Hybrid Scoring Key Questions

March 2024

This presentation is intended to address the key questions the agency has received about implementing hybrid scoring.

1. What prompted the move to hybrid scoring?
2. How did TEA communicate these changes to stakeholders?
3. How does hybrid scoring fit into the assessment process?
4. Why are we seeing differences in ECR scores?

Key questions around hybrid scoring

1. What prompted the move to hybrid scoring?
2. How did TEA communicate these changes to stakeholders?
3. How does hybrid scoring fit into the assessment process?
4. Why are we seeing differences in ECR scores?

1 In 2023, STAAR was redesigned to improve alignment to the classroom experience.

	In effective classrooms, teachers are…	*The STAAR redesign will…*
1	Coherently building students' **background knowledge and vocabulary** in all subject areas...	Prioritize **cross-curricular passages** in RLA that reference topics that students have learned about in other classes
2	Asking students to **write about what they read using evidence from text**…	Include **writing in all RLA tests**, reflecting our updated TEKS, and having **students write text-based responses**
3	Providing **various open-ended formats** for students to respond to questions…	Add new**, non-multiple-choice questions** that are more like questions teachers ask in class
4	Supporting the learning needs of all students by providing **appropriate accommodations**…	Move to online assessments that provide a **full suite of robust accommodations** for students with specific learning needs

5 Moving to **online assessments** supports all the changes above and provides faster test results to support accelerated learning.

1 The STAAR redesign added more open-ended questions that are similar to the questions teachers ask in class.

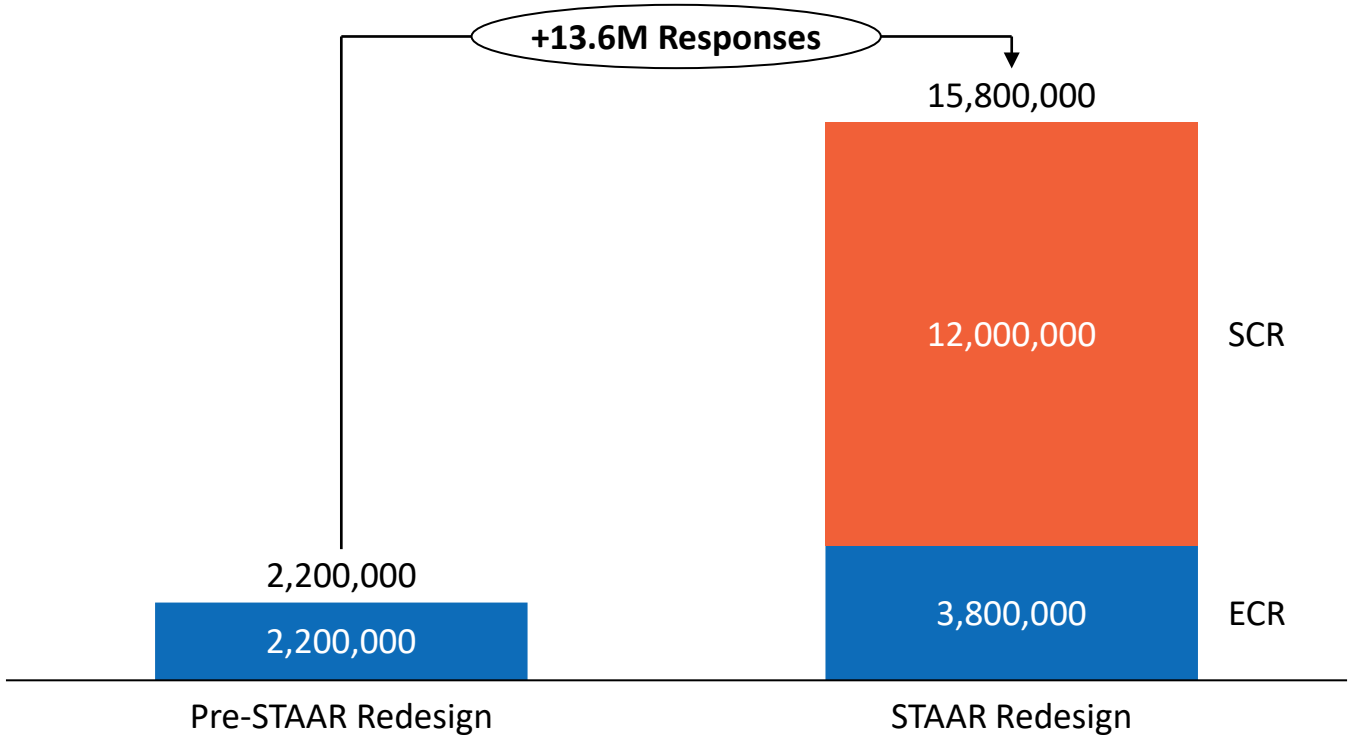
By making the test more aligned with the classroom experience, this increased the number of constructed response questions students access.

Content Area	Item Type*	Pre-STAAR Redesign	Post-STAAR Redesign
STAAR RLA	SCR	-	1-2
(G4, G7, E1, E2)	ECR	1	1
STAAR RLA	SCR	-	1-2
(Remaining titles)	ECR	-	1
STAAR Science	SCR	-	1-2
STAAR Social Studies	SCR	-	1-2

**SCR = Short Constructed Response; ECR = Extended Constructed Response*

1 The significant increase in written responses required the move to hybrid scoring to meet budget and timeline limitations.

With **6-7x more constructed responses** to grade annually for STAAR, maintaining full human scoring would have cost **$15-20M more per year**.



Key questions around hybrid scoring

1. What prompted the move to hybrid scoring?
2. How did TEA communicate these changes to stakeholders?
3. How does hybrid scoring fit into the assessment process?
4. Why are we seeing differences in ECR scores?

2 TEA communicated the transition with stakeholders in advance.

Aug. 2022

1

Testimony at House Public Education Committee

- Commissioner testified that automated scoring is a necessary step to control costs while ensuring reliability
- Relying solely on human scoring for the increased number of constructed responses as part of STAAR redesign would require $15-20M/year

Sept. 2023

2

Statewide Announcement to Testing Coordinators

- TEA announced the implementation of SY23-24 hybrid scoring during the annual test coordinator training
- This presentation was recorded and posted in the Texas Assessment Program Learning Management System for district personnel access

Nov. 2023

3

Presentations at Conferences and with Stakeholder Groups

- TEA presented about the hybrid scoring model at the 2023 Texas Assessment Co nference
- TEA provided information in other stakeholder group presentations, e.g., Texas Science Education Leadership Association (TSELA), Texas Social Studies Supervisors Association (TSSSA), Texas Council of Teachers of English Language Arts (TCTELA), Coalition of Reading and English Supervisors of Texas (CREST)

