# Power of Predictability: Rigorous and Reliable Assessments

Presented By

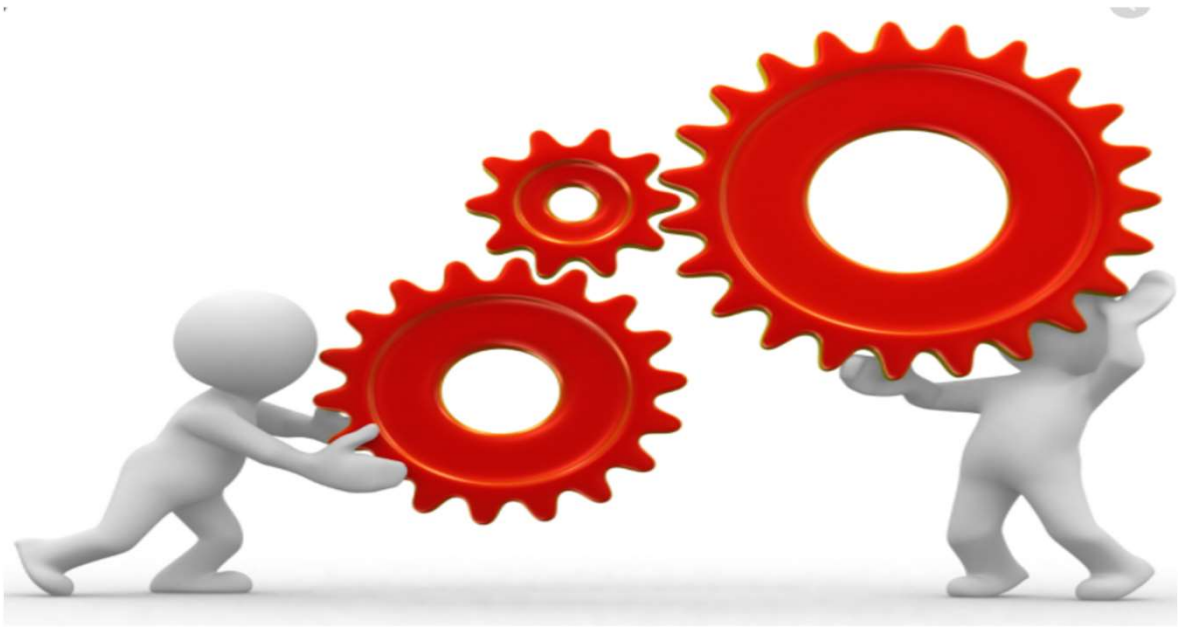
**Rebeca Cooper, Ed.D.**
**Assistant Professor of Education**
**Education Leadership**
**College of Arts, Sciences, and Education**
**Texas A&M University Texarkana**


1

## Learning Intentions

- Define Valid and Reliable Assessments
- Distinguish between Question Difficulty & Question Quality
- Understand the Correlation Alpha Coefficient
- Compare local assessments to STAAR
- Explore a different kind of Item Analysis
- Learn how other disricts are continuously improving assessment systems and processes to improve instruction

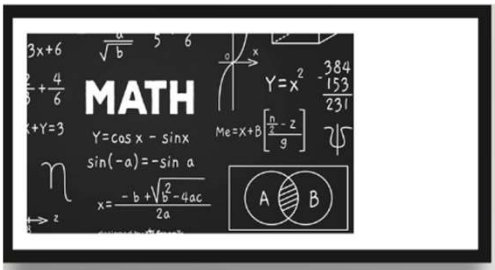
2

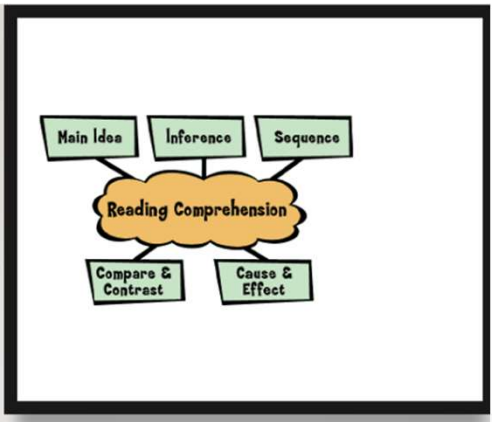


3

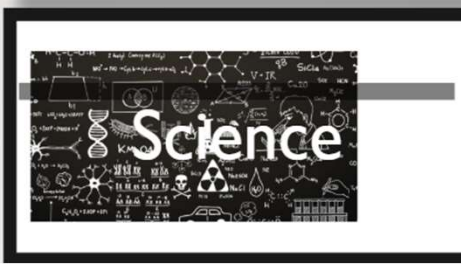
## Assessment

4



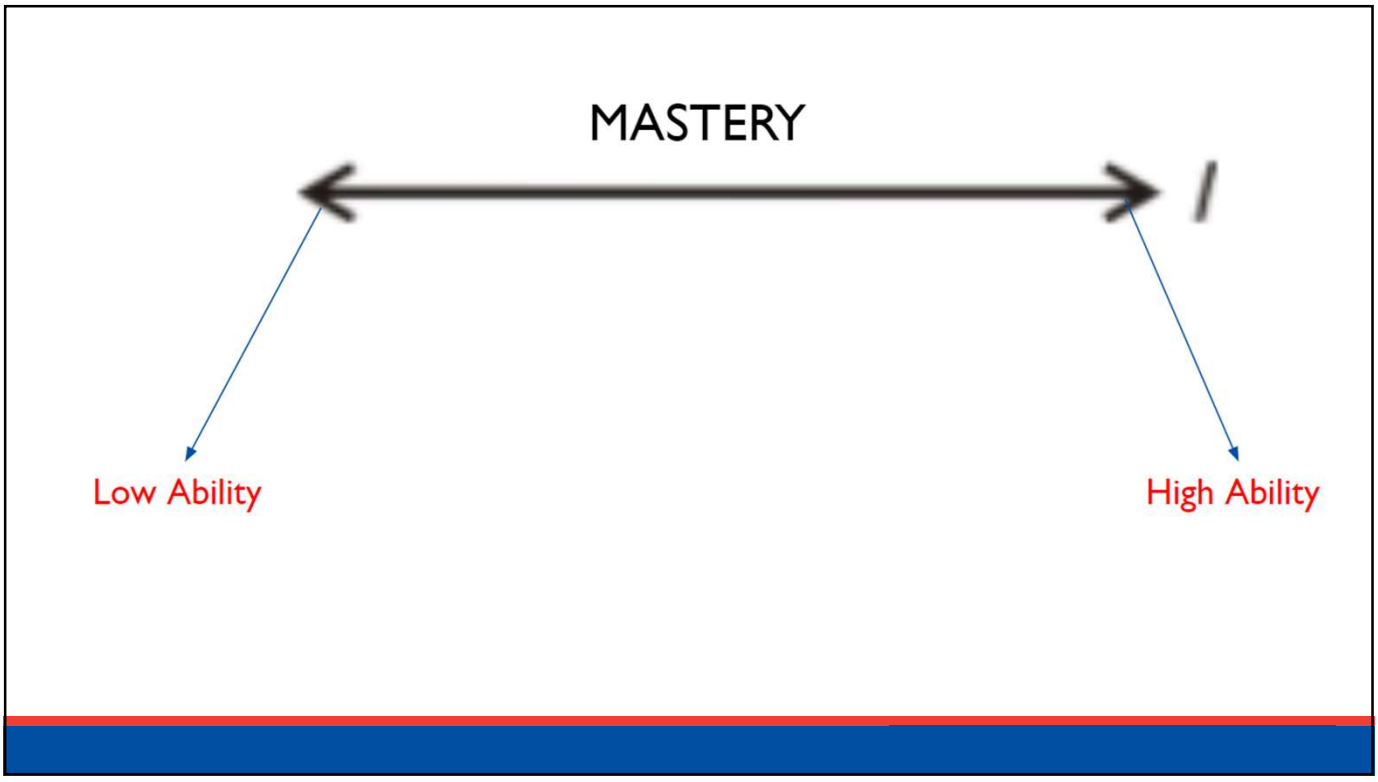





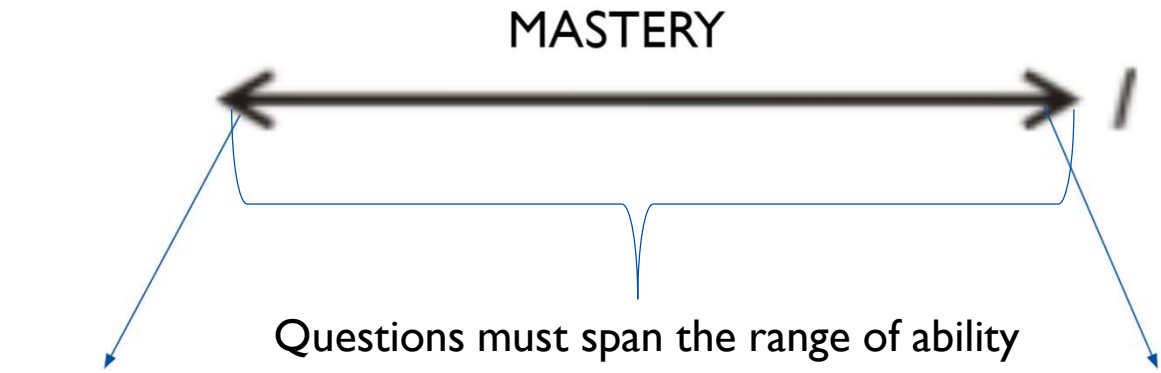



WHAT ARE WE REALLY MEASURING?

5

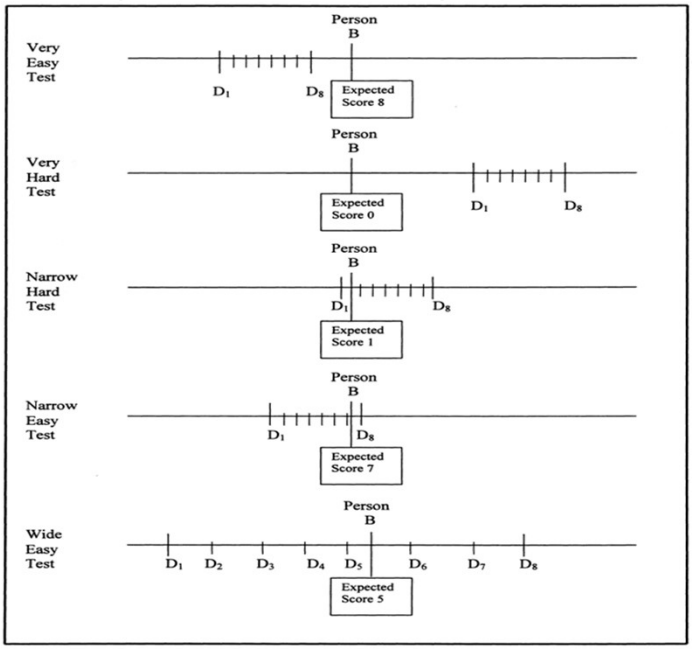



6




7

How scores depend on the level and spread of item difficulty




| Raw Score |
|---|
| 8 |
| 0 |
| 1 |
| 7 |
| 5 |

8

The quality of educational decisions is only as good as the information that leads to them

Student Performance By Demographics

| | # Tested | # Failed | # Passed | % Passing |
|---|---|---|---|---|
| All | 84 | 9 | 75 | 89.29% |
| Female | 41 | 2 | 39 | 95.12% |
| Male | 43 | 7 | 36 | 83.72% |
| Asian | 1 | 0 | 1 | 100% |
| Black or African American | 1 | 1 | 0 | 0% |
| Hispanic/Latino | 27 | 4 | 23 | 85.19% |
| Other | 2 | 1 | 1 | 50% |
| White | 53 | 3 | 50 | 94.34% |

"Almost everyone realizes that 50% on an easy test does not mean as much as making 50% on a hard test. Some even realize that 75% on a narrow test does not imply as much ability as 75% on a WIDE test." (Wright, 1994 p. 5)

9

The quality of educational decisions is only as good as the information that leads to them

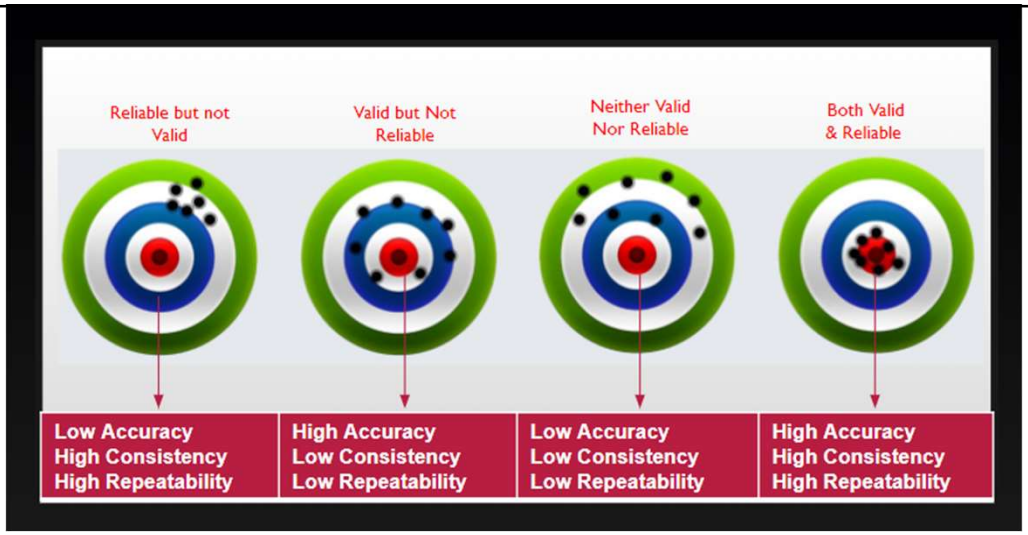
If appropriate information is collected, or if it is collected without precision, the decisions that follow it will be inaccurate.

