportunities and STAAR, with a more pronounced relationship in the latter opportunities. The correlations, considered criterion validity evidence of TTAP scores, are moderately high.

Table 16. Correlation Coefficients Between TTAP and STAAR Scale Scores

Assessment	Opp.	Opp. I	Opp. II	Opp. III	STAAR
Grade 5 Science	Opp. I	1	0.698	0.744	0.723
	Opp. II	-	1	0.757	0.721
	Opp. III	-	-	1	0.809
	STAAR	-	-	-	1
Grade 5 Science Spanish	Opp. I	1	0.689	0.700	0.673
	Opp. II	-	1	0.694	0.720
	Opp. III	-	-	1	0.743
	STAAR	-	-	-	1
Grade 6 Mathematics	Opp. I	1	0.727	0.736	0.719
	Opp. II	-	1	0.799	0.779
	Opp. III	-	-	1	0.846
	STAAR	-	-	-	1
Grade 7 Mathematics	Opp. I	1	0.677	0.723	0.712
	Opp. II	-	1	0.760	0.741
	Opp. III	-	-	1	0.811
	STAAR	-	-	-	1
Grade 8 Social Studies	Opp. I	1	0.741	0.761	0.752
	Opp. II	-	1	0.789	0.785
	Opp. III	-	-	1	0.845
	STAAR	-	-	-	1

5. Special Study Summaries

In the 2022–2023 school year, four special studies were conducted to assess the technical quality of TTAP, comparing it to STAAR, and providing valuable insights for TTAP test design and score reporting. The findings from these studies are detailed in this section. To further enhance TTAP reliability and validity and adhere to the highest standards of technical rigor, a series of forthcoming special studies are planned for the upcoming school year. The planned future studies will ensure TEA continuously refines TTAP based on empirical evidence and makes informed decisions regarding test design and score reporting.

Special Study 1: Comparing the Psychometric Properties of TTAP Opportunity III and STAAR

The objective of this study was to investigate the comparability of TTAP Opportunity III scores with those of STAAR. Simulations were conducted to compare the technical characteristics of TTAP Opportunity III and STAAR. Comparisons focused on ability recovery, classification accuracy, performance-level distribution, and reliability of both assessments. The analyses revealed that, in general, TTAP Opportunity III and STAAR exhibit similar technical properties with respect to the aforementioned aspects. However, STAAR demonstrated marginally higher correlations between the true theta and estimated theta for all tests, slightly higher classification accuracies for all tests except for the grade 7 mathematics test, and marginally higher reliabilities for all tests except for the grade 8 social studies test. Except where noted, these differences are minimal and only observable at the thousandths place. Additionally, as TTAP Opportunity III tests are shorter than STAAR, the former proves to be more efficient in achieving a similar level of reliability and classification accuracy. This finding aligns with results from earlier studies (Armstrong et al., 2004; Pustule, 1999).

This study shows preliminary positive evidence that TTAP Opportunity III and STAAR would likely provide comparable interpretations of student ability if administered within the same testing window. However, it is crucial to acknowledge a limitation of this simulation study, which assumes uniformity in student characteristics, motivation, and administration conditions across assessments. Subsequent analysis of this high-stakes data would enhance the robustness of the conclusions.

Special Study 2: A TTAP-to-STAAR ROC Study

Through-year assessments are multiple-administration assessments that are designed to meet the intended uses and purposes of interim and summative assessments and bridge these two assessments into a single, coherent assessment system. One purpose of through-year assessments is to predict if a student will achieve a performance level of importance on a large-scale summative assessment (Perlie, Marion, & Gong, 2009). In previous work (Schneider, Liu, & Robinson, 2022), Cambium Assessment, Inc. (CAI) found that the receiver operating characteristic (ROC) curves functioned effectively to identify cuts and provide predictions that give teachers information regarding whether a student is likely or unlikely to meet passing standards. The findings from this previous work underscore the recommendation to annually recalibrate the cuts using the ROC methodology, especially while navigating the uncertainties stemming from evolving achievement trends, as observed during the pandemic.

In this study, CAI applied the ROC methodology to each of the three TTAP opportunities to predict students' STAAR performance levels using data from the 2022–2023 school year. The predictions were evaluated in terms of sensitivity, specificity, and area under the curve (AUC). The results indicate that the ROC method, when applied to TTAP score scales, yields sensitivity, specificity, and AUC values that generally meet or exceed desired thresholds, affirming the reliability of the predictions. The predictions will provide teachers and students useful information in terms of whether a student is likely or unlikely to meet passing standards. This prediction method is being implemented in 2023–2024 TTAP tests to provide students their predicted STAAR performance levels for each opportunity. Subsequent analyses of how well the ROC method predicts students' actual STAAR performance levels and how teachers and students use these predicted levels will be considered in special studies of Year 3.