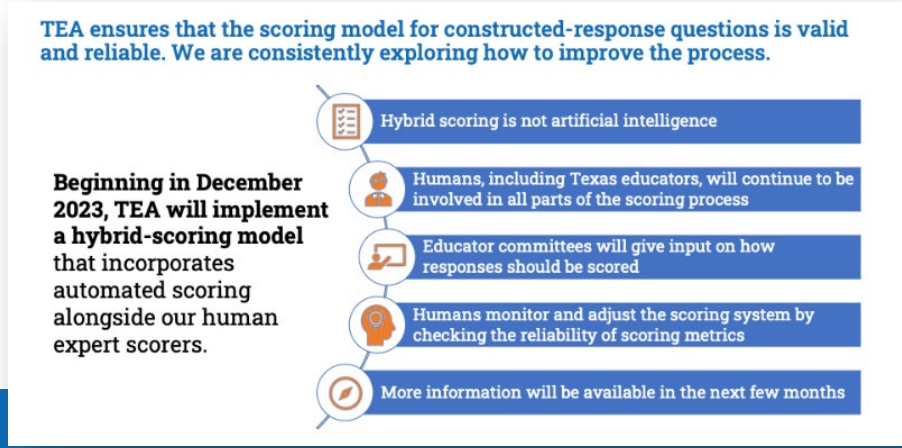
Dec. 2023

4

Release of STAAR Scoring Process Document

- The detailed scoring process document was published on TEA's website

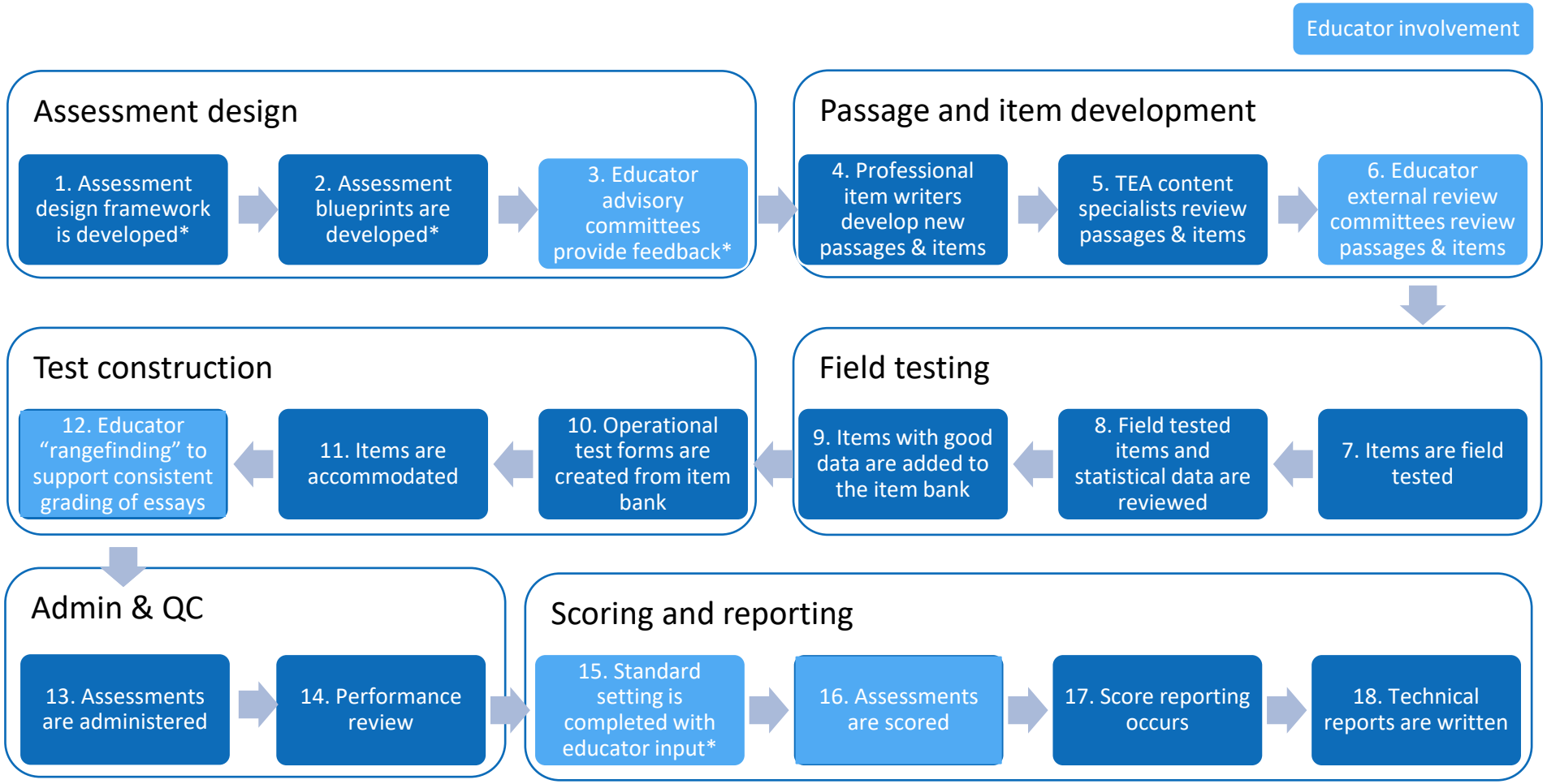
Example slide from TEA presentations:



Key questions around hybrid scoring

1. What prompted the move to hybrid scoring?
2. How did TEA communicate these changes to stakeholders?
3. How does hybrid scoring fit into the assessment process?
4. Why are we seeing differences in ECR scores?

3 Creating high-quality state assessments is a rigorous process that includes educators across multiple steps.

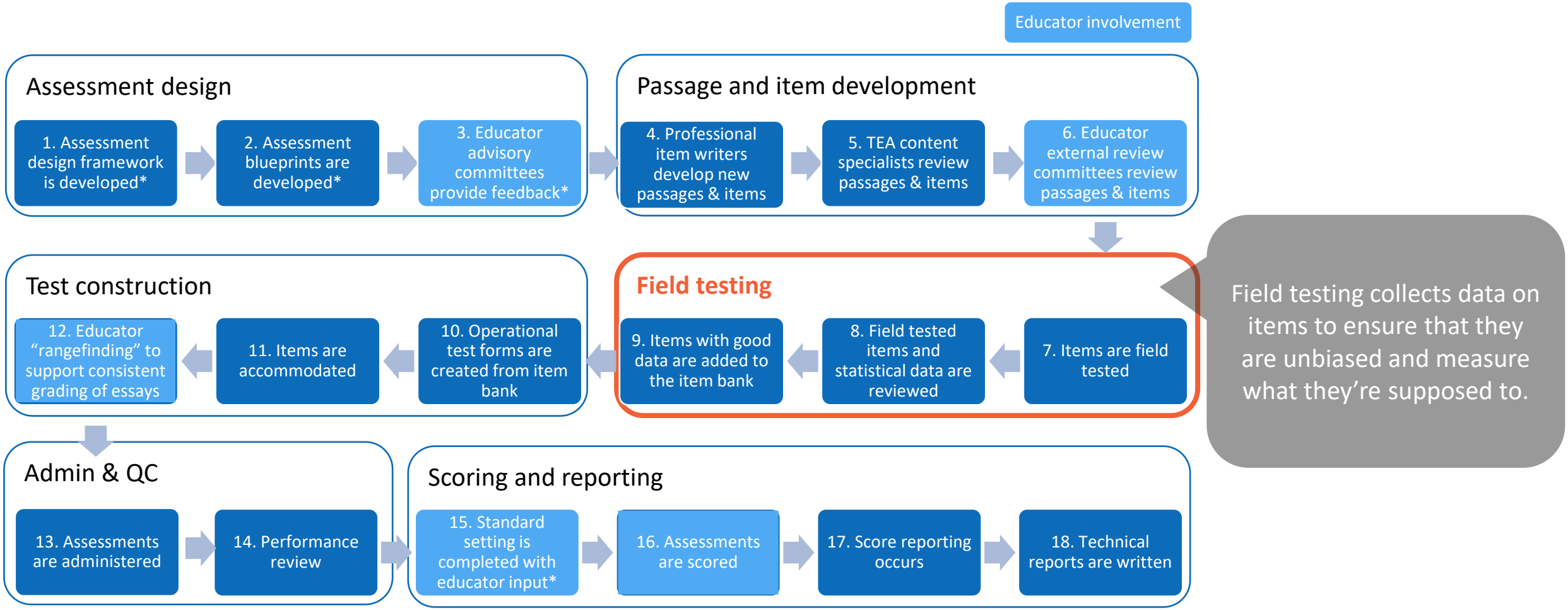


It takes over **two years** from assessment design to scoring and reporting.

Thus, any changes to assessments must be planned far in advance.

*Does not occur every year

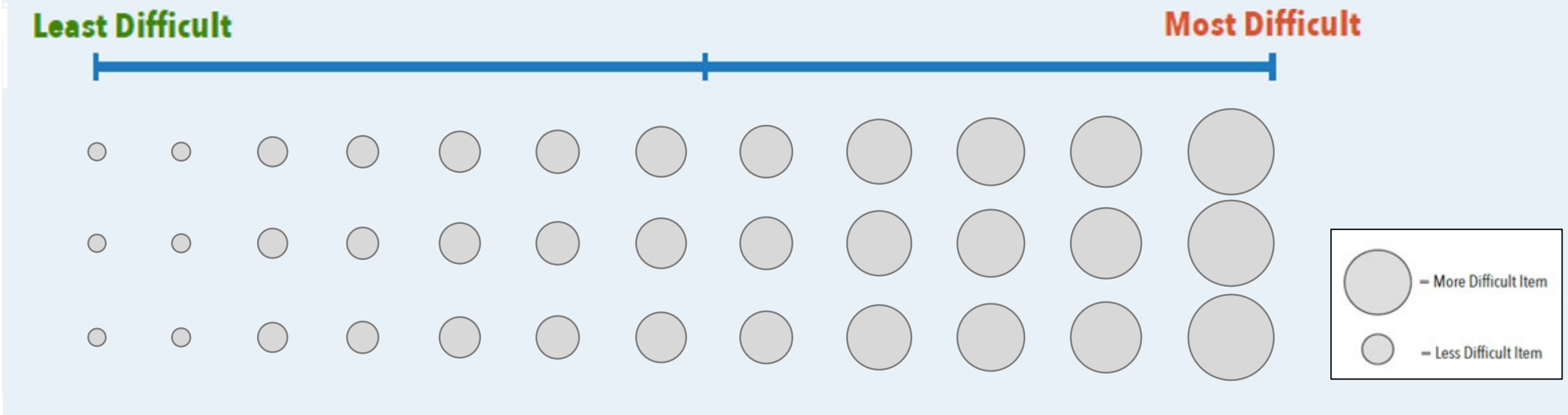
3 Field testing serves as a building block for assessment construction and is not related to the move to hybrid scoring.