# Validity & Reliability

OnDataSuite™

10


11

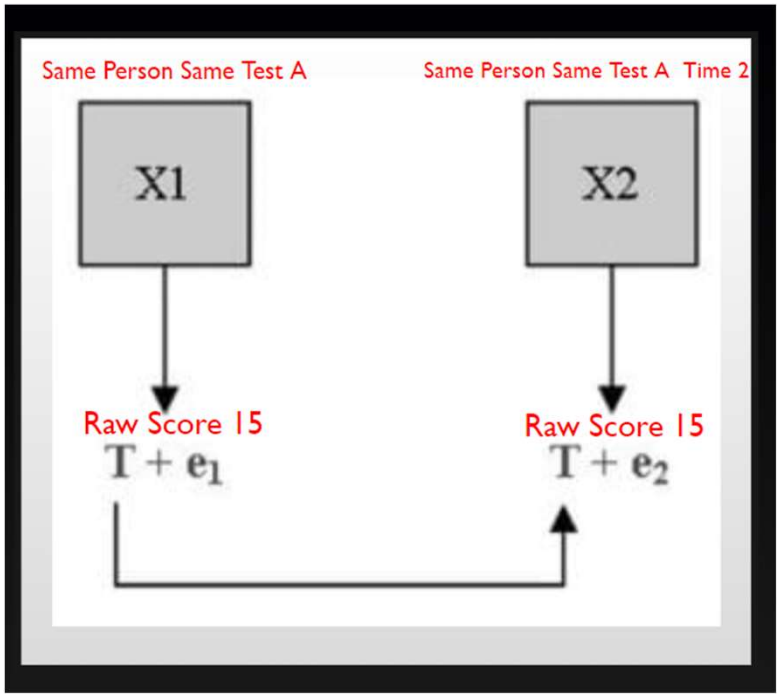



ACCURACY + CONSISTENCY + REPEATABILITY
=
PREDICTABILITY

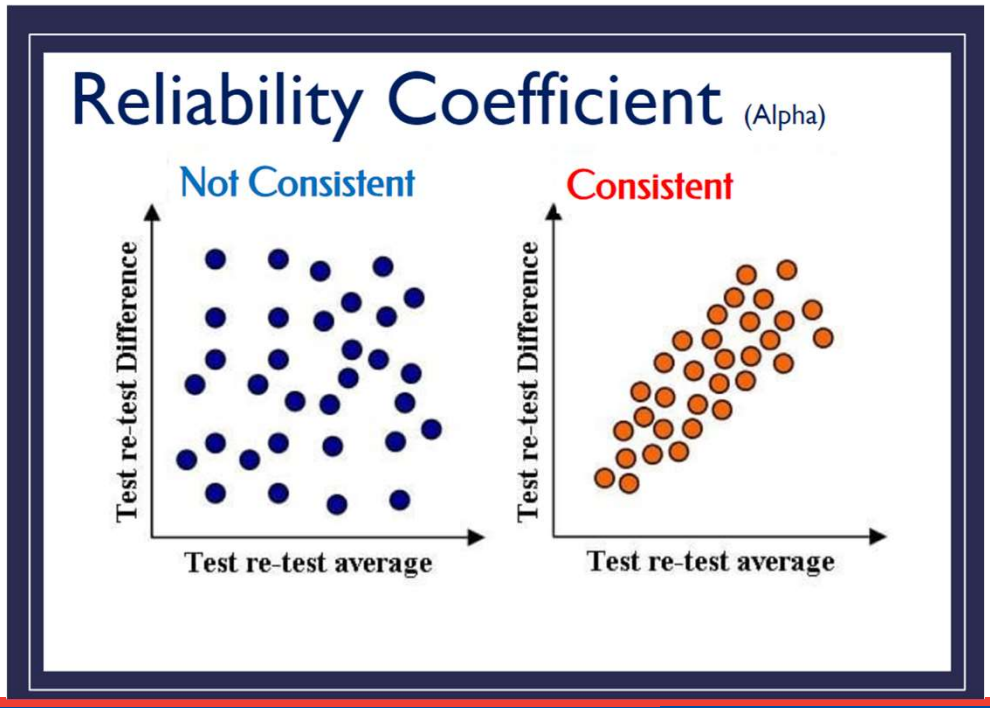
OnDataSuite™

12

SAME SCORE = RELIABILITY

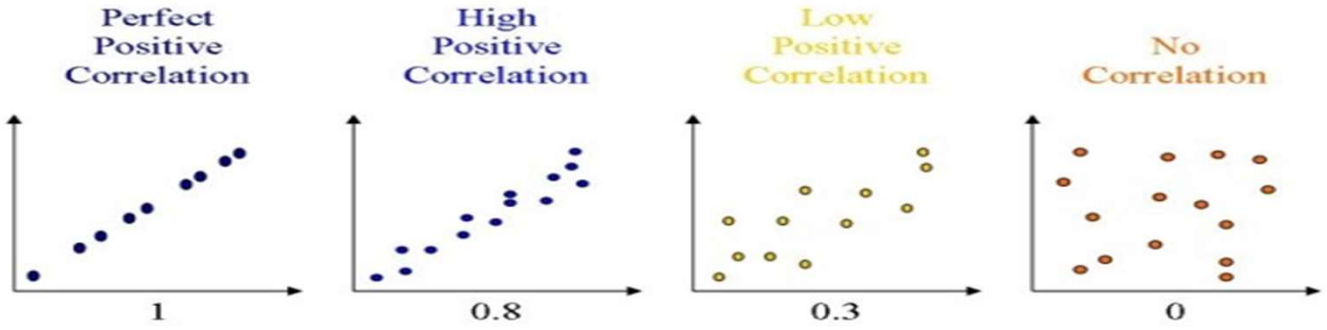



13




14

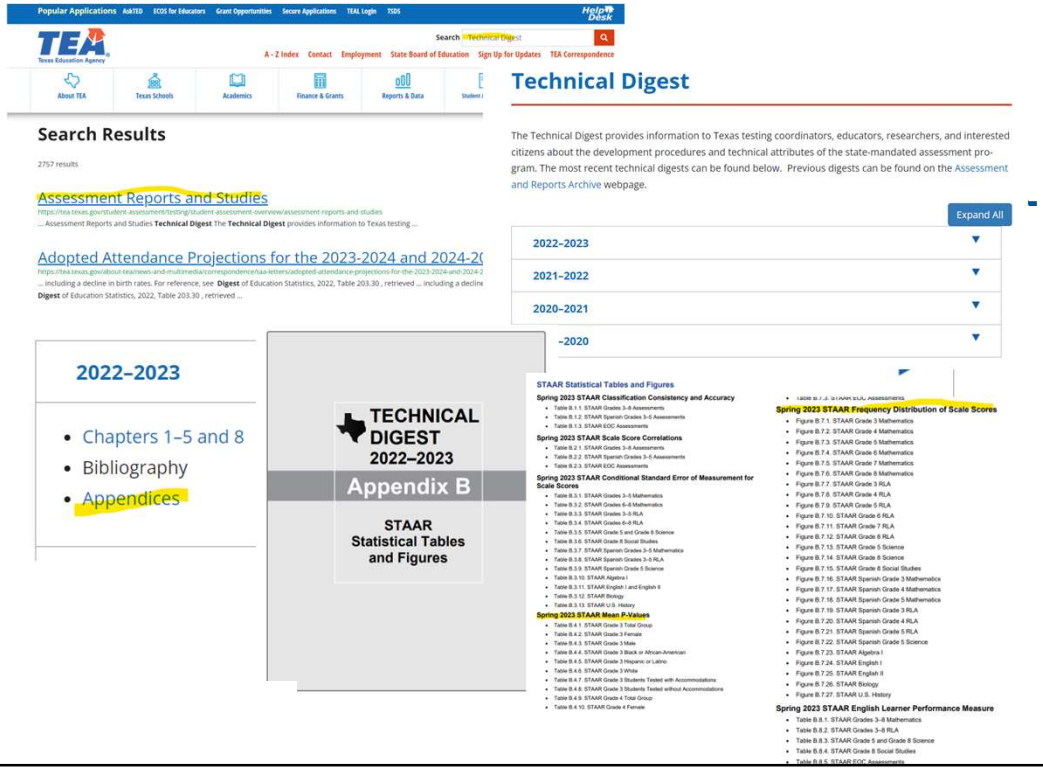
# Reliability Coefficient (Alpha)




- .70 to .79 = Adequate
- .80 to .89 = Good
- .90 and above = Excellent

(TEA, 2019 Technical Digest)