Special Study 3: A Comparison of Through-Year Cumulative Scoring Methods

This study compared eight different approaches to producing a summative score using TTAP test scores. The first method simply used the last TTAP test score. The second method used the maximum score from all three opportunities as the summative score. The third, fourth, and fifth methods were linear composite scores and used the following general equation: composite score = test 1 • weight 1 + test 2 • weight 2 + test 3 • weight 3. These linear composites scores were similar to weighted averages. For example, if the weights were all set to ⅓, then the composite score would be a simple average. The three composite scores differed in how the weights were defined.

The first composite method, opportunity-to-learn weighted (OW) score, set the weights proportional to typical opportunity-to-learn estimates with typical cumulative instructional time. This resulted in larger weights for the last test and smallest weights for the first test. The next composite method adjusted the OW method using test length. This approach, OW hybrid, corrected for differences in test length between the first two TTAP tests and the third test. This approach resulted in even greater weight being assigned to the last test and even less weight assigned to the first test. The last composite method set weights proportional to the reliability of the test scores. This method, the reliability weighted (RW) score, gave more weight to the third test because that test was longer and more reliable. These three composite scores were produced for all students; however, the scores were created a second time using a help-not-hurt (HNH) rule, which means the linear composite score is used only if it is higher than the third TTAP score. This resulted in two versions of each composite score, one without the HNH rule and one with the HNH rule, for a total of six composite methods.

After producing scores from the eight different approaches (Opportunity III score, maximum score, three non-HNH composite scores, and three HNH composite scores) for each student, the eight scores were compared to STAAR scores to see how similar the scores were in terms of reliability, bias, root mean square error, correlations, and classification decisions. The maximum scores showed the closest similarity to STAAR scores, followed by the HNH composite scores; however, it should be noted that TTAP Opportunity III was administered almost one month prior to STAAR. In an effort to correct for this timing issue, a linear regression was used to project Opportunity III scores to coincide with the timing of STAAR scores. This made the projected TTAP Opportunity III score distribution very similar to the STAAR score distribution.

Next, all the composite scores were reproduced using the projected TTAP Opportunity III scores. In this case, the maximum scores and the HNH composite scores overestimated STAAR scores, but the HNH composite scores overestimated to a much lesser degree. Overall, the maximum score performed best when using only the observed scores; however, using the projected TTAP Opportunity III scores, the third TTAP scores performed best, followed by the HNH composite scores. The non-HNH composite scores performed well in terms of classification agreement with STAAR scores but suffered from a negative bias, meaning they always underestimated STAAR scores.

Each scoring method has benefits and disadvantages, so selecting between the options requires one to weigh the net benefits of each. It is important to examine the tradeoffs wholistically. Table 20 attempts to weigh the tradeoffs and the unknowns of each method. Recommendations are provided in the last column. The research questions and recommendations were influenced by discussions with the technical advisor committee.

Table 17. Tradeoffs of Different Cumulative Scoring Methods

Method	Tradeoffs: Advantages	Tradeoffs: Disadvantages	Unknowns	Recommendations
TTAP Opportunity III Score	• Avoids Opp. I & Opp. II missing data • Most comparable to STAAR (CAI, 2023) • Opp. III is informed by Opp. II • Only Opp. III has to meet federal requirements	• Does not give credit for higher possible score obtained in Opp. I or Opp. II • Like STAAR, Opp. III is required	Not sure how sensitive Opp. III is to biased starting thetas in MST	• The sensitivity of MST to biased starting thetas should be analyzed. • Evidence of predictive validity should be collected.
Maximum Score	• Allows mastery to be "banked" • Most closely replicates STAAR score distributions • Aligns with Superscoring of College Readiness Assessments	• All three opportunities must meet federal requirements • Variable test lengths mean variable score reliabilities and non-equivalence may not meet federal requirements • May overestimate end-of-year ability	• Not sure if variable score reliabilities of maximum scores will meet federal requirements • Not sure of the unintended consequences on incentives	Decide if equal test lengths is feasible; if not, research federal requirements and guidelines to determine if variable reliabilities/non-equivalence can be justified.
Composite Scores* with HNH Rule	• Composite scores are marginally more reliable than Opp. III alone • Uses all three scores • Adults believe students will provide effortful responses • Innovative	• All three opportunities must meet federal requirements • Computationally complex • Mix of scales (Opp. III + composite) vary in reliability and may not be equivalent • Policy for missing scores must be made and validated	Not sure if summative scores containing a mix of scales with variable score reliabilities will meet federal requirements	Examine the degree to which scores are equivalent. Conduct experimental research to examine the effects of missing scores and differing policies have on score accuracy.
Composite Scores* without HNH Rule	• Most equivalent and reliable within missing score patterns • Uses all three scores • Adults believe students will provide effortful responses • Innovative	• All three opportunities must meet federal requirements • Scores are usually biased low, especially when growth is high • Composite score is commonly lower than Opportunity III scores • Missing scores means weights are not equivalent for all students	Not sure if there are viable solutions to missing scores that can make them equivalent across missing score patterns Not sure if the negative bias in composite scores and classification can be corrected	Conduct experimental research to examine the effects that missing scores and differing policies have on score accuracy. Explore approaches for reducing negative score bias and classification bias of composite scores, especially when growth is variable.