*Does not occur every year

3 The purpose of field testing is to populate a bank that is diverse and free of bias.

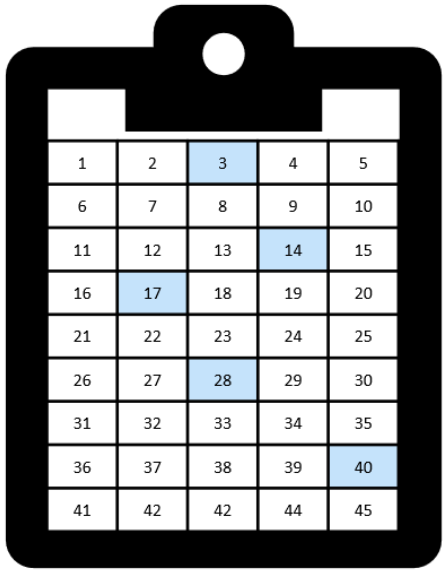
To best measure student performance, TEA needs to build an item bank that represents all items from the least to most difficult spectrum:



When we field test, we are testing the item, not the students.

3 Most field test items are embedded within the STAAR assessments, with the exception of extended constructed response (ECR) items.

A student typically interacts with field test items within their STAAR test. They would not know which items are a part of their assessment and which items are field test items. Field test items do not count towards a student's score.



STAAR tests have embedded field test items.

Extended constructed response (ECR) items cannot be placed within a STAAR test due to its length.

Therefore, Stand Alone Field Tests (SAFT) have been established for ECRs. This is the same process for field testing ECRs before and after the STAAR redesign.

Selected students take an assessment that includes the ECR field test mixed in with other items. This is to simulate a testing environment and provide items that help link the field test items back to the STAAR scale.

3 After the field test event, each constructed response item is scored against a rubric by two humans.

Short Constructed Response (SCR)

Rubrics are developed alongside passages and reviewed during educator item review meetings

→ Rubrics range from 0-2 points

- RLA writing SCRs use 0-1 point rubrics
- RLA reading SCRs use 0-2 point rubrics
- Science and Social Studies SCRs use 0-2 point rubrics

→ There are 2 human scorers involved

- The first human's score is the score used
- The second human's score is kept for quality assurance purposes

Extended Constructed Responses (ECR)

Rubrics were established during STAAR redesign and are static

→ Rubrics are 5 points in total

- Development and organization (max 3 pts)
- Language and conventions (max 2 pts)

→ Scores from each human scorer is combined for a max score of 10

- If the scores are adjacent, scores are summed up
- If not, it goes through adjudication process where new score received is doubled

Rubrics are employed to ensure consistency in scoring for items that are open-ended. (See examples in the appendix.)

3 All of the humans involved in the scoring process are highly trained and calibrated.

To qualify as a rater, one must have a 4-year college degree and experience teaching at the assigned grade level.

Raters undergo rigorous training to learn how to use the standardized rubric to score student responses.

To pass training, they must accurately rate example responses that have already been scored.

Each certified rater's performance is calibrated at regular intervals to ensure that all responses are graded consistently across Texas.

More on this later

3 Field test constructed response items continue to be scored against a rubric by two humans.

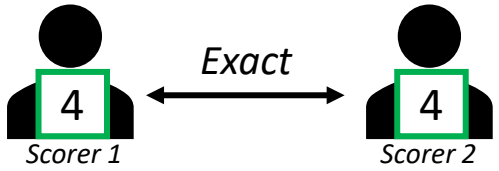
Short Constructed Response (SCR)

Two humans score each response.

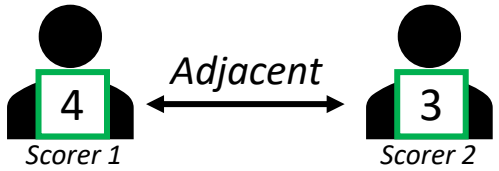
Scorer 1's rating is the score of record.

Scorer 2's rating is used for the purpose of auditing / quality control.

Extended Constructed Responses (ECR)

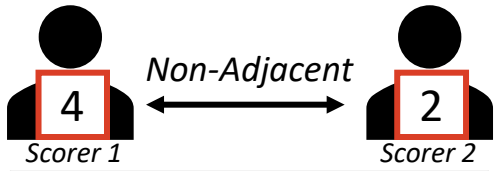


= 8



= 7

Scenario 1: *If two raters show exact or adjacent (within 1 point) agreement on a response, then the scores are summed together to create the score of record.*

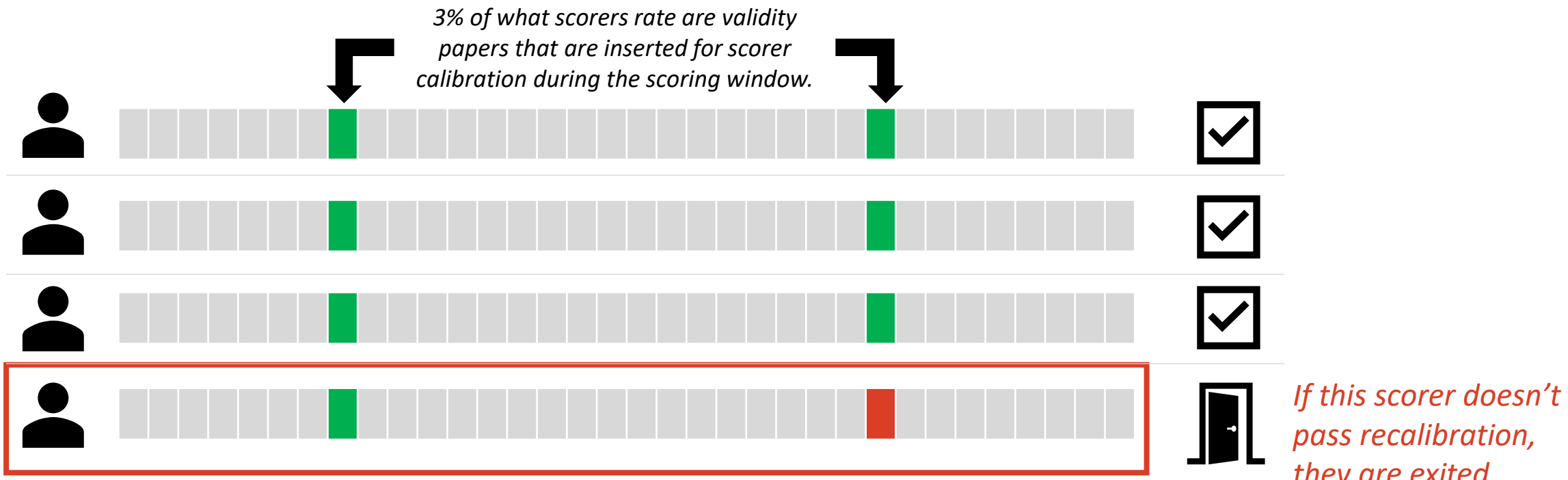


x 2

= 8

Scenario 2: *If two raters show more than 1 point difference in rating a response, then a scoring leader takes over and assigns the score of record.*

3 Within an administration period, any rater that does not meet standard is removed from scoring.



If a scorer does not maintain at <u>least at 65% exact agreement and 95% adjacent agreement</u> during the scoring window, they cannot remain as a scorer for that admin. If they fail re-calibrations, they will have to try again in the next administration.

3 Pre-equating is a part of the test construction process that occurs after field testing.

Educator involvement

Assessment design

1. Assessment design framework is developed*
2. Assessment blueprints are developed*
3. Educator advisory committees provide feedback*

Passage and item development

4. Professional item writers develop new passages & items
5. TEA content specialists review passages & items
6. Educator external review committees review passages & items

Field testing

7. Items are field tested
8. Field tested items and statistical data are reviewed
9. Items with good data are added to the item bank

Test construction

10. Operational test forms are created from item bank
11. Items are accommodated
12. Educator "rangefinding" to support consistent grading of essays

Admin & QC

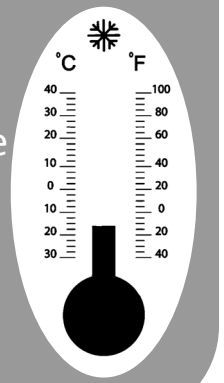
13. Assessments are administered
14. Performance review

Scoring and reporting

15. Standard setting is completed with educator input*
16. Assessments are scored
17. Score reporting occurs
18. Technical reports are written

Equating is the statistical process by which the performance of items from different administrations can be compared by placing the items on the same scale.

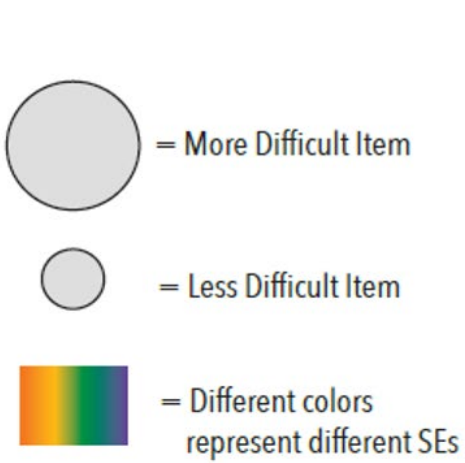
Equating ensures students taking the STAAR receive the correct scale score.