15



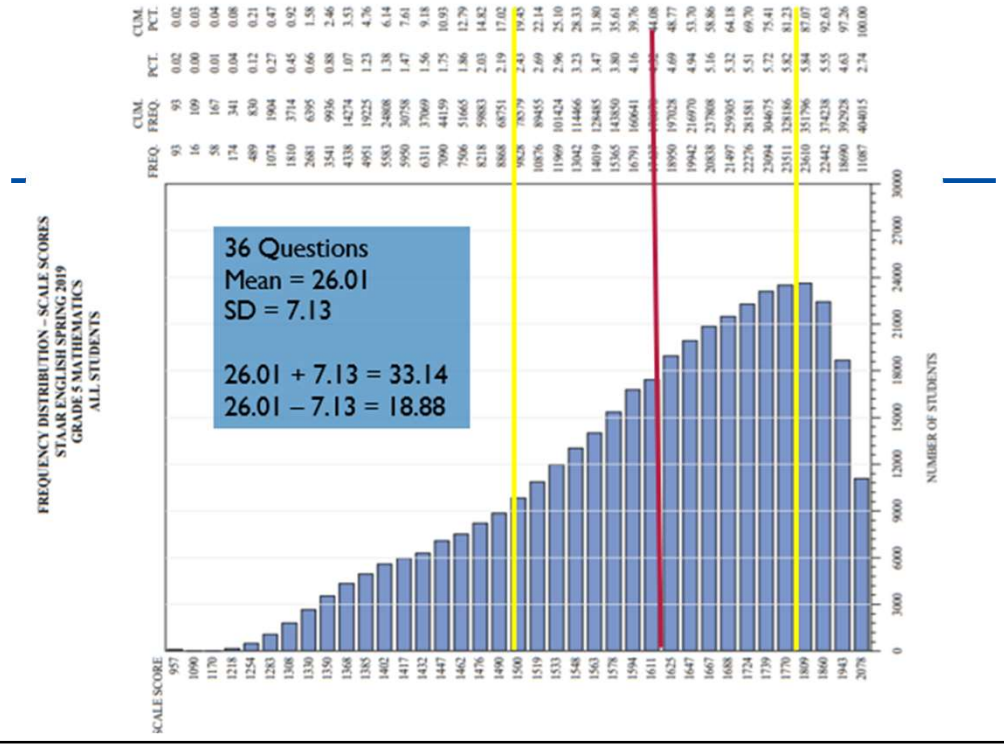

16

**Table B.4.24. Spring 2019 STAAR English Grade 5**
**Students Tested without Accommodations**

| Subject | Reporting Category | Score Point[1] | N | Mean | SD | Alpha[2] | SEM | Mean P-Value[3] |
|---|---|---|---|---|---|---|---|---|
| MATHEMATICS | OVERALL TEST | 36 | 370,981 | 26.01 | 7.13 | 0.90 | 2.26 | 72.25 |
| | Numerical Representations and Relationships | 6 | 370,981 | 4.66 | 1.41 | 0.61 | 0.89 | 77.73 |
| | Computations and Algebraic Relationships | 17 | 370,981 | 12.82 | 3.59 | 0.83 | 1.47 | 75.39 |
| | Geometry and Measurement | 9 | 370,981 | 5.91 | 2.16 | 0.68 | 1.22 | 65.62 |
| | Data Analysis and Personal Financial Literacy | 4 | 370,981 | 2.62 | 1.05 | 0.40 | 0.81 | 65.60 |
| READING | OVERALL TEST | 38 | 372,736 | 27.60 | 7.12 | 0.88 | 2.44 | 72.64 |
| | Understanding/Analysis Across Genres | 8 | 372,736 | 6.48 | 1.61 | 0.62 | 0.99 | 80.99 |
| | Understanding/Analysis of Literary Texts | 16 | 372,736 | 11.22 | 3.39 | 0.77 | 1.64 | 70.12 |
| | Understanding/Analysis of Informational Texts | 14 | 372,736 | 9.90 | 2.93 | 0.74 | 1.50 | 70.74 |
| SCIENCE | OVERALL TEST | 36 | 372,983 | 26.25 | 6.60 | 0.87 | 2.35 | 72.91 |
| | Matter and Energy | 6 | 372,983 | 4.33 | 1.36 | 0.50 | 0.96 | 72.18 |
| | Force, Motion, and Energy | 8 | 372,983 | 5.73 | 1.73 | 0.57 | 1.13 | 71.62 |
| | Earth and Space | 10 | 372,983 | 7.07 | 2.21 | 0.67 | 1.28 | 70.71 |
| | Organisms and Environments | 12 | 372,983 | 9.11 | 2.54 | 0.74 | 1.29 | 75.96 |

*Notes*:
1. Total number of Score Points (May exceed the number of items for tests/reporting categories with essay questions)
2. Stratified Alpha Reliability computed for tests/objectives involving essay questions, KR-20 reliability computed for all others
3. Mean of percent correct (0~100%) for the multiple-choice and gridded items only

17

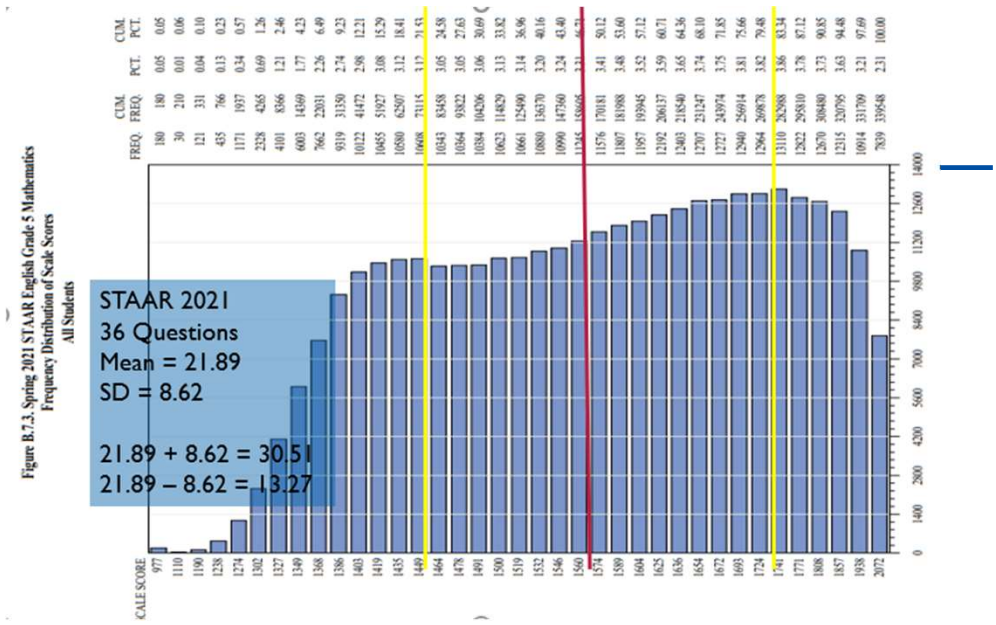



18

**Table B.4.17. Spring 2021 STAAR English Grade 5**
**Total Group**

| Subject | Reporting Category | Score Point[1] | N | Mean | SD | Alpha[2] | SEM | Mean P-Value[3] |
|---|---|---|---|---|---|---|---|---|
| MATHEMATICS | OVERALL TEST | 36 | 339,548 | 21.89 | 8.62 | 0.92 | 2.48 | 60.82 |
| | Numerical Representations and Relationships | 6 | 339,548 | 3.41 | 1.78 | 0.68 | 1.01 | 56.89 |
| | Computations and Algebraic Relationships | 17 | 339,548 | 10.88 | 4.27 | 0.84 | 1.68 | 63.98 |
| | Geometry and Measurement | 9 | 339,548 | 5.17 | 2.32 | 0.70 | 1.26 | 57.44 |
| | Data Analysis and Personal Financial Literacy | 4 | 339,548 | 2.43 | 1.30 | 0.61 | 0.81 | 60.86 |
| READING | OVERALL TEST | 38 | 332,353 | 25.48 | 8.30 | 0.91 | 2.51 | 67.05 |
| | Understanding/Analysis Across Genres | 8 | 332,353 | 5.09 | 2.00 | 0.64 | 1.20 | 63.63 |
| | Understanding/Analysis of Literary Texts | 16 | 332,353 | 10.67 | 3.51 | 0.78 | 1.65 | 66.70 |
| | Understanding/Analysis of Informational Texts | 14 | 332,353 | 9.72 | 3.56 | 0.84 | 1.44 | 69.42 |
| SCIENCE | OVERALL TEST | 36 | 337,483 | 22.61 | 7.27 | 0.88 | 2.57 | 62.81 |
| | Matter and Energy | 6 | 337,483 | 4.00 | 1.51 | 0.55 | 1.02 | 66.61 |
| | Force, Motion, and Energy | 8 | 337,483 | 4.58 | 1.88 | 0.57 | 1.23 | 57.26 |
| | Earth and Space | 10 | 337,483 | 6.50 | 2.35 | 0.67 | 1.36 | 65.00 |
| | Organisms and Environments | 12 | 337,483 | 7.53 | 2.86 | 0.74 | 1.47 | 62.78 |

19



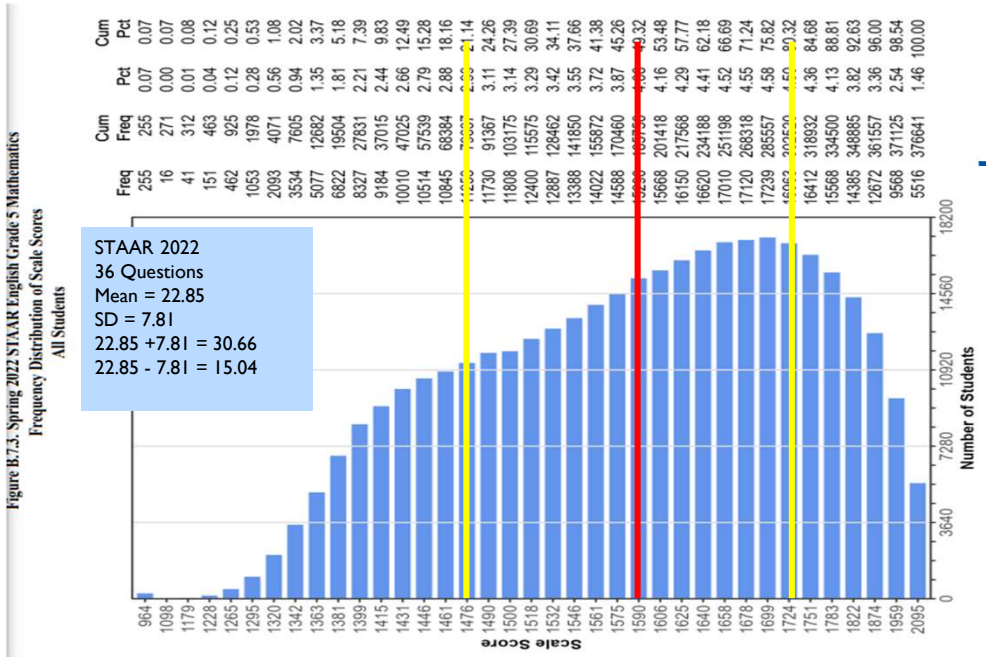

20

**Table B.4.17. Spring 2022 STAAR English Grade 5 Total Group**

| Subject | Reporting Category | Score Point[1] | N | Mean | SD | Alpha[2] | SEM | Mean P-Value[3] |
|---|---|---|---|---|---|---|---|---|
| MATHEMATICS | OVERALL TEST | 36 | 376,641 | 22.85 | 7.81 | 0.90 | 2.46 | 63.47 |
| | Numerical Representations and Relationships | 6 | 376,641 | 4.10 | 1.55 | 0.61 | 0.97 | 68.26 |
| | Computations and Algebraic Relationships | 17 | 376,641 | 10.58 | 3.93 | 0.82 | 1.69 | 62.21 |
| | Geometry and Measurement | 9 | 376,641 | 5.87 | 2.28 | 0.69 | 1.26 | 65.26 |
| | Data Analysis and Personal Financial Literacy | 4 | 376,641 | 2.30 | 1.12 | 0.46 | 0.82 | 57.57 |
| READING | OVERALL TEST | 38 | 369,356 | 27.87 | 7.86 | 0.91 | 2.35 | 73.34 |
| | Understanding/Analysis Across Genres | 8 | 369,356 | 5.82 | 2.01 | 0.71 | 1.08 | 72.7 |
| | Understanding/Analysis of Literary Texts | 16 | 369,356 | 11.68 | 3.41 | 0.80 | 1.52 | 72.99 |
| | Understanding/Analysis of Informational Texts | 14 | 369,356 | 10.37 | 3.15 | 0.80 | 1.42 | 74.1 |
| SCIENCE | OVERALL TEST | 36 | 376,142 | 23.61 | 7.53 | 0.89 | 2.52 | 65.57 |
| | Matter and Energy | 6 | 376,142 | 4.06 | 1.63 | 0.63 | 0.99 | 67.59 |
| | Force, Motion, and Energy | 8 | 376,142 | 5.23 | 1.80 | 0.55 | 1.20 | 65.34 |
| | Earth and Space | 10 | 376,142 | 6.17 | 2.41 | 0.68 | 1.37 | 61.67 |
| | Organisms and Environments | 12 | 376,142 | 8.16 | 2.98 | 0.78 | 1.40 | 67.97 |

*Notes:*
1. Total number of Score Points (may exceed the number of items for tests/reporting categories with essay questions).
2. Stratified Alpha Reliability computed for tests involving essay questions, KR-20 reliability computed for all others.
3. Mean of percent correct (0~100%) for the multiple-choice and gridded items only.