Note: Of the composite score methods, OW hybrid performed best when considering all metrics.

Special Study 4: CAT Feasibility Study for TTAP

This study investigated the performance of item-level computer adaptive tests (CAT) using the CAI item selection algorithm and different configurations of an existing TTAP item pool. The objective of this study was to determine how feasible CAT would be using an existing item pool and to gain a better understanding of the tradeoffs of CAT and MST. Both CAT and MST are adaptive, however, MST adapts once per test, while CAT adapts each time an item is selected; hence, CAT is more adaptive than MST and may offer greater measurement efficiencies.

This study simulated the three TTAP testing opportunities across two growth conditions (realistic growth versus no growth) using two pool configurations for CAT: a small pool without overlapping items and a large item pool with significant overlap of items. Key metrics of CAT, MST, and STAAR were compared including ability recovery, classification accuracy, performance-level distribution, reliability, item exposure, and utility.

The analyses revealed that, in general, both CAT and MST designs are feasible and exhibited similar technical properties. Most differences between tests were small to negligible, especially near the middle of the ability distribution. CAT performance depended on the item pool configuration. This study demonstrated that using minimal item exposure rules with a very small pool (40 items), CAT did not perform well. In contrast, a 300-item pool was sufficient to meet the blueprint requirements of a CAT 20-item mathematics test, and a 500-item pool sufficed for a CAT 30-item test. In these conditions, CAT was comparable to MST.

CAT showed small improvements over MST when they were simulated with large item pools with significant overlap (600–700 items). The improvements in measurement precision and ability recovery increased near the tails of the ability distribution among the lowest and highest scoring students. This improvement in precision increased classification accuracy at the lowest and highest cut scores. Although not a focus of this study, this improvement in precision is expected to produce more precise growth measures.

Under the no-growth condition, latent ability distributions for all opportunities were simulated to match those of STAAR. In this case, the reliability of CAT met or exceeded minimal reliability thresholds for low-stakes tests (> 0.80) and most exceeded the threshold (> 0.85) for high-stakes tests; however, the reliability of all tests (CAT and MST) was lowered when Opportunity I and II ability distributions were shifted down to mirror within-year growth. The reliability of CAT and MST Opportunity I dropped below the 0.80 reliability level suggesting that the adaptivity of CAT and MST was hampered by the lack of easier items in the item pool. This suggests that item pools could be improved by increasing the frequency of easier items in Opportunity I and Opportunity II.

Another goal of this study was to understand the tradeoffs of MST versus CAT. TEA field tests sufficient numbers of items each year to support both MST and CAT. MST offers curated forms potentially lowering item development costs, but MST requires complex form construction resulting in lower item utilization rates. CAT selects optimal forms per student eliminating form selection costs and increasing item utilizations; however, annually releasing and replacing the full CAT item pool is costly and problematic.

Table 21 summarizes the tradeoffs of MST versus CAT.

Table 18. Tradeoffs of MST versus CAT

MST	CAT
More efficient than STAAR	More efficient than STAAR
Releasing Opportunity III items *only* would not be a significant cost to TEA.	Releasing an Opportunity III item pool would be a significant cost to TEA; likely a smaller Opportunity I pool would need to be released for viability.
Forms need to be developed in conjunction with simulations in a high-stakes context. Variations in form difficulty from year to year and across opportunities have to be assessed for comparability of outcomes.	Once a pool is baselined, minor adjustments to the pool that maintain the mean and standard deviation of the items deliver comparable outcomes.
Students can be routed to a form in error when the separation between forms is not optimal and their ability may not be recovered.	Does not have the same degree of routing risk because it adapts at the item level.
Allows technology-enhanced items to be validated prior to release as is currently implemented for STAAR.	Requires technology-enhanced items to be validated during field-testing due to low sample sizes.
Opportunity I forms likely need to be adjusted to allow students to score comparably across time at the tails of the distribution.	Provides a better return on investment per item since more field-test items are viable.
Forms can be built to maximize multiple classification decisions.	Provides greater score precision in the tails of the distribution, which will likely provide better growth measures for very low- and very high-scoring students.