*Does not occur every year

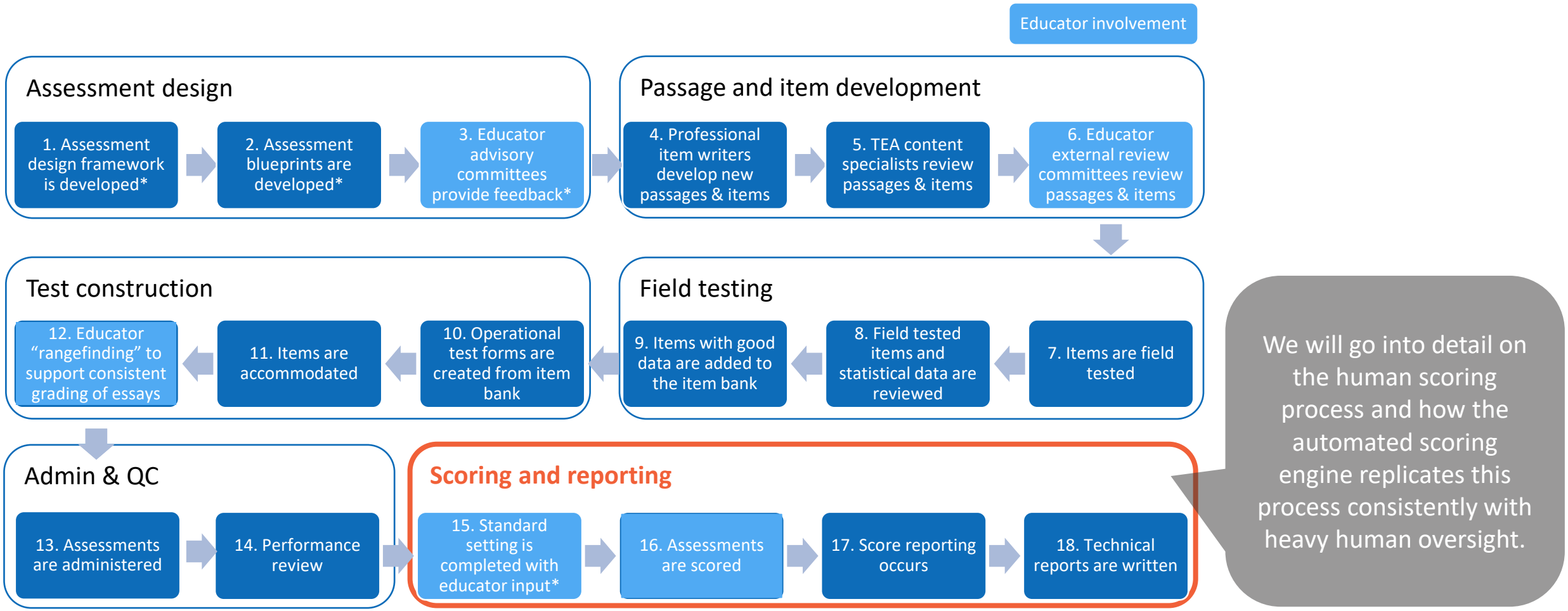
3 Equating ensures that the STAAR test is the same level of difficulty each year.

While individual items can be easier or harder in a given year, the mix of item difficulty is balanced across years by using field test results.



Click to see STAAR FAQ video: "How do we know the STAAR test is the same level of difficulty from year to year?".

3 Scoring and reporting is the final process of the assessment lifecycle.



*Does not occur every year

3 Prior to hybrid scoring, all constructed response items were scored the same way as field testing, but the process has changed.

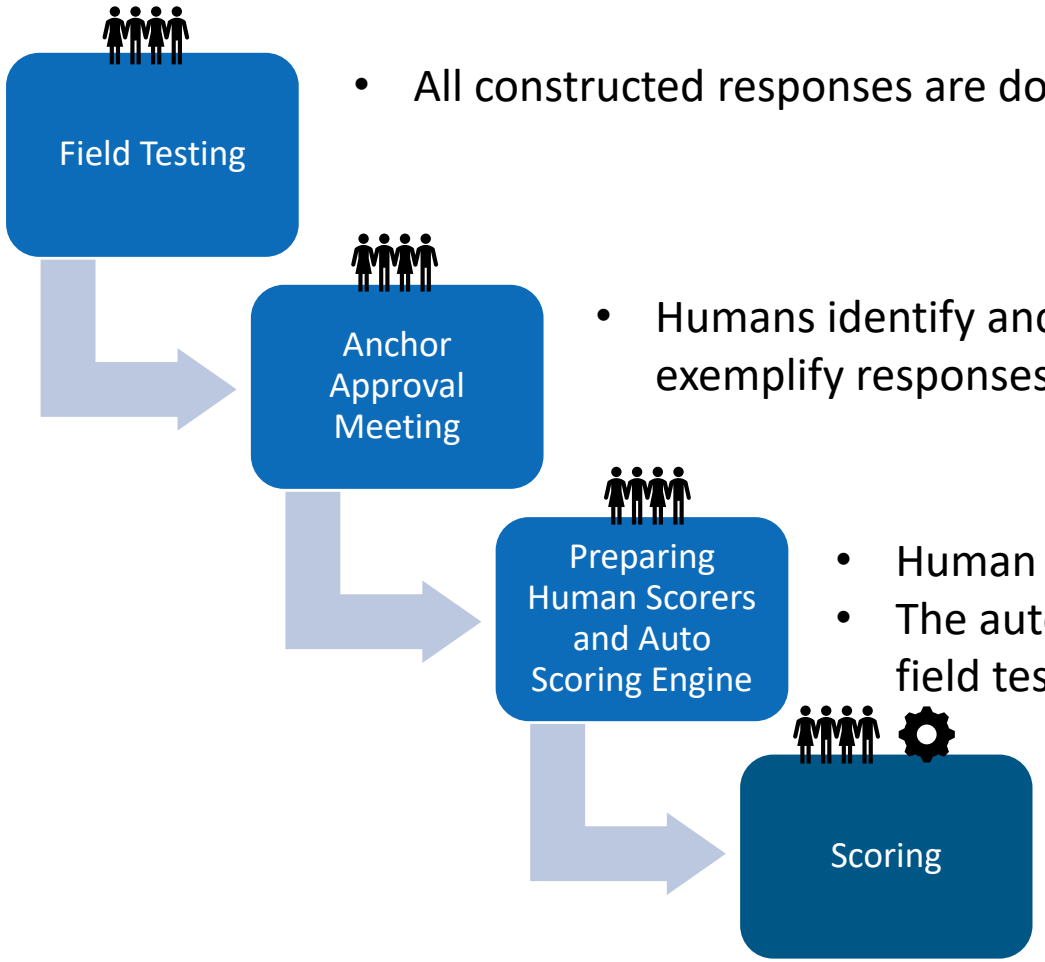
Scoring Process Today

Item Type	Field Testing	Operational Scoring
Non-Constructed Response Items	**Machine Scored**	**Machine Scored**
Short Constructed Response (SCR) Items	**2 human scorers** (one score of record, one for auditing)	**Hybrid Scoring** All scores go through auto scoring engine; 25% are double human scored (one score of record, one for auditing).
Extended Constructed Response (ECR) Items	**2 human scorers** (scores are combined for total score)	**Hybrid Scoring** All scores go through auto scoring engine; 25% are double human scored (scores are combined for total score).

More in next slides

While this process has changed, TEA is ensuring that the quality of scoring remains the same.

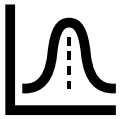
3 Up until the CR scoring event itself, each part of the CR scoring prep process relies solely on human input.



- All constructed responses are double human scored.
- Humans identify anchor responses which are field test responses that exemplify responses at each rubric score point.
- Human scorers are trained through the anchor responses.
- The automated scoring engine is programmed through ~3,000 hand-scored field test responses and human-identified anchor responses.
- Hybrid scoring: Human raters support the automated scoring engine through calibration checks and scoring non-standard responses during the administration window.

3 The auto scoring engine (ASE) goes through a rigorous programming process that is led and checked by humans.

For each item being scored…



The engine uses a sample of ~3,000 human scored responses from the field test for programming.

The engine analyzes the responses to identify common patterns and is programmed to emulate how humans would score.

TEA evaluates the performance for each item and compares it to how humans would score.

The engine is monitored throughout the scoring cycle to ensure that it remains calibrated to the anchor set.

Similar to human scorers who need to be constantly calibrated throughout the scoring window, there is a parallel process for the ASE.

3 TEA conducted a proof-of-concept study with STAAR Spring 2023 operational data before implementing hybrid scoring.

- Spring 2023 constructed response items were scored entirely by humans. The study was conducted after score reports were sent out to districts.
- The study "re-scored" constructed response items with the automated scoring engine and compared how closely the engine performed to humans.
- Five subgroups were examined (Male, Female, Black, Hispanic/Latino, White).
- The proof-of-concept study was successful and found that the automated scoring engine met the performance criteria to be implemented operationally.
- A detailed technical report on the Spring 2023 study can be found on the Assessment Reports and Studies webpage.