21

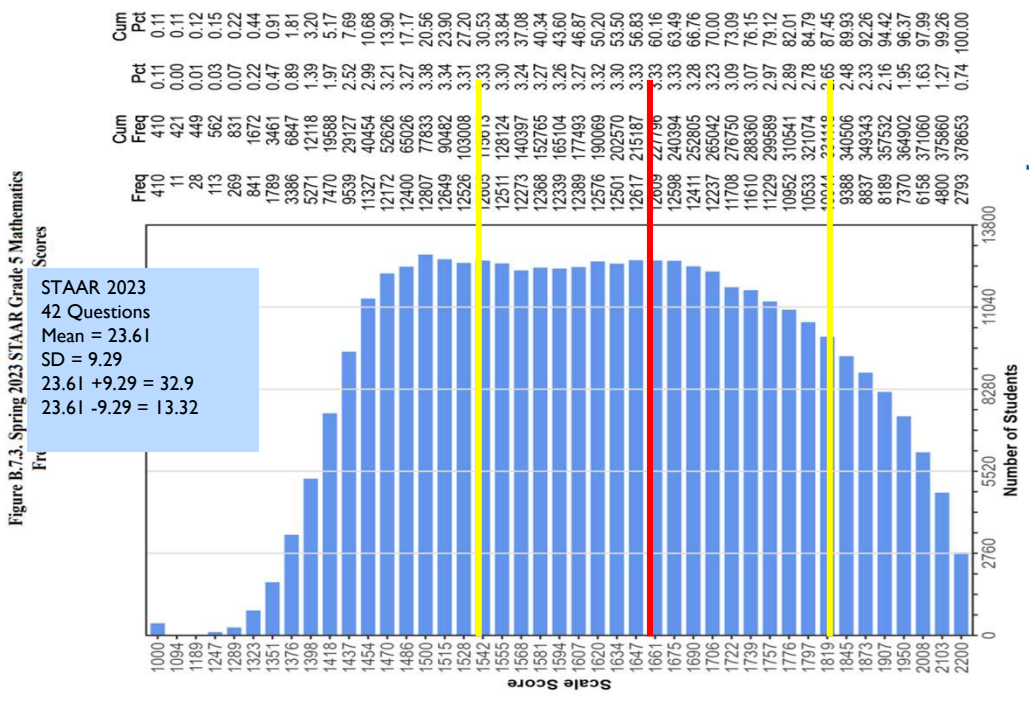



22

**Table B.4.17. Spring 2023 STAAR Grade 5 Total Group**

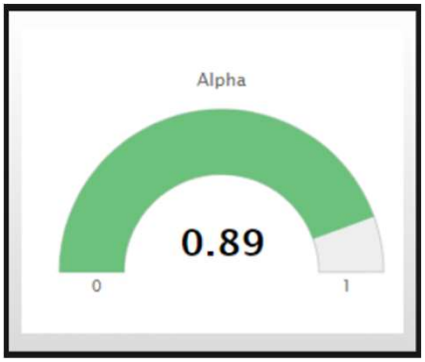
| Subject | Reporting Category | Score Point[1] | N | Mean | SD | Alpha[2] | SEM | Mean P-Value[3] |
|---|---|---|---|---|---|---|---|---|
| Mathematics | OVERALL TEST | 42 | 378,653 | 23.61 | 9.29 | 0.90 | 2.88 | 56.18 |
| | Numerical Representations and Relationships | 9 | 378,653 | 5.62 | 2.11 | 0.64 | 1.26 | 58.74 |
| | Computations and Algebraic Relationships | 20 | 378,653 | 11.18 | 4.93 | 0.84 | 2.00 | 58.04 |
| | Geometry and Measurement | 9 | 378,653 | 4.40 | 2.35 | 0.63 | 1.43 | 48.48 |
| | Data Analysis and Personal Financial Literacy | 4 | 378,653 | 2.40 | 1.15 | 0.45 | 0.85 | 60.11 |
| RLA | OVERALL TEST | 52 | 372,636 | 30.96 | 10.49 | 0.92 | 2.98 | 64.39 |
| | Reading | 26 | 372,636 | 16.77 | 5.01 | 0.82 | 2.11 | 65.57 |
| | Writing | 26 | 372,636 | 14.19 | 6.14 | 0.89 | 2.08 | 62.71 |
| Science | OVERALL TEST | 39 | 378,696 | 20.93 | 7.48 | 0.87 | 2.74 | 52.98 |
| | Matter and Energy | 6 | 378,696 | 3.16 | 1.46 | 0.48 | 1.05 | 51.28 |
| | Force, Motion, and Energy | 9 | 378,696 | 4.64 | 2.32 | 0.68 | 1.32 | 53.34 |
| | Earth and Space | 11 | 378,696 | 4.88 | 2.39 | 0.56 | 1.59 | 45.89 |
| | Organisms and Environments | 13 | 378,696 | 8.25 | 2.80 | 0.73 | 1.47 | 59.34 |

23




24

# Reliability Coefficient (Alpha)


- .70 to .79 = Adequate
- .80 to .89 = Good
- .90 and above = Excellent

"As a general rule, reliability coefficients from 0.70 to 0.79 are considered adequate, those from 0.80 to 0.89 are considered good, and those at 0.90 or above are considered excellent."

*— TEA Technical Digest 2018 - 2019*

25


26

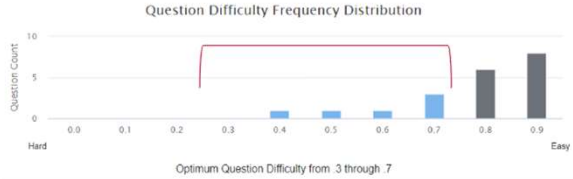
Question Difficulty

Question Quality

27

| Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 |
|---|---|---|---|---|---|---|
| 0.79 | 0.86 | 0.99 | 0.58 | 0.42 | 0.92 | 0.69 |
| Q8 | Q9 | Q10 | Q11 | Q12 | Q13 | Q14 |
| 0.93 | 0.82 | 0.9 | 0.93 | 0.89 | 0.86 | 0.88 |
| Q15 | Q16 | Q17 | Q18 | Q19 | Q20 | |
| 0.88 | 0.94 | 0.95 | 0.95 | 0.79 | 0.75 | |

Optimum Question Difficulty from .3 through .7




# QUESTION DIFFICULTY

p-value = The probability of the correct response to the question.

Avoid extremely easy or difficult items such that classical item difficulty (p-value) is within a range of 0.20-0.90.

| Subject | Reporting Category | Score Point[1] | N | Mean | SD | Alpha[2] | SEM | Mean P-Value[3] |
|---|---|---|---|---|---|---|---|---|
| ALGEBRA I | OVERALL TEST | 54 | 396,086 | 29.33 | 12.46 | 0.94 | 3.16 | 54.31 |

Texas Education Agency, (2013). State of Texas Assessments of Academic Readiness (STAAR™): Vertical Scale Technical Report. Austin, Texas. Texas Education Agency Retrieved from:

28

# QUESTION DIFFICULTY

p-value = The probability of the correct response to the question.

| Grade | | 2019 | | 2021 | 2022 | | 2023 |
|---|---|---|---|---|---|---|---|
| 3 | Math | 68.68 | | 57.19 | 61.58 | | 48.34 |
| | Reading | 67.53 | | 62.22 | 67.90 | | 59.79 |
| 4 | Math | 64.44 | | 56.44 | 60.79 | | 52.60 |
| | Reading | 66.13 | | 61.70 | 69.93 | | 54.93 |
| 5 | Math | 70.34 | | 60.82 | 63.47 | | 56.18 |
| | Reading | 71.27 | | 67.05 | 73.34 | | 64.39 |
| | Science | 71.36 | | 62.81 | 65.57 | | 52.98 |
| 6 | Math | 56.19 | | 50.87 | 51.74 | | 49.39 |
| | Reading | 64.50 | | 62.27 | 64.36 | | 54.11 |

COVID

STAAR REDESIGN

29

| Grade | | 2019 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| 7 | Math | 53.53 | 46.19 | 48.61 | 47.96 |
| | Reading | 66.91 | 65.78 | 69.72 | 60.57 |
| 8 | Math | 63.54 | 52.85 | 57.61 | 49.96 |
| | Reading | 70.15 | 66.80 | 71.18 | 55.64 |
| | Science | 67.13 | 61.31 | 64.29 | 51.92 |
| | Social Studies | 59.23 | 54.31 | 55.62 | 49.91 |
| EOC | Algebra I | 64.96 | 54.31 | 59.21 | 52.10 |
| | ENG I | NA | | | 58.98 |
| | ENG II | NA | | | 59.02 |
| | BIO | 64.98 | 62.31 | 61.72 | 52.70 |
| | US History | 72.33 | 69.61 | 70.08 | 57.44 |

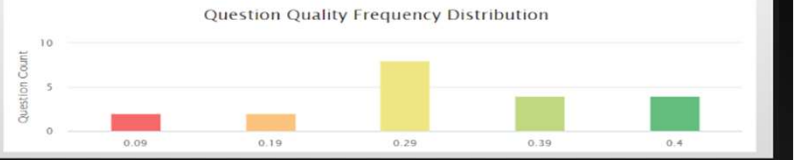
30

# QUESTION QUALITY

Point Biserial Correlation

Point Biserial Correlations should be greater than or equal to 0.20.

| Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 |
|---|---|---|---|---|---|---|
| 0.33 | 0.27 | 0.03 | 0.44 | 0.07 | 0.53 | 0.12 |
| Q8 | Q9 | Q10 | Q11 | Q12 | Q13 | Q14 |
| 0.19 | 0.3 | 0.21 | 0.25 | 0.34 | 0.23 | 0.22 |
| Q15 | Q16 | Q17 | Q18 | Q19 | Q20 | |
| 0.29 | 0.27 | 0.3 | 0.48 | 0.47 | 0.23 | |




.4 and above — Very good Questions
.3 to .39 — Good Questions
.2 to .29 — Fairly Good Questions
.1 to .19 — Marginal Questions
.09 or below — Poor Questions

The point biserial may be weak on items with very high or very low p-values. If all students get the item correct (or conversely incorrect), this item doesn't provide enough useful information to help distinguish between students with higher performance and students with lower performance on the entire test.

31

# Item Analysis

| | | Pupil Progression | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Target | Binary inc. addition, 2s complement | Hexadecimal | Octal | Parts of a Processor | Special Registers | Buses | Memory | Input / Output Devices | Different Computer Architectures | OVERALL | Grade | Difference |
| | | 77% | 80% | 70% | 39% | 50% | 36% | 59% | 68% | 95% | 63% | | |
| | c | 44% | 50% | 50% | 33% | 75% | 83% | 75% | 100% | 100% | 64% | C | 0 |
| | c | 78% | 100% | 100% | 67% | 50% | 0% | 75% | 100% | 100% | 72% | B | 1 |
| | c | 33% | 75% | 100% | 33% | 75% | 0% | 75% | 100% | 100% | 60% | C | 0 |
| | c | 100% | 100% | 100% | 33% | 0% | 0% | 0% | 0% | 100% | 51% | D | -1 |
| | c | 100% | 88% | 75% | 17% | 50% | 50% | 75% | 100% | 75% | 70% | C | 0 |
| | c | 100% | 100% | 100% | 67% | 88% | 0% | 75% | 100% | 100% | 81% | A | 2 |
| | c | 67% | 63% | 25% | 17% | 50% | 0% | 75% | 50% | 100% | 49% | E | -2 |
| | c | 100% | 88% | 75% | 50% | 50% | 33% | 25% | 100% | 100% | 70% | C | 0 |
| | c | 67% | 63% | 50% | 67% | 38% | 67% | 75% | 0% | 100% | 58% | D | -1 |
| | c | 100% | 75% | 75% | 17% | 50% | 67% | 50% | 0% | 100% | 62% | C | 0 |
| | c | 56% | 75% | 25% | 33% | 25% | 100% | 50% | 100% | 75% | 58% | D | -1 |

OnDataSuite™

32

# QUESTION REVIEW

| Question 3 | | |
|---|---|---|
| Question Difficulty | 0.99 | Outside of Optimum Range |
| Question Quality | 0.03 | Poor |

| Question 5 | | |
|---|---|---|
| Question Difficulty | 0.42 | Within Optimum Range |
| Question Quality | 0.07 | Poor |

| Question 7 | | |
|---|---|---|
| Question Difficulty | 0.69 | Within Optimum Range |
| Question Quality | 0.12 | Marginal |



Texas Education Agency

## Technical Digest

The Technical Digest provides information to Texas testing coordinators, educators, researchers, and interested citizens about the development procedures and technical attributes of the state-mandated assessment program. The most recent technical digests can be found below. Previous digests can be found on the Assessment and Reports Archive webpage.