Note: MST = multistage tests, CAT = computer-adaptive tests

References

Armstrong, R. D., Jones, D. H., Koppel, N. B., & Pashley, P. J. (2004). Computerized adaptive testing with multiple-form structures. *Applied Psychological Measurement, 28*(3), 147164.

Chen, J., Wood, C., Gianopulos, G. (2023). *Comparing the Psychometric Properties of TTAP Opp3 and STAAR.* Unpublished manuscript.

Dorans, N. J., Pommerich, M., & Holland, P. W. (2007). *Linking and aligning scores and scales*. New York, NY: Springer.

Patsula, L. N. (1999). *A comparison of computerized adaptive testing and multistage testing.* (Doctoral dissertation). University of Massachusetts Amherst, MA.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice,* (28). 5–13. 10.1111/j.1745–3992.2009.00149

Phillips, S. E., & Camara, W. J. (2006). Legal and ethical issues. *Educational measurement*, 733-755.

Rotou, O., Patsula, L., Steffen, M., & Rizavi, S. (2007). Comparison of Multistage Tests with Computerized Adaptive and Paper-and-Pencil Tests. *ETS Research Report Series*, *2007*(1), i-27.

Samejima, F. (1977). A use of the information function in tailored testing. *Applied Psychological Measurement, 1*(3)*,* 233–247.

Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modifications. *Applied Psychological Measurement, 18*(3), 229–244.

Schneider, C., Liu, Y., & Robinson, J. (2022). STAAR Grades 3–8 RLA and Mathematics: An interim-to-summative ROC study [Unpublished manuscript]. Cambium Learning Group.

Zheng, Y., Nozawa, Y., Gao, X., & Chang, H. (2012). Multi-stage adaptive testing for a large-scale classification test: Design, heuristic assembly, and comparison with other testing modes. ACT.

Appendix A: Data Cleaning and Merging Rules

a) TTAP Data Files

The following cleaning rules are applied for the TTAP Database of Record (DOR) data files within each opportunity. Appendix B includes a data dictionary to explain each exclusion variable, possible value, and rule applied for inclusion or exclusion.

- Keep students with appropriate test status values.
  - Using the variable "*status*" include values of "*scored*" and "*completed*".
- Remove students who have not attempted the test.
  - Using the variable "*Overall_Attempted*" keep values of "*Y*".
- Remove private schools.
  - Using "*RTS_REGION_EXTERNALID*" keep values between 1 and 20.
    - Private schools are denoted under a region identifier with a value of 21.
    - Demo schools are listed under region 99.
- Remove students who tested off-grade.
  - e.g., for grade 6 mathematics summaries, keep only students with an "*RTS_EnrlGrdCd*" = 6
- Remove demo students.
  - Using the variable "IsDemo" keep values of 0.
- Separate English and Spanish for grade 5 science.
  - For grade 5 science use the variable "*segment_2_formID"* to determine if the student took an English or Spanish version of the TTAP assessment.
- Within a given grade and subject, if a duplicate "*RTS_EXTERNALID*" occurs, keep the first observation.

b) Summative Data Files

The following cleaning rules are applied for the summative assessments data files:

- Remove private schools.
  - Using "ESCREGIONNUMBER", keep values between 1 and 20.
    - Private schools are denoted under a region identifier with a value of 21.
- For grades 3–8, remove students who tested off-grade.
  - Using "ENROLLEDGRADE" to select valid grade(s).
- Select language.
  - Using "SCIENCELANGUAGEVERSION" to select "E" for English and "S" for Spanish versions for grade 5 science.
- Keep only records with a score code of S.
  - For grades 5–8
    - Using "SCORECODE-MATHEMATICS" of "S" for valid mathematics records.

        - Using "SCORECODE-SOCIALSTUDIES" of “S” for valid social study records.
        - Using "SCORECODE-SCIENCE" of “S” for valid science records.
- Keep only records with respective DISCREPANCYINDICATOR value of 0.
    - Using “DISCREPANCYINDICATORMATHEMATICS” for mathematics.
    - Using “DISCREPANCYINDICATORSCIENCE” for science.
    - Using “DISCREPANCYINDICATORSOCIALSTUDIES” for social studies.
- Remove duplicated records by subject, grade, and student ID; keep the first observation.