*Report can be found under "Additional Reports and Studies".

Example: *Spring STAAR 2023 ECR Conventions – Exact Agreement for ASE Model 2*

Item	Human-Human	Human-Engine	Difference
1	72%	74%	2%
2	71%	72%	1%
3	67%	66%	-1%
4	67%	72%	5%
5	67%	73%	6%
6	71%	76%	5%

TEA analyzed a number of criteria, including exact agreement (shown here), adjacent agreement, and score distributions to ensure the engine performs as expected.

3 Similarly, STAAR December 2023 data confirms the ASE performed as expected.

December 2023 SCR Exact Agreement Results

Grade	Item	Subject	Human-Human Exact Agreement	Human-Engine Exact Agreement	Difference
9	1	Biology	96.5%	97.1%	0.6%
9	2	Biology	95.8%	95.1%	-0.7%
9	3	Read	82.6%	86.8%	4.3%
9	4	USH	86.2%	92.4%	6.2%
9	5	USH	95.1%	97.1%	2.0%
9	6	Write	95.6%	97.3%	1.7%
10	8	Read	76.2%	79.1%	2.9%
10	9	Write	92.2%	97.0%	4.8%

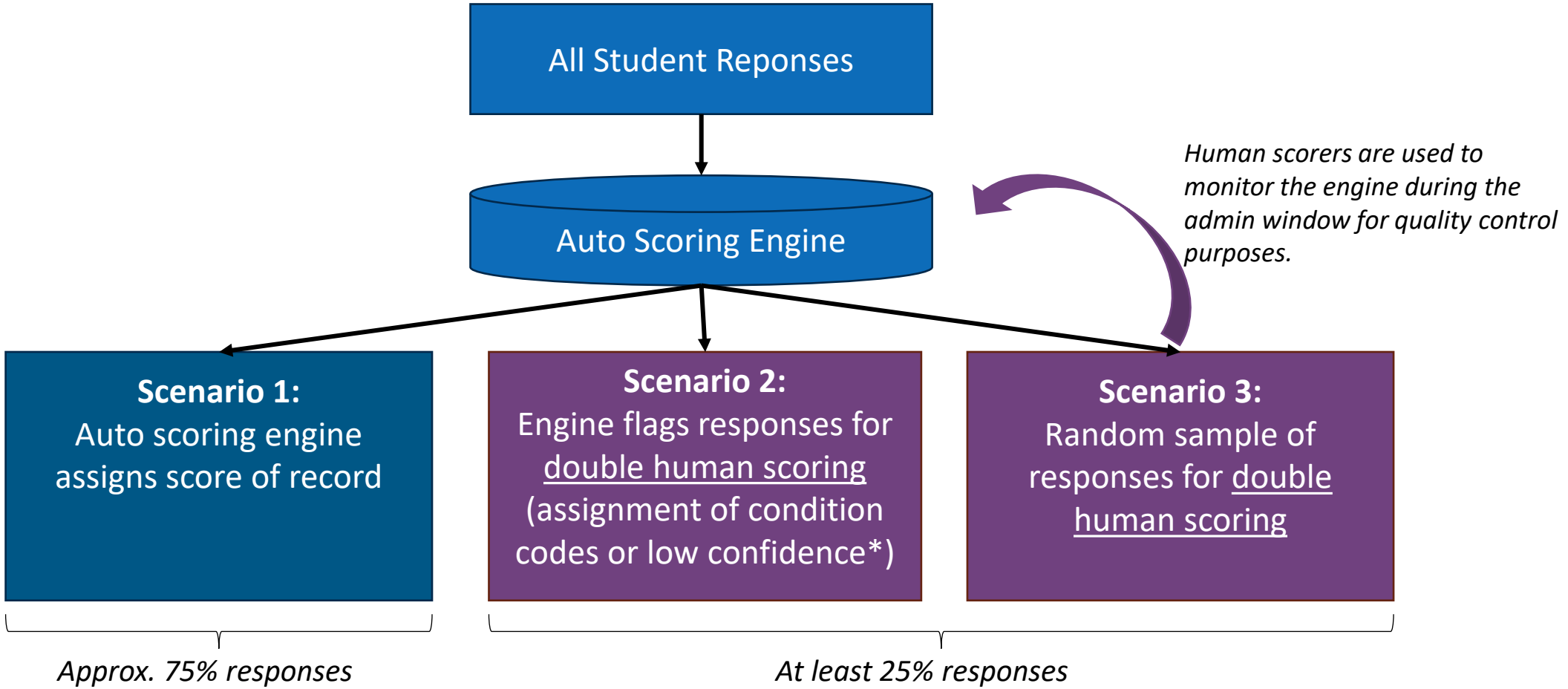
December 2023 ECR Exact Agreement Results

Grade	Item	Subject	Human-Human Agreement	Human-Engine Agreement	Difference
9	7	Read	Convention 69.3% Ideas 66.5%	Convention 73.9% Ideas 70.5%	4.6% 4.0%
10	10	Read	Convention 78.4% Ideas 73.5%	Convention 87.9% Ideas 77.1%	9.5% 3.6%

- TEA used the hybrid scoring approach to score all constructed responses in December 2023.
- Constructed response items saw similar human-engine exact agreement to human-human exact agreement.
- TEA will continue to monitor hybrid scoring during every administration to ensure ASE produces accurate scores.

Lower exact agreement levels are expected with ECRs compared to SCRs due to larger range of possible points.

3 The Texas hybrid scoring model uses an automated scoring engine to augment the work of human scorers.

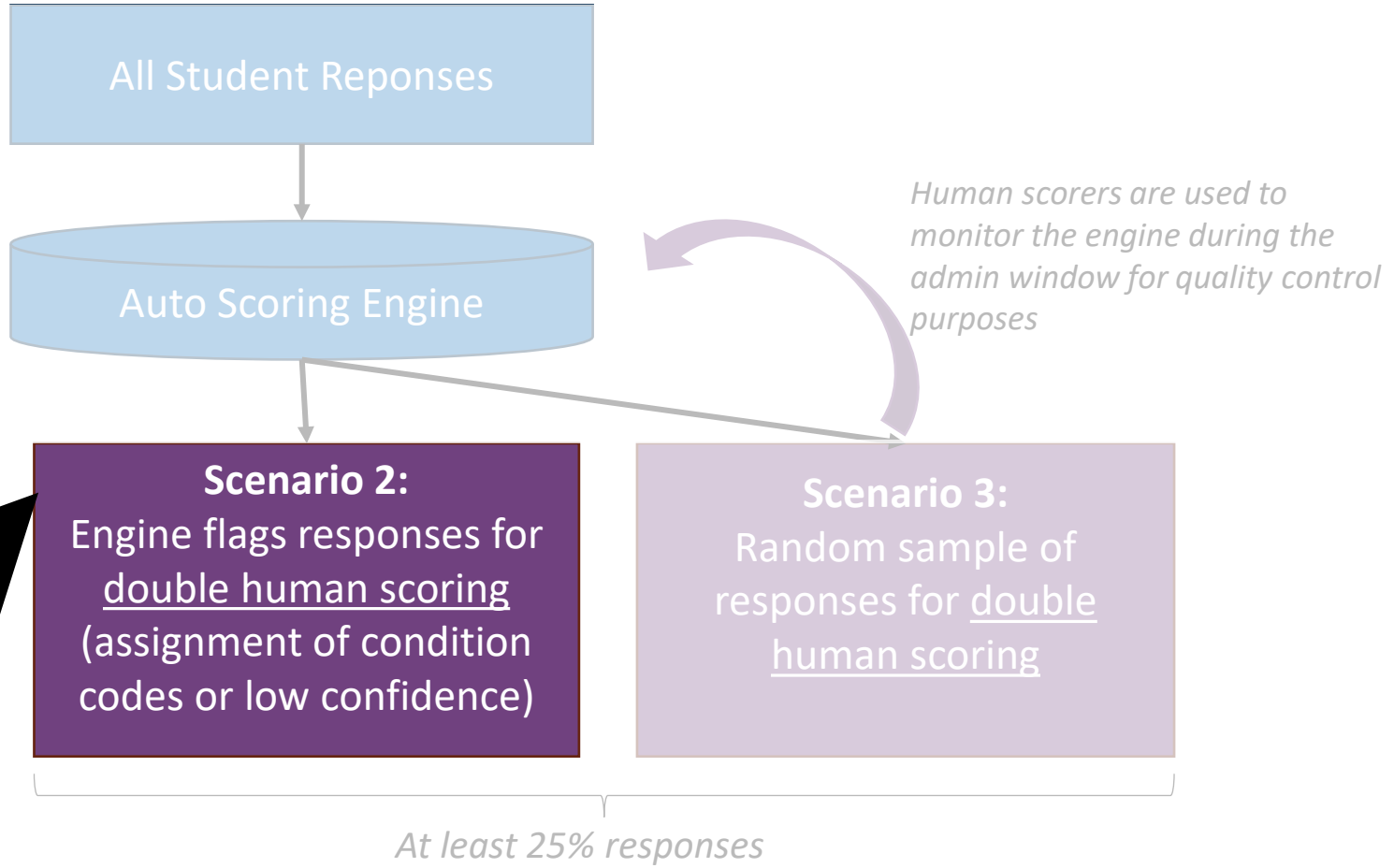


Note: Any student responses that are routed for human scoring maintain the score assigned by humans as the score of record. Human scoring will also go through the adjudication process if needed.
*Condition codes that get sent for human scoring are those flagged for unusual patterns; low confidence responses are often those responses that are on the border between two score points.