Expand All

**2022–2023**

- Chapters 1–6 and 8
- Bibliography
- Appendices

**2021–2022**

**2020–2021**

**2019–2020**

Technical Digest 2022–2023

Test Development Activities
Groups Involved
Item Development and Review
Pilot Testing
Field Testing and Data Review
Security
Quality-Control Procedures
Performance Assessments

**Table 2.2.** Item and Passage Review Guidelines

| Category | Guidelines |
|---|---|
| Reporting Category/Student Expectation Item Match | • The item measures what it is supposed to assess.<br>• The item poses a clearly defined problem or task. |
| Appropriateness (Interest Level) | • The item or passage is well written and clear.<br>• The point of view is relevant to students taking the test.<br>• The subject matter is of fairly wide interest to students at the grade being tested.<br>• The artwork is clear, correct, and appropriate. |
| Appropriateness (Format) | • The format is appropriate for the intended grade.<br>• The format is interesting to the student.<br>• The item is formatted so it is not unnecessarily difficult. |
| Appropriateness (Answer Choices) | • The answer choices are reasonably parallel in structure.<br>• The answer choices are worded clearly and concisely.<br>• The answer choices do not eliminate each other.<br>• There is only one correct answer. |
| Appropriateness (Difficulty of Distractors) | • Each distractor is plausible.<br>• There is a rationale for each distractor.<br>• Each distractor is relevant to the knowledge and understanding being measured.<br>• Each distractor is at a difficulty level appropriate for both the objective and the intended grade. |
| Opportunity to Learn | • The item is a good measure of the curriculum.<br>• The item is suitable for the grade or course. |

## Technical Digest

The Technical Digest provides information to Texas testing coordinators, educators, researchers, and interested citizens about the development procedures and technical attributes of the state-mandated assessment program. The most recent technical digests can be found below. Previous digests can be found on the Assessment and Reports Archive webpage.

Expand All

**2022–2023**

- Chapters 1–6 and 8
- Bibliography
- Appendices

**2021–2022**

**2020–2021**

**2019–2020**

Technical Digest 2022–2023

Test Development Activities
Groups Involved
Item Development and Review
Pilot Testing
Field Testing and Data Review
Security
Quality-Control Procedures
Performance Assessments

Technical Digest 2022–2023

| Category | Guidelines |
|---|---|
| Sensitivity Concerns and Freedom from Bias | • The item or passage does not assume racial, class, or gender values or suggest such stereotypes.<br>• The item does not provide an advantage or disadvantage to any group of students because of their personal characteristics, such as race, gender, socioeconomic status, or religion.<br>• The item or passage avoids needless reference to topics that are extremely controversial or upsetting.<br>• The item or passage addresses sensitive topics in a careful, fair, and balanced way.<br>• The item fairly represents cultural, ethnic, social, and political diversity. |

35

District Name: Learning ISD
District ID: 999001

Show Statistical Details | Tools

### Assessment Analysis

| | | | |
|---|---|---|---|
| Test Name | test description | Number of Questions | 20 |
| Date Given | November 10, 2021 | Number of Students | 330 |
| Subject Area | ELA | Date Data Compiled | April 30, 2024 |
| Grade Level | 8 | Teacher | teacher first name teacher last name |

Test Rating


Student Performance By Demographics

| | # Tested | # Failed | # Passed | % Passing |
|---|---|---|---|---|
| All | 213 | 17 | 196 | 92.02% |
| Female | 112 | 6 | 106 | 94.64% |
| Male | 101 | 11 | 90 | 89.11% |
| Asian | 19 | 0 | 19 | 100% |
| Black or African American | 5 | 1 | 4 | 80% |
| Hispanic/Latino | 83 | 11 | 72 | 86.75% |
| Two or More Races | 4 | 0 | 4 | 100% |
| White | 102 | 5 | 97 | 95.1% |

36

## Statistical Details

| | State STAAR Test | My Test |
|---|---|---|
| **Total Questions** | 56 | 20 |
| **Reliability Coefficient (Alpha)** | 0.93 | 0.66 |
| **Average Raw Score** | 30.86 | 14.36 |
| **Standard Deviation** | 11.62 | 2.59 |
| **Mean P-Value** | 55.64 | 73.73 |

STAAR Source: TEA, My Test Source: OnTarget

"As a general rule, reliability coefficients from 0.70 to 0.79 are considered adequate, those from 0.80 to 0.89 are considered good, and those at 0.90 or above are considered excellent."
*— TEA Technical Digest 2018 - 2019*

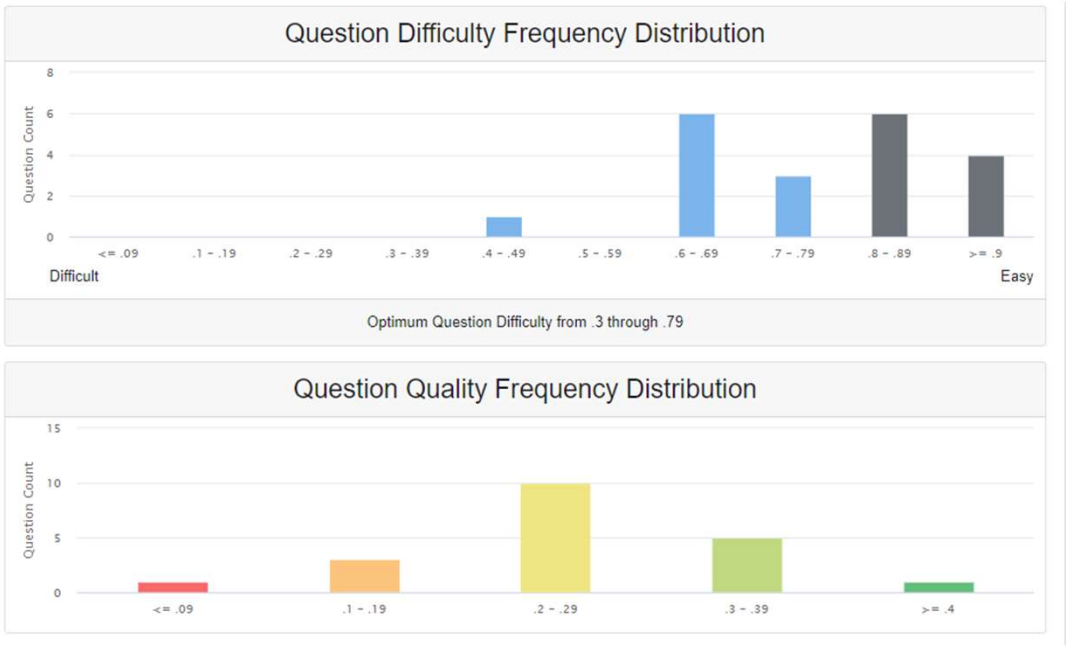
37

# Item Analysis

| | Target | Pupil Progression | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Binary inc. addition, 2s complement | Hexadecimal | Octal | Parts of a Processor | Special Registers | Buses | Memory | Input / Output Devices | Different Computer Architectures | OVERALL | Grade | Difference |
| | | 77% | 80% | 70% | 39% | 50% | 36% | 59% | 68% | 95% | 63% | | |
| | c | 44% | 50% | 50% | 33% | 75% | 83% | 75% | 100% | 100% | 64% | C | 0 |
| | c | 78% | 100% | 100% | 67% | 50% | 0% | 75% | 100% | 100% | 72% | B | 1 |
| | c | 33% | 75% | 100% | 33% | 75% | 0% | 75% | 100% | 100% | 60% | C | 0 |
| | c | 100% | 100% | 100% | 33% | 0% | 0% | 0% | 0% | 100% | 51% | D | -1 |
| | c | 100% | 88% | 75% | 17% | 50% | 50% | 75% | 100% | 75% | 70% | C | 0 |
| | c | 100% | 100% | 100% | 67% | 88% | 0% | 75% | 100% | 100% | 81% | A | 2 |
| | c | 67% | 63% | 25% | 17% | 50% | 0% | 75% | 50% | 100% | 49% | E | -2 |
| | c | 100% | 88% | 75% | 50% | 50% | 33% | 25% | 100% | 100% | 70% | C | 0 |
| | c | 67% | 63% | 50% | 67% | 38% | 67% | 75% | 0% | 100% | 58% | D | -1 |
| | c | 100% | 75% | 75% | 17% | 50% | 67% | 50% | 0% | 100% | 62% | C | 0 |
| | c | 56% | 75% | 25% | 33% | 25% | 100% | 50% | 100% | 75% | 58% | D | -1 |

OnDataSuite™

38




39

## Item Analysis

| Question | Difficulty | Quality | Analyzed |
|---|---|---|---|
| 1 | 0.62 | Marginal (0.14) | × |
| 2 | 0.93 | Marginal (0.18) | × |
| 3 | 0.94 | Fairly Good (0.22) | × |
| 4 | 0.83 | Marginal (0.14) | × |
| 5 | 0.78 | Fairly Good (0.2) | × |
| 6 | 0.87 | Fairly Good (0.29) | × |
| 7 | 0.88 | Fairly Good (0.21) | × |
| 8 | 0.95 | Good (0.35) | × |
| 9 | 0.61 | Fairly Good (0.21) | × |
| 10 | 0.82 | Fairly Good (0.29) | × |
| 11 | 0.66 | Fairly Good (0.2) | × |
| 12 | 0.63 | Fairly Good (0.27) | × |
| 13 | 0.69 | Good (0.35) | × |
| 14 | 0.79 | Good (0.32) | × |
| 15 | 0.75 | Very Good (0.46) | × |
| 16 | 0.88 | Good (0.35) | × |
| 17 | 0.85 | Fairly Good (0.29) | × |
| 18 | 0.92 | Good (0.37) | × |
| 19 | 0.41 | Poor (0.07) | × |
| 20 | 0.64 | Fairly Good (0.2) | × |

Optimum Question Difficulty from .3 through .79

40

Texas Education Agency

## Technical Digest

The Technical Digest provides information to Texas testing coordinators, educators, researchers, and interested citizens about the development procedures and technical attributes of the state-mandated assessment program. The most recent technical digests can be found below. Previous digests can be found on the Assessment and Reports Archive webpage.