Once the TTAP and STAAR data files are cleaned separately, they are merged by student ID. The merged data files have been used to generate the statistics for this report.

Appendix B: Demographic Variable Recode

The table below indicates the values for each demographic variable used in the summaries and how they are recoded for analyses.

Summative Data Variables	Values and Definitions	Recode for Analysis
SEX-CODE	M = Male F = Female	M = Mal F = Female
ETHNICITY/ RACEREPORTING CATEGORY	H = Hispanic/Latino I = American Indian or Alaska Native A = Asian B = Black or African American P = Native Hawaiian or Other Pacific Islander W = White T = Two or More Races N = No Information Provided	H = Hispanic/Latino I = American Indian or Alaska Native A = Asian B = Black or African American P = Native Hawaiian or Other Pacific Islander W = White T = Two or More Races N = No Information Provided
ECONOMIC -DISADVANTAGE -CODE	1 = Eligible for free meals under the National School Lunch and Child Nutrition Program 2 = Eligible for reduced-price meals under the National School Lunch and Child Nutrition Program 9 = Other economic disadvantage 0 = Not identified as economic disadvantaged	1, 2, 9 = Economically Disadvantaged 0 = Otherwise
TITLE-I-PART-A -INDICATOR-CODE	6 = Student attends campus with schoolwide program 7 = Student participates in program at targeted assistance school 8 = Student is previous participant in program at targeted assistance school (not a current participant) 9 = Student does not attend a Title I, Part A school but receives Title I, Part A services because the student is homeless 0 = Student does not currently participate in and has not previously participated in program at current campus	6, 7, 9 = Title-I Part A 0, 8 = Otherwise
MIGRANT -INDICATOR -CODE	1 = Yes 0 = No	1 = Migrant 0 = Otherwise

EMERGENT BILINGUALIN DICATORCODE	C - Identified as Emergent Bilingual (EB)/English learner (EL) F - Monitored 1st Year (M1), reclassified from EB/EL S - Monitored 2nd Year (M2), reclassified from EB/EL T - Monitored 3rd Year (M3), reclassified from EB/EL R - Monitored 4th Year (M4), reclassified from EB/EL E - Former EB/EL (Post Monitoring) 0 - Non-Emergent Bilingual (Non-EB)/Non-English learner (Non-EL)	C = Emergent Bilingual 0, E, F, S, T, R = Otherwise
BILINGUAL-INDICATOR-CODE	2 = Transitional bilingual/early exit 3 = Transitional bilingual/late exit 4 = Dual language immersion/two-way 5 = Dual language immersion/one-way 0 = Student is not participating in a state-approved full bilingual program	2, 3, 4, 5 = Bilingual 0 = Otherwise
ESL-INDICATOR-CODE	2 = ESL/content-based 3 = ESL/pull-out 0 = Student is not participating in a state-approved ESL program	2, 3 = ESL 0 = Otherwise
SPECIAL-ED-INDICATOR-CODE	1 = Student is participating in a special education program 0 = Student is not participating in a special education program	1= Special Ed 0 = Otherwise
GIFTED-TALENTED-INDICATOR-CODE	1 = Yes 0 = No	1 = Gifted and Talented 0 = Otherwise
AT-RISK-INDICATOR-CODE	1 = Yes 0 = No	1 = At Risk 0 = Otherwise

Appendix C: DOR Extract Variable Dictionary

Database of Record (DOR) Extract Variables	Values and Definitions	Rules for Inclusion and Exclusion
Status	Status of the opportunity. Possible values are: completed, submitted, scored, reported, expired, invalidated, and reset.	Keep values of *scored* and *completed.*
Overall_Attempted	Attempted indicates if the student met the attemptedness criteria for the given assessment. Possible values are: *Y* and *N* (some blanks may occur with certain status values).	Keep values of *Y.*
RTS_REGION_EXTERNALID	Numeric identifier (external ID) for the region to which the student belongs. Private schools are denoted with a region identifier of *21* and demo schools are listed under a region identifier of *99*.	Keep values between *1* and *20.*
RTS_EnrlGrdCd	The grade in which a student is registered in the Test Information Distribution Engine (TIDE). Possible values are: EE, PK, KG, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, OS.	For grades 3–8, remove off-grade testers. For EOC remove *OS.*
isDemo	The demo variable indicates if the record is for a demo student or actual student.	Keep values of *0*.