3 The ASE assigns condition codes to some responses, which are each routed to two trained human scorers.

Condition codes indicate that a response uses just a few words, uses mostly duplicated text, is written in another language, consists primarily of text from the passage, uses vocabulary that does not overlap with the vocabulary in the subset of responses used to program the ASE, or uses language patterns that are reflective of off-topic or off-task responses.

The purpose of this routing is to ensure that these unusual responses receive fair and accurate scoring. The score assigned by the human scorer is kept as the score of record for any student response that is routed for human scoring.



Note: Any student responses that are routed for human scoring maintain the score assigned by humans as the score of record. Human scoring will also go through the adjudication process if needed

3 Automated scoring technology* is over a decade old and is widely used, including in Texas.

10+

years

amount of time technology for automated scoring engine has been around

180K+

Texas students

annually use the Texas Success Initiative Assessment (TSIA) to meet their graduation requirement, which relies on automated scoring technology

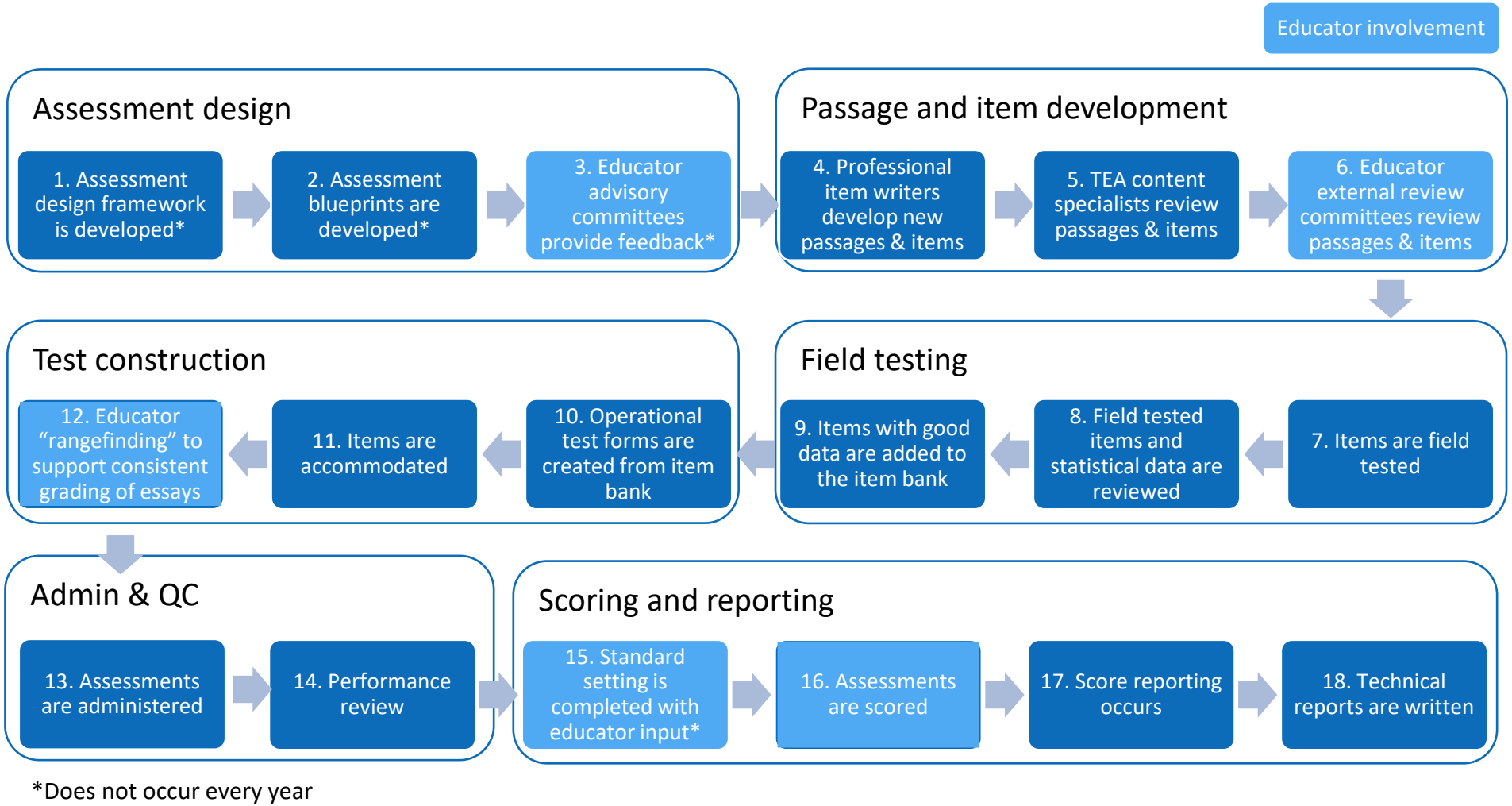
21+

states

currently employ auto scoring for their state assessments

*This kind of technology is different from AI in that AI is a computer using progressive learning algorithms to adapt, allowing the data to do the programming and essentially teaching itself. Instead, the automated scoring engine is a closed database with student response data accessible only by TEA and, with strict contractual privacy controls, its assessment contractors, Cambium and Pearson.

3 The Texas Assessment Program continues to strive for assessment development transparency.



*Does not occur every year

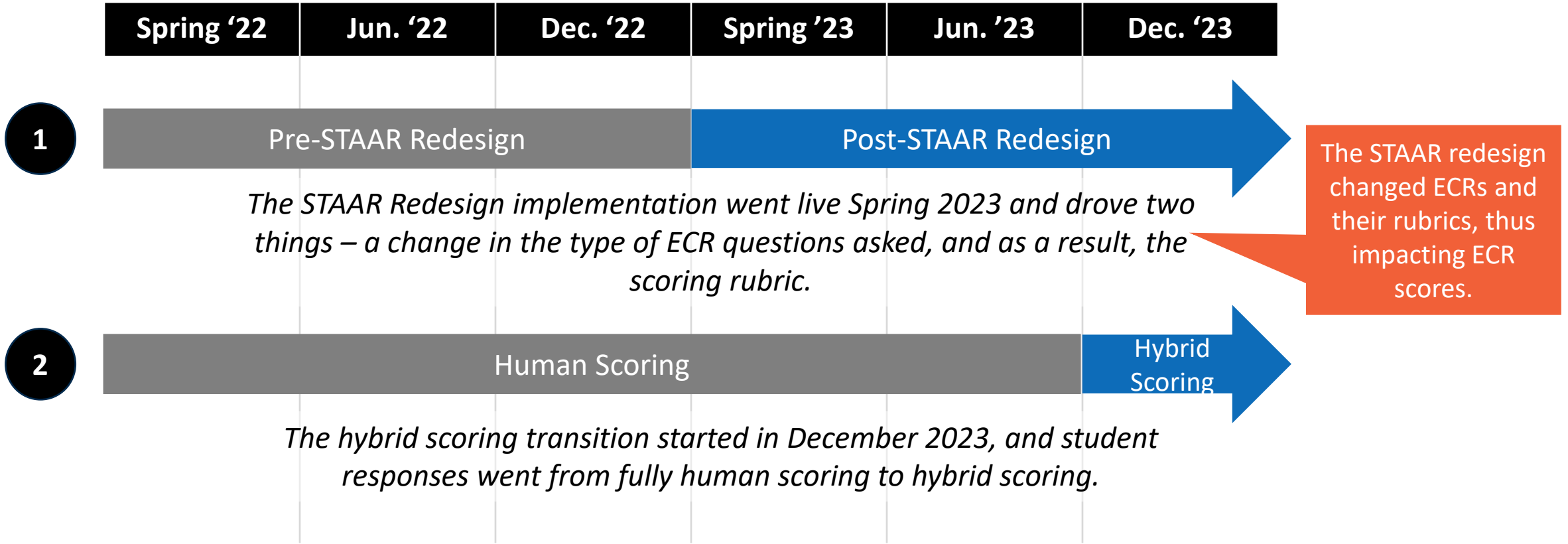
It is uncommon for states to release all test items for primary spring test administrations on an annual basis, yet Texas has made that commitment.

June and December STAAR tests cannot be released annually because items are re-used in later tests. To release June and December tests annually, TEA would need to develop and field test 3x the number of items.

Key questions around hybrid scoring

1. What prompted the move to hybrid scoring?
2. How did TEA communicate these changes to stakeholders?
3. How does hybrid scoring fit into the assessment process?
4. Why are we seeing differences in ECR scores?