Expand All

**2022–2023**

- Chapters 1–6 and 8
- Bibliography
- Appendices

**2021–2022**

**2020–2021**

**2019–2020**

Technical Digest 2022–2023

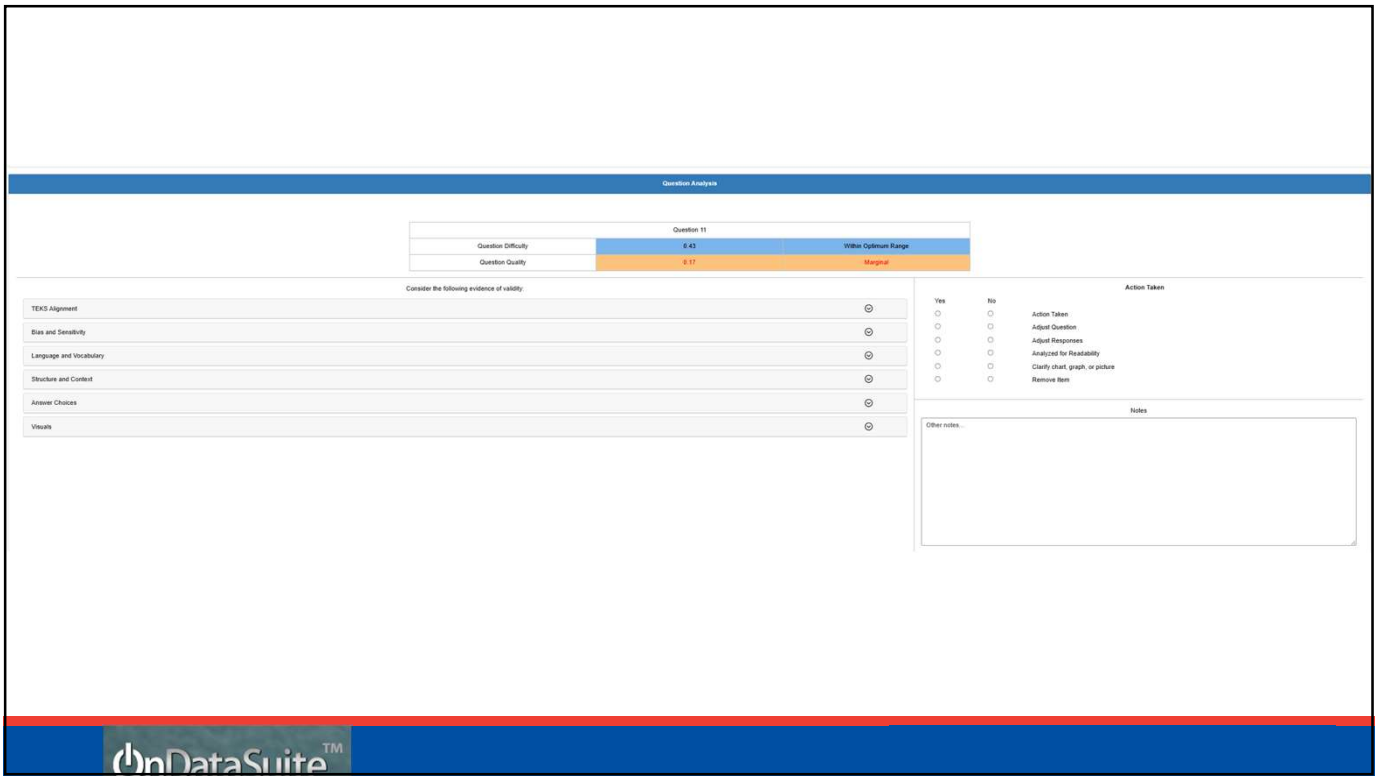


**Table 2.2.** Item and Passage Review Guidelines

| Category | Guidelines |
|---|---|
| Reporting Category/Student Expectation Item Match | • The item measures what it is supposed to assess.<br>• The item poses a clearly defined problem or task. |
| Appropriateness (Interest Level) | • The item or passage is well written and clear.<br>• The point of view is relevant to students taking the test.<br>• The subject matter is of fairly wide interest to students at the grade being tested.<br>• The artwork is clear, correct, and appropriate. |
| Appropriateness (Format) | • The format is appropriate for the intended grade.<br>• The format is interesting to the student.<br>• The item is formatted so it is not unnecessarily difficult. |
| Appropriateness (Answer Choices) | • The answer choices are reasonably parallel in structure.<br>• The answer choices are worded clearly and concisely.<br>• The answer choices do not eliminate each other.<br>• There is only one correct answer. |
| Appropriateness (Difficulty of Distractors) | • Each distractor is plausible.<br>• There is a rationale for each distractor.<br>• Each distractor is relevant to the knowledge and understanding being measured.<br>• Each distractor is at a difficulty level appropriate for both the objective and the intended grade. |
| Opportunity to Learn | • The item is a good measure of the curriculum.<br>• The item is suitable for the grade or course. |

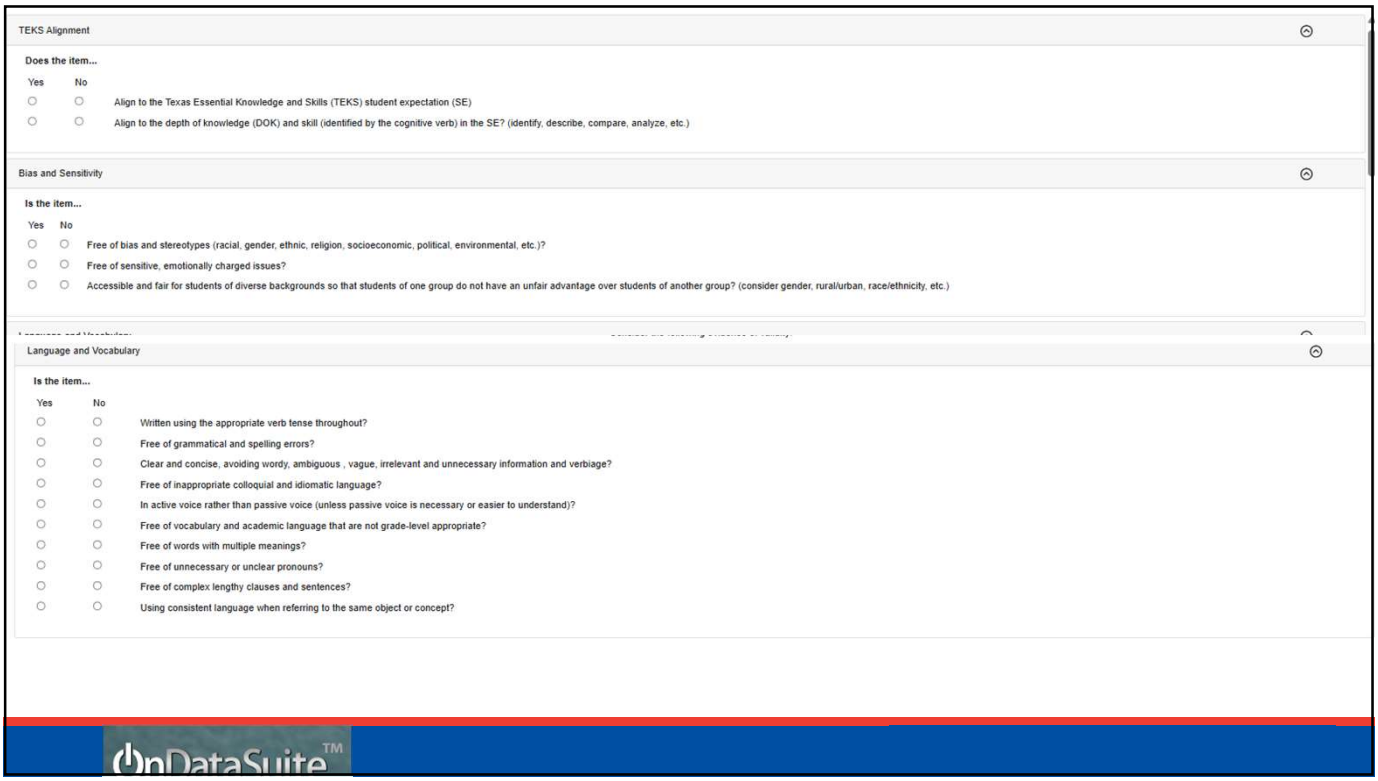
OnDataSuite™

41

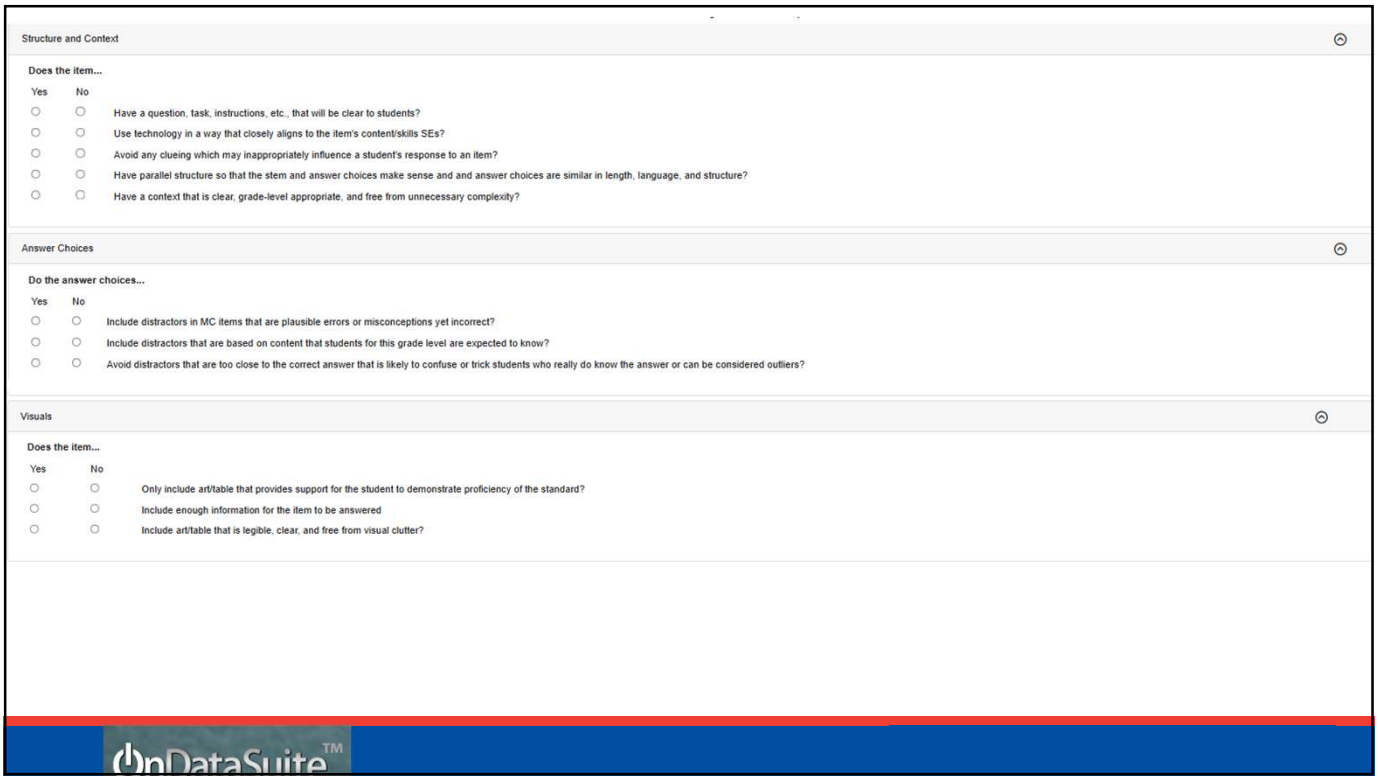



OnDataSuite™

42


TEKS Alignment
Does the item...
Yes
No
Align to the Texas Essential Knowledge and Skills (TEKS) student expectation (SE)
Align to the depth of knowledge (DOK) and skill (identified by the cognitive verb) in the SE? (identify, describe, compare, analyze, etc.)
Bias and Sensitivity
Is the item...
Yes
No
Free of bias and stereotypes (racial, gender, ethnic, religion, socioeconomic, political, environmental, etc.)?
Free of sensitive, emotionally charged issues?
Accessible and fair for students of diverse backgrounds so that students of one group do not have an unfair advantage over students of another group? (consider gender, rural/urban, race/ethnicity, etc.)
Language and Vocabulary
Is the item...
Yes
No
Written using the appropriate verb tense throughout?
Free of grammatical and spelling errors?
Clear and concise, avoiding wordy, ambiguous , vague, irrelevant and unnecessary information and verbiage?
Free of inappropriate colloquial and idiomatic language?
In active voice rather than passive voice (unless passive voice is necessary or easier to understand)?
Free of vocabulary and academic language that are not grade-level appropriate?
Free of words with multiple meanings?
Free of unnecessary or unclear pronouns?
Free of complex lengthy clauses and sentences?
Using consistent language when referring to the same object or concept?
OnDataSuite™

43


Structure and Context
Does the item...
Yes
No
Have a question, task, instructions, etc., that will be clear to students?
Use technology in a way that closely aligns to the item's content/skills SEs?
Avoid any clueing which may inappropriately influence a student's response to an item?
Have parallel structure so that the stem and answer choices make sense and and answer choices are similar in length, language, and structure?
Have a context that is clear, grade-level appropriate, and free from unnecessary complexity?
Answer Choices
Do the answer choices...
Yes
No
Include distractors in MC items that are plausible errors or misconceptions yet incorrect?
Include distractors that are based on content that students for this grade level are expected to know?
Avoid distractors that are too close to the correct answer that is likely to confuse or trick students who really do know the answer or can be considered outliers?
Visuals
Does the item...
Yes
No
Only include art/table that provides support for the student to demonstrate proficiency of the standard?
Include enough information for the item to be answered
Include art/table that is legible, clear, and free from visual clutter?
OnDataSuite™

44

# Why

Continuous Improvement – One District's Perspective

- Math Unit Assessments **didn't predict** our 2022 STAAR **performance**
- **Added Unit Assessments** for elementary
- Went **online** for all unit assessments with **interactive item types** included

45

# What we noticed through our analysis

Summer 2022

- Curriculum team energized by understanding statistical analysis
- Team eager to update assessments
- Unit Assessments were too easy
- Opportunities to improve questions selected from our test banks
- STAAR Released Items showing poor quality so reevaluating our curriculum supports

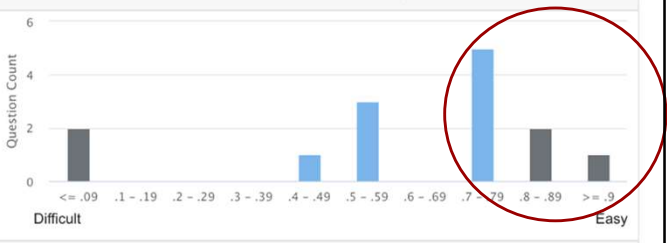
46

## Item Analysis

| Question | Difficulty | Quality | Analyzed |
|---|---|---|---|
| 1 | 0.82 | Good (0.33) | ✖ |
| 2 | 0.9 | Very Good (0.44) | ✖ |
| 3 | 0.77 | Good (0.35) | ✖ |
| 4 | 0.75 | Good (0.35) | ✖ |
| 5 | Griddable | | |
| 6 | 0.59 | Good (0.33) | ✖ |
| 7 | 0.51 | Fairly Good (0.26) | ✖ |
| 8 | 0.55 | Fairly Good (0.21) | ✖ |
| 9 | 0.77 | Fairly Good (0.23) | ✖ |
| 10 | Griddable | | |
| 11 | 0.8 | Very Good (0.4) | ✖ |
| 12 | 0.74 | Very Good (0.41) | ✖ |
| 13 | 0.47 | Marginal (0.15) | ✖ |
| 14 | 0.7 | Good (0.32) | ✖ |

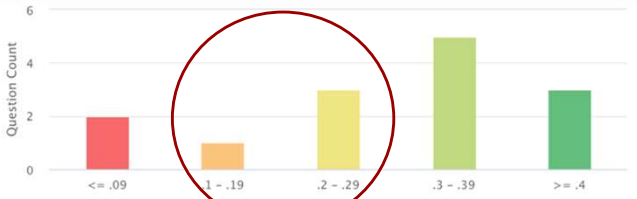
Optimum Question Difficulty from .3 through .79

## Question Difficulty Frequency Distribution




Optimum Question Difficulty from .3 through .79

## Question Quality Frequency Distribution



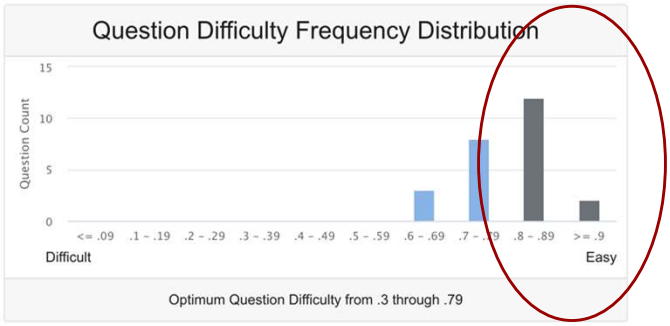

47

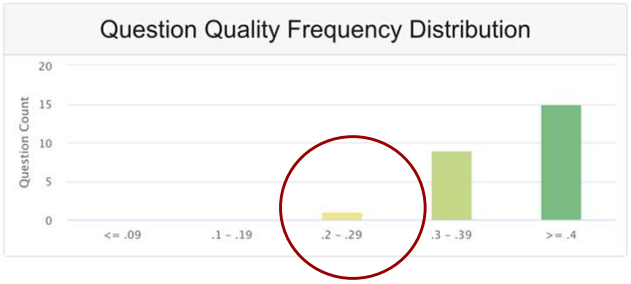
| | Meets | | | Masters | | |
|---|---|---|---|---|---|---|
| | UAs | STAAR | Difference | UAs | STAAR | Difference |
| **6th Math** | 64 | 55 | **-9** | 30 | 25 | **-5** |

## Statistical Details

| | State STAAR Test | My Test |
|---|---|---|
| **Total Questions** | 50 | 25 |
| **Reliability Coefficient (Alpha)** | 0.93 | 0.85 |
| **Average Raw Score** | 31.16 | 20.13 |
| **Standard Deviation** | 10.98 | 4.48 |
| **Mean P-Value** | 62.31 | 80.51 |

48

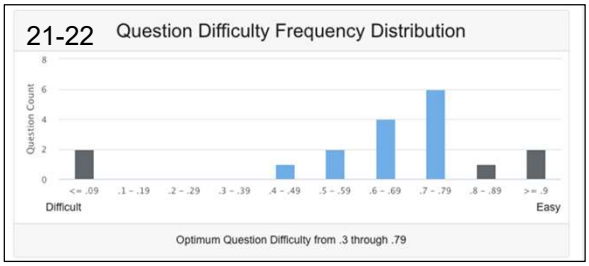


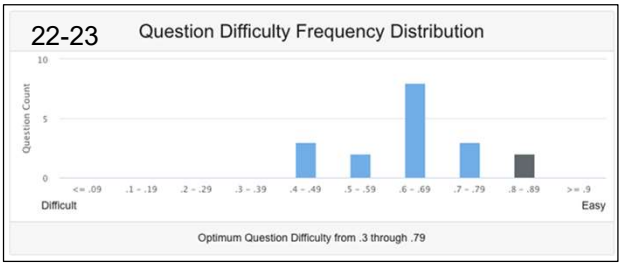




49

# Action steps

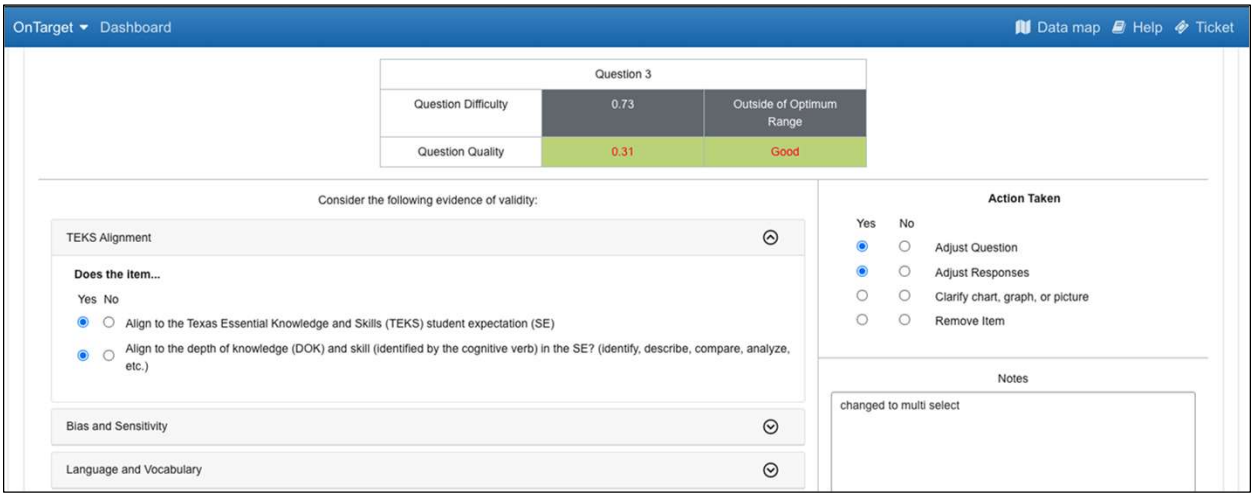
| | 21-22 | 22-23 |
|---|---|---|
| **Reliability Coefficient** | .76 | .81 |
| **Average Raw Score** | 12.2 | 15 |
| **Standard Deviation** | 3.59 | 4.68 |
| **Mean P Value** | 67.91 | 62.98 |