4 There are two timelines and changes to account for when comparing ECR scores.



4 STAAR Redesign: Based on stakeholder feedback, ECRs were redesigned to ask students to use evidence from text.

Pre-STAAR Redesign

Writing responses to standalone prompts

WRITTEN COMPOSITION: Expository

Read the following quotation.

> The reason most people never reach their goals is that they don't define them, or ever seriously consider them as believable or achievable. Winners can tell you where they are going, what they plan to do along the way, and who will be sharing the adventure with them.
>
> —*Denis Waitley*

Successful people often set clear goals and understand the specific steps needed to achieve them. Think carefully about this statement.

Write an essay explaining the best way to achieve a goal.

Be sure to —

- clearly state your thesis
- organize and develop your ideas effectively
- choose your words carefully
- edit your writing for grammar, mechanics, and spelling

Spring 2022 English 1 EOC Example

STAAR Redesign Implementation

Writing responses using evidence from text

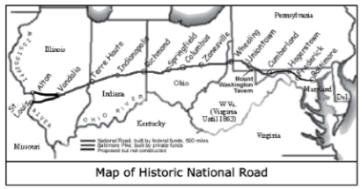
Read the selection and choose the best answer to each question.

Back in Time: The National Road

by Rickie Longfellow

1 The National Road, in many places known as Route 40, was built between 1811 and 1834 to reach the western settlements. It was the first federally funded road in U.S. history. George Washington and Thomas Jefferson believed that a trans-Appalachian road was necessary for unifying the young country. In 1806 Congress authorized construction of the road and President Jefferson signed the act establishing the National Road. It would connect Cumberland, Maryland, to the Ohio River.

2 In 1811 the first contract was awarded and the first 10 miles of road built. By 1818 the road was completed to Wheeling, and mail coaches began using the road. By the 1830s the federal government conveyed part of the road's responsibility to the states through which it runs. Tollgates and tollhouses were then built by the states, with the federal government taking responsibility for road repairs.



Map of Historic National Road

3 As work on the road progressed, a settlement pattern developed that is still visible. Original towns and villages are found along the National Road, many barely touched by the passing of time. The road, also called the Cumberland Road, National Pike, and other names, became Main Street in these early settlements, earning the nickname "The Main Street of America." The height of the National Road's popularity came in 1825 when it was celebrated in song, story, painting, and poetry. During the 1840s popularity soared again. Travelers and drovers, westward bound, crowded the inns and taverns along the route. Huge Conestoga wagons hauled produce from frontier farms to the East Coast, returning

30

GUEST, GUEST

Read the article "Back in Time: The National Road." Based on the information in the article, write a response to the following:

Explain why travelers enjoy using the National Road today.

Write a well-organized essay that uses specific evidence from the article to support your answer.

Remember to —

- clearly state your thesis
- organize your writing
- develop your ideas in detail
- use evidence from the selection in your response
- use correct spelling, capitalization, punctuation, and grammar

Manage your time carefully so that you can —

- review the selection
- plan your response
- write your response
- revise and edit your response

Write your response in the box provided.

Spring 2023 English EOC Example

4 STAAR Redesign: In addition, the educator-approved new rubrics for writing prompts introduce a possible score of zero.

Pre-STAAR Redesign

Lowest score that could be assigned to a response was a 2.

Zeros were only reserved for unscorable responses (i.e., blanks, random characters).

STAAR Redesign Implementation

Students could earn a zero through the rubric or through a non-scoreable response.

English I
Constructed Response Scoring Guide

English I Passage with Extended Constructed Response

Informational Writing Rubric

Score Point	Organization and Development of Ideas
3	• **Controlling idea/Thesis is clear and fully developed** The controlling idea/thesis is clearly identifiable. The focus is consistent throughout, creating a response that is unified and easy to follow. • **Organization is effective** A purposeful structure that includes an effective introduction and conclusion is evident. The organizational structure is appropriate and effectively supports the development of the controlling idea/thesis. The sentences, paragraphs, or ideas are logically connected in purposeful and highly effective ways. • **Evidence is specific, well chosen, and relevant** The response includes relevant text-based evidence that is clearly explained and consistently supports and develops the controlling idea/thesis. For pairs in grades 6 through EII, evidence is drawn from both texts. The response reflects a thorough understanding of the writing purpose. • **Expression of ideas is clear and effective** The writer's word choice is specific, purposeful, and enhances the response. Almost all sentences and phrases are effectively crafted to convey the writer's ideas and contribute to the overall quality of the response and the clarity of the message.
2	• **Controlling idea/Thesis is present and partially developed** A controlling idea/thesis is presented, but it may not be clearly identifiable because it is not fully developed. The focus may not always be consistent and may not always be easy to follow. • **Organization is limited** A purposeful structure that includes an introduction and conclusion is present. An organizational structure may not be consistent and may not always support the logical development of the controlling idea/thesis. Sentence-to-sentence connections and clarity may be lacking. • **Evidence is limited and may include some irrelevant information** The response may include text-based evidence to support the controlling idea/thesis, but it may be insufficiently explained, and/or some evidence may be irrelevant to the controlling idea/thesis. For pairs, evidence is drawn from at least one of the texts. The response reflects partial understanding of the writing purpose. • **Expression of ideas is basic** The writer's word choice may be general and imprecise and at times may not convey the writer's ideas clearly. Sentences and phrases are at times ineffective and may interfere with the writer's intended meaning and weaken the message.

English I
Constructed Response Scoring Guide

1	• **Controlling idea/Thesis is evident but not developed** A controlling idea/thesis is present but not developed appropriately in response to the writing task. • **Organization is minimal and/or weak** An introduction or conclusion may be present. An organizational structure that supports logical development is not always evident or is not appropriate to the task. • **Evidence is insufficient and/or mostly irrelevant** Little text-based evidence is presented to support the controlling idea/thesis, or the evidence presented is mostly extraneous and/or repetitious. Explanation of any evidence presented is insufficient and may be only vaguely related to the writing task. For pairs in grades 6 through EII, evidence is drawn from only one text. The response reflects a limited understanding of the writing purpose. • **Expression of ideas is ineffective** The writer's word choice is vague or limited and may impede the quality and clarity of the essay. Sentences and phrases are often ineffective, interfere with the writer's intended meaning, and impact the strength and clarity of the message.
0	• A controlling idea/thesis may be evident. • The response lacks an introduction and conclusion. An organizational structure is not evident. • Evidence is not provided or is irrelevant. The response reflects a lack of understanding of the writing purpose. • The expression of ideas is unclear and/or incoherent. ***Please note that if a response receives a score point 0 in the Development and Organization of Ideas trait, the response will also earn 0 points in the Conventions trait.***

Score Point	Conventions
2	Student writing demonstrates **consistent command** of grade-level-appropriate conventions, including correct: • sentence construction • punctuation • capitalization • grammar • spelling The response has few errors, but those errors do not impact the clarity of the writing.

English I
Constructed Response Scoring Guide

1	Student writing demonstrates **inconsistent command** of grade-level-appropriate conventions, including limited use of correct: • sentence construction • punctuation • capitalization • grammar • spelling The response has several errors, but the reader can understand the writer's thoughts.
0	Student writing demonstrates **little to no command** of grade-level-appropriate conventions, including infrequent use of or no evidence of correct: • sentence construction • punctuation • capitalization • grammar • spelling The response has many errors, and these errors impact the clarity of the writing and the reader's understanding of the writing.

Example rubric from the 2023 English I Constructed Response Scoring Guide

4 To make a proper comparison, we isolated the scoring data to only include retesters who existed in each test administration.

		Pre-STAAR Redesign			Post-STAAR Redesign		
		Human Scoring					Hybrid Scoring
		Spring '22	**Jun. '22**	**Dec. '22**	**Spring '23**	**Jun. '23**	**Dec. '23**
English 1	Zero on ECR	**9%**	**10%**	**11%**	**50%**	**68%**	**72%**
	Approaches or Above on EOC	**21%**	**16%**	**22%**	**33%**	**20%**	**32%**
	Number of Retesters	*82,755*	*6,932*	*125,320*	*95,551*	*53,654*	*106,926*
English 2	Zero on ECR	**7%**	**9%**	**9%**	**62%**	**68%**	**84%**
	Approaches or Above on EOC	**28%**	**8%**	**25%**	**28%**	**19%**	**24%**
	Number of Retesters	*59,617*	*44,687*	*89,918*	*64,857*	*34,859*	*76,763*

Percent of Retesters Who Received a Zero Score on the Extended Constructed Response Item, Percent of Retesters Who Achieved Approaches Grade Level or Better on the EOC as a Whole, and Number of Retesters

4 The STAAR redesign resulted in notably higher zero rates on ECRs in 2023 but did not impact overall test performance.

		Pre-STAAR Redesign			Post-STAAR Redesign		
		Human Scoring				Hybrid Scoring	
		Spring '22	Jun. '22	Dec. '22	Spring '23		
English 1	Zero on ECR	9%	10%	11%	50%		
	Approaches or Above on EOC	21%	16%	22%	33%		
	Number of Retesters	*82,755*	*6,932*	*125,320*	*95,551*		
English 2	Zero on ECR	7%	9%	9%	62%		
	Approaches or Above on EOC	28%	8%	25%	28%		
	Number of Retesters	*59,617*	*44,687*	*89,918*	*64,857*		

Due to the change in ECR questions and introducing a new possible score of zero on the rubric, Spring 2023 has higher zero rates than Spring 2022. No change in scoring had occurred.

However, our test equating process (see slide 19) means that while individual items can be easier or harder in a given year, the mix of item difficulty is balanced across years, so overall performance on the test stayed the same or increased.

Percent of Retesters Who Received a Zero Score on the Extended Constructed Response Item, Percent of Retesters Who Achieved Approaches Grade Level or Better on the EOC as a Whole, and Number of Retesters

4 The variation in zero rates seen during the move to hybrid scoring are much smaller and are considered normal.

		Pre-STAAR Redesign			Post-STAAR Redesign		
		Human Scoring					Hybrid Scoring
		Spring '22	Jun. '22	Dec. '22	Spring '23	Jun. '23	Dec. '23
English 1	Zero on ECR				**50%**	**68%**	**72%**
	Approaches or Above on EOC				**33%**	**20%**	**32%**
	Number of Retesters	*82,755*	*6,932*	*125,320*	*95,551*	*53,654*	*106,926*
English 2	Zero on ECR				**62%**	**68%**	**84%**
	Approaches or Above on EOC				**28%**	**19%**	**24%**
	Number of Retesters	*59,617*	*44,687*	*89,918*	*64,857*	*34,859*	*76,763*

Individual items (like ECRs) can be easier or harder in a given year, so we expect to see some variation in ECR zero rates across tests

However, the mix of item difficulty is balanced across years through equating (see slide 19). Overall performance on the test has stayed largely consistent

Percent of Retesters Who Received a Zero Score on the Extended Constructed Response Item, Percent of Retesters Who Achieved Approaches Grade Level or Better on the EOC as a Whole, and Number of Retesters

4 For transparency purposes, TEA provided LEAs with additional December 2023 English 1 and English 2 ECR results information.

Unlike Spring STAAR tests, which are released annually, June and December STAAR tests are not released so items can be re-used in later tests. To support transparency, TEA developed a two-step process in March to give districts more insight into student performance on December 2023 ECRs:

Step 1 (444 LEAs): Upon request, TEA provided frequency distributions of reasons for an ECR receiving a score of 0 within the requesting district –

Statewide E1/E2 ECR Score of 0 Breakdown	
14%	**No response**
30%	**Not scored due to a condition code** (not enough words, duplicated text, written in another language, consisting mostly of copied text from the passage, or writing is off-topic or off-task)
55%	**Received a score of 0 according to the rubric** (eligible for Step 2)

Step 2 (109 LEAs): After Step 1, LEAs can also opt to schedule an appointment to view the responses in person

Only responses that receive a score of zero are made available for viewing;

Received positive feedback from district personnel on this process

Thank you!

- More details on the hybrid scoring study performed on 2023 STAAR data can be found in "Assessment Reports and Studies" under the section titled Additional Reports and Studies.
- If you have questions about hybrid scoring, contact TEA using the Student Assessment Help Desk.

Appendix

ECR Rubric Example: English 1

English I
Constructed Response Scoring Guide

English I Passage with Extended Constructed Response

Informational Writing Rubric

Score Point	Organization and Development of Ideas
3	• **Controlling idea/Thesis is clear and fully developed** The controlling idea/thesis is clearly identifiable. The focus is consistent throughout, creating a response that is unified and easy to follow. • **Organization is effective** A purposeful structure that includes an effective introduction and conclusion is evident. The organizational structure is appropriate and effectively supports the development of the controlling idea/thesis. The sentences, paragraphs, or ideas are logically connected in purposeful and highly effective ways. • **Evidence is specific, well chosen, and relevant** The response includes relevant text-based evidence that is clearly explained and consistently supports and develops the controlling idea/thesis. For pairs in grades 6 through EII, evidence is drawn from both texts. The response reflects a thorough understanding of the writing purpose. • **Expression of ideas is clear and effective** The writer's word choice is specific, purposeful, and enhances the response. Almost all sentences and phrases are effectively crafted to convey the writer's ideas and contribute to the overall quality of the response and the clarity of the message.
2	• **Controlling idea/Thesis is present and partially developed** A controlling idea/thesis is presented, but it may not be clearly identifiable because it is not fully developed. The focus may not always be consistent and may not always be easy to follow. • **Organization is limited** A purposeful structure that includes an introduction and conclusion is present. An organizational structure may not be consistent and may not always support the logical development of the controlling idea/thesis. Sentence-to-sentence connections and clarity may be lacking. • **Evidence is limited and may include some irrelevant information** The response may include text-based evidence to support the controlling idea/thesis, but it may be insufficiently explained, and/or some evidence may be irrelevant to the controlling idea/thesis. For pairs, evidence is drawn from at least one of the texts. The response reflects partial understanding of the writing purpose. • **Expression of ideas is basic** The writer's word choice may be general and imprecise and at times may not convey the writer's ideas clearly. Sentences and phrases are at times ineffective and may interfere with the writer's intended meaning and weaken the message.

English I
Constructed Response Scoring Guide

1	• **Controlling idea/Thesis is evident but not developed** A controlling idea/thesis is present but not developed appropriately in response to the writing task. • **Organization is minimal and/or weak** An introduction or conclusion may be present. An organizational structure that supports logical development is not always evident or is not appropriate to the task. • **Evidence is insufficient and/or mostly irrelevant** Little text-based evidence is presented to support the controlling idea/thesis, or the evidence presented is mostly extraneous and/or repetitious. Explanation of any evidence presented is insufficient and may be only vaguely related to the writing task. For pairs in grades 6 through EII, evidence is drawn from only one text. The response reflects a limited understanding of the writing purpose. • **Expression of ideas is ineffective** The writer's word choice is vague or limited and may impede the quality and clarity of the essay. Sentences and phrases are often ineffective, interfere with the writer's intended meaning, and impact the strength and clarity of the message.
0	• A controlling idea/thesis may be evident. • The response lacks an introduction and conclusion. An organizational structure is not evident. • Evidence is not provided or is irrelevant. The response reflects a lack of understanding of the writing purpose. • The expression of ideas is unclear and/or incoherent. ***Please note that if a response receives a score point 0 in the Development and Organization of Ideas trait, the response will also earn 0 points in the Conventions trait.***

Score Point	Conventions
2	Student writing demonstrates **consistent command** of grade-level-appropriate conventions, including correct: • sentence construction • punctuation • capitalization • grammar • spelling The response has few errors, but those errors do not impact the clarity of the writing.

English I
Constructed Response Scoring Guide

1	Student writing demonstrates **inconsistent command** of grade-level-appropriate conventions, including limited use of correct: • sentence construction • punctuation • capitalization • grammar • spelling The response has several errors, but the reader can understand the writer's thoughts.
0	Student writing demonstrates **little to no command** of grade-level-appropriate conventions, including infrequent use of or no evidence of correct: • sentence construction • punctuation • capitalization • grammar • spelling The response has many errors, and these errors impact the clarity of the writing and the reader's understanding of the writing.

More RLA scoring resources can be found here: https://tea.texas.gov/student-assessment/testing/staar/staar-reading-language-arts-resources

SCR Rubric Examples: US History and Grade 8 Science

United States History Prompt

Prompt: What was President Theodore Roosevelt's Big Stick policy **AND** what was an example of this policy?

Item-Specific Rubric

Score: 2

Response includes specific details in reference to a description and one example:

Description:

- The Big Stick policy used military readiness and diplomacy to protect the Western Hemisphere from foreign intervention.

Examples:

- Roosevelt used this diplomacy to restrain European countries from threatening Latin American countries.
- Roosevelt issued this policy to enforce the Monroe Doctrine and become the international police power of the Americas.
- The United States increasingly used force to justify intervention in several countries, including securing the Panama Canal Zone, Cuba, Nicaragua, Haiti, and the Dominican Republic.

Score: 1

Response provides only half of the correct details.

Score: 0

Does not provide a response, or the response is incorrect or irrelevant.

Grade 8 Science
Short Constructed Response Scoring Guide

Grade 8 Science Short Constructed Response

Prompt

Sodium sulfate (Na_2SO_4) is used to produce many products.

Which elements are represented in the formula **AND** how many atoms of each element are represented in the formula?

Read the question carefully. Then enter your answer in the box provided.

Item-Specific Rubric

Score: 2

The student response includes:

- There are a total of 7 atoms representing three elements in the formula

AND

- The elements are sodium (2 atoms), sulfur (1 atom), and oxygen (4 atoms)

Score: 1

The student answers half of the question correctly.

Score: 0

The response is incorrect or irrelevant.

More social studies scoring resources can be found here: https://tea.texas.gov/student-assessment/testing/staar/staar-social-studies-resources
More science scoring resources can be found here: https://tea.texas.gov/student-assessment/testing/staar/staar-science-resources