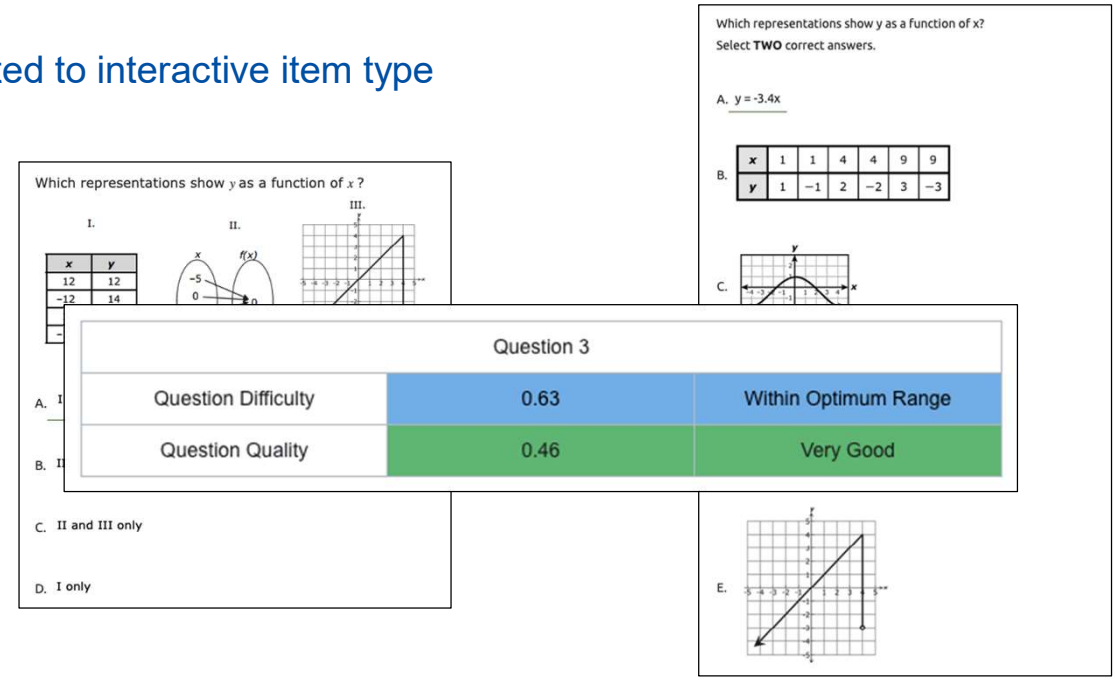

50

## Question Analysis

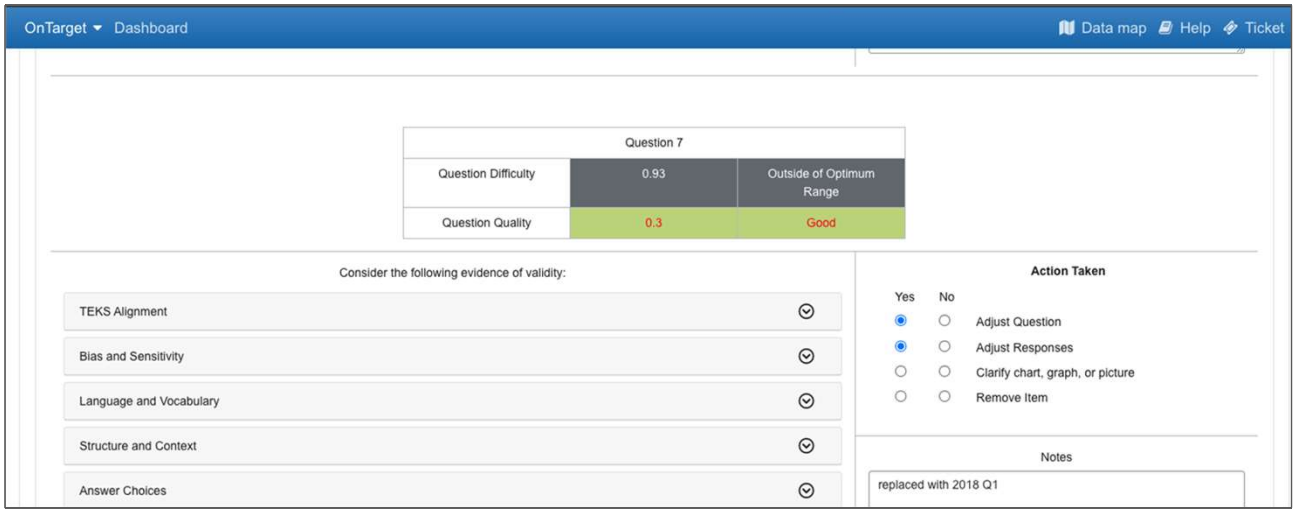



51

## Adjusted to interactive item type

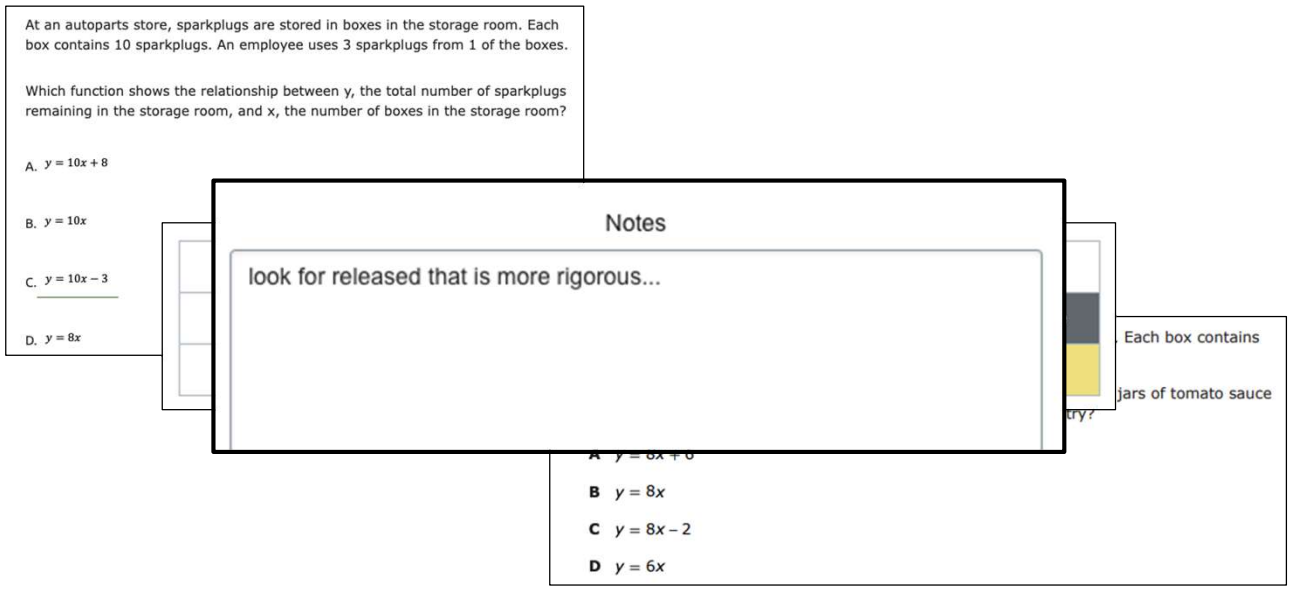



52




53

## Changed question to another released item




54

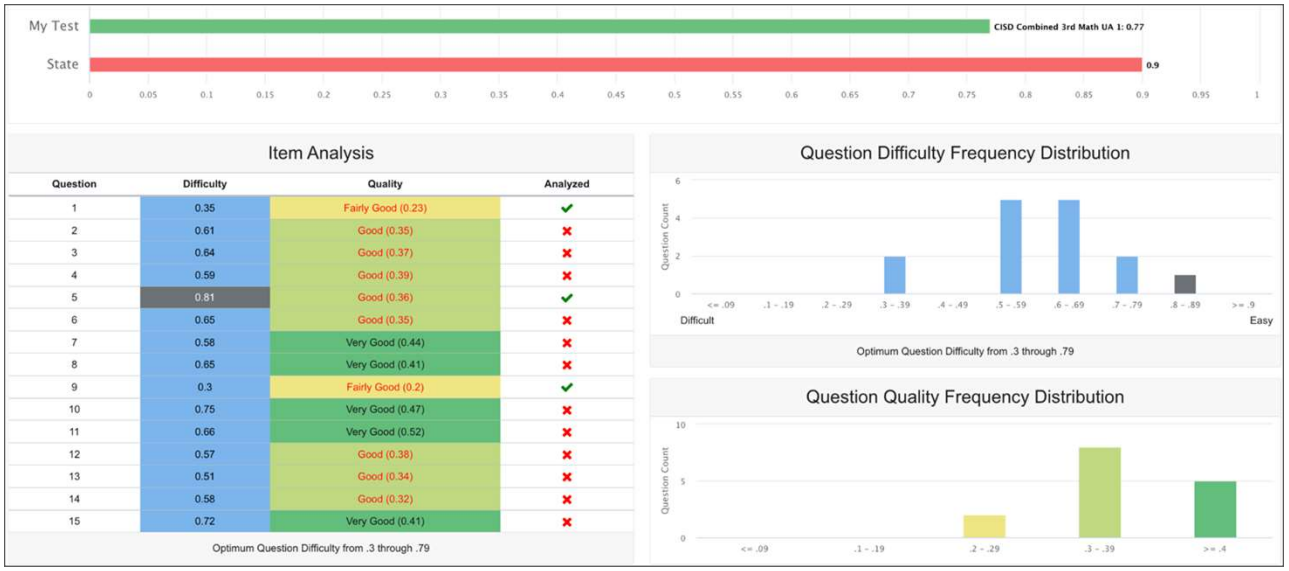
# What are districts doing with this year?

Fall 2024

- Analyzing our UAs
- Making notes for assessment changes in following year
- Reflecting on curriculum & instructional support

55

# 3rd Math UA Analysis




| Question | Difficulty | Quality | Analyzed |
|---|---|---|---|
| 1 | 0.35 | Fairly Good (0.23) | ✔ |
| 2 | 0.61 | Good (0.35) | ✖ |
| 3 | 0.64 | Good (0.37) | ✖ |
| 4 | 0.59 | Good (0.39) | ✖ |
| 5 | 0.81 | Good (0.36) | ✔ |
| 6 | 0.65 | Good (0.35) | ✖ |
| 7 | 0.58 | Very Good (0.44) | ✖ |
| 8 | 0.65 | Very Good (0.41) | ✖ |
| 9 | 0.3 | Fairly Good (0.2) | ✔ |
| 10 | 0.75 | Very Good (0.47) | ✖ |
| 11 | 0.66 | Very Good (0.52) | ✖ |
| 12 | 0.57 | Good (0.38) | ✖ |
| 13 | 0.51 | Good (0.34) | ✖ |
| 14 | 0.58 | Good (0.32) | ✖ |
| 15 | 0.72 | Very Good (0.41) | ✖ |

56

## Statistical Details

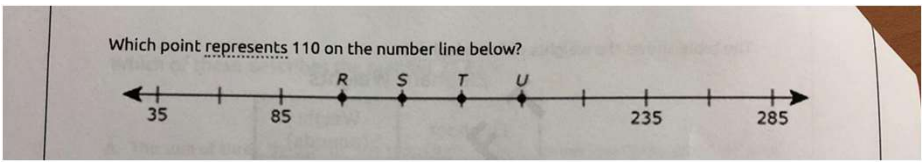
| | State STAAR Test | My Test |
|---|---|---|
| **Total Questions** | 32 | 15 |
| **Reliability Coefficient (Alpha)** | 0.9 | 0.77 |
| **Average Raw Score** | 18.3 | 11.65 |
| **Standard Deviation** | 7.62 | 3.76 |
| **Mean P-Value** | 57.19 | 59.8 |

STAAR Source: TEA, My Test Source: OnTarget

"As a general rule, reliability coefficients from 0.70 to 0.79 are considered adequate, those from 0.80 to 0.89 are considered good, and those at 0.90 or above are considered excellent."

*— TEA Technical Digest 2018 - 2019*

57




Test Name: CISD Combined 3rd Math UA 1 | Number of Questions: 15
Date Given: September 12 2022 | Number of Students: 1989
Subject Area: | Date Data Compiled: October 27, 2022
Grade Level: 3 | Combined Assessment: CISD Combined 3rd Math UA 1

Question Analysis

Question 9

| | | |
|---|---|---|
| Question Difficulty | 0.3 | Within Optimum Range |
| Question Quality | 0.2 | Fairly Good |

Consider the following evidence of validity:

TEKS Alignment

Does the item...

Yes / No

Align to the Texas Essential Knowledge and Skills (TEKS) student expectation (SE)

Align to the depth of knowledge (DOK) and skill (identified by the cognitive verb) in the SE? (identify, describe, compare, analyze, etc.)

Bias and Sensitivity

Language and Vocabulary

Structure and Context

Answer Choices

Visuals

Action Taken

Yes / No

Adjust Question

Tweak the points on the number line OR replace the question.

The student is expected to represent a number on a number line as being between two consecutive multiples of 10; 100; 1,000; or 10,000 and use words to describe relative size of numbers in order to round whole numbers.

58

| Question 1 | | |
|---|---|---|
| Question Difficulty | 0.35 | Within Optimum Range |
| Question Quality | 0.23 | Fairly Good |

Consider the following evidence of validity:

Action Taken

TEKS Alignment

**Does the item...**

Yes No

- Align to the Texas Essential Knowledge and Skills (TEKS) student expectation (SE)
- Align to the depth of knowledge (DOK) and skill (identified by the cognitive verb) in the SE? (identify, describe analyze, etc.)

Bias and Sensitivity

**Is the item...**

Yes No

- Free of bias and stereotypes (racial, gender, ethnic, religion, socioeconomic, political, environmental, etc.)?
- Free of sensitive, emotionally charged issues?
- Accessible and fair for students of diverse backgrounds so that students of one group do not have an unfair a students of another group? (consider gender, rural/urban, race/ethnicity, etc.)

Notes

4th grade reading level according to coh.metrix

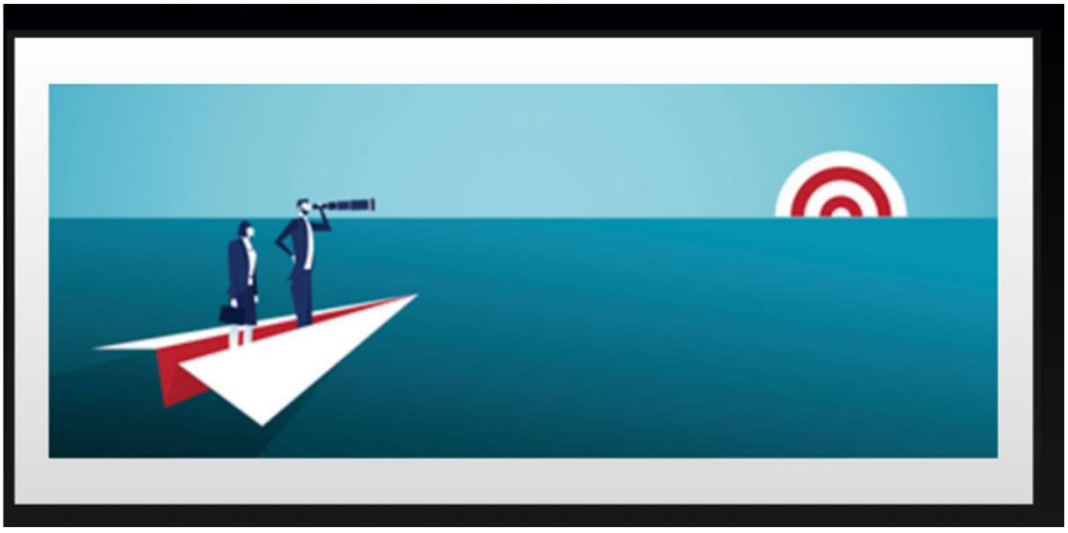
Answer choices are written in all word form.

2019 State performance 32%
District performance 35%

Instruction question - how are we exposing our students to math problems in all word form with the word NOT?

59


60



# POWER OF PREDICTABILITY

61

## References


Andrich, D. & Marais, I. (2019). A course in Rasch measurement theory: Measuring in the Educational, Social, and Health Sciences. Springer. https://www.springer.com/gp/book/9789811374951

Texas Education Agency. (2019a), STAAR 2018-2019 Technical Digest. Austin, Texas: Texas Education Agency. Retrieved from https://tea.texas.gov/student-assessment/testing/student-assessment-overview/technical-digest-2018-2019

Texas Education Agency. (2019b), 2019 STAAR Mean P-Values and Internal Consistency Values by Reporting Category and Content Area. Austin, Texas: Texas Education Agency. Retrieved from www.tea.state.tx.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=25769805361&libID=257698 05361.

Texas Education Agency. (2019c), 2019 STAAR Score Distributions and Statistics by Content Area and Grade. Austin, Texas: Texas Education Agency. Retrieved from https://tea.texas.gov/sites/default/files/digest19-appendB-STAAR-7- Stats%20and%20Dist.pdf

Varma, S. (N.D.) *Preliminary Item Statistics Using Point-Biserial Correlation and P-Values. Educational Data Systems.* *https://eddata.com/wpcontent/uploads/2015/11/EDS_Point_Biserial.pdf*

Wright, B. D. & Stone M.A. (1979). *Best test design.* Mesa Press. https://research.acer.edu.au/measurement/1/

Classroom Assessment A Practical Guide for Educators by Craig A Published by Domenic Logan


62