MCAS Constructed Responses for High School Biology

Welcome to this presentation on Scoring MCAS Constructed Responses for High School Biology.

My name is Steve Long, and I am a member of the Science Test Development Team at DESE.

This presentation is geared towards classroom teachers and curriculum coordinators who are interested in learning more about how constructed response items are scored. It was originally shared as webinar on February 1, 2024.

In addition to the PowerPoint presentation, you will want to print out the participant packet which includes an overview of each item, along with the scoring guide, scoring notes, selected anchor papers, and the student responses for you to score. This packet will be referred to throughout the presentation.

This presentation will provide a brief overview of how items are developed and scored, focusing primarily on how constructed responses are scored.

There will be a short scorer training on two of our Biology items, one about chicken combs and natural selection, and the other about plants and insects performing photosynthesis and cell respiration.

Then, you will have an opportunity to score some actual student responses and compare your scores to those made by the scoring team.

We will wrap up the presentation by sharing some web resources.

This life cycle diagram traces the item development process from a new item to an operational item.

Along the way, each new item goes through multiple layers of review: including two educator committees selected from districts across the commonwealth, the Assessment Development Committee, shown in purple, which reviews and edits item content and the Bias and Sensitivity Committee, shown in blue, which reviews items for bias and sensitivity issues.

Once an item has been accepted by both educator committees, it goes into a Field Test Eligible bank.

Before the item is field tested, it gets reviewed by content “experts” to identify any potential content issues and by our publications team to improve clarity and grammar.

After field testing, Constructed Response items are benchmarked, where we use actual student responses to finalize the scoring notes and to select papers for training scorers.

Once benchmarking and scoring are complete, the two educator committees review students’ performance on each item.

If the item statistics are poor or if one of the answer options pulled too many responses, the item can be either be rejected or re-edited and sent back to the beginning of the life cycle for review.

Once an item has been accepted, it goes into the Operational Eligible Bank where it can be selected to appear on a future test administration as a scored item.

It generally takes about 2 to 3 years for a new item to work its way through the life cycle from a newly written item to an operational item.

In constructed response items, students are asked to apply a concept, use evidence from data, explain their reasoning, or show their how they reached a conclusion.

This slide shows two examples of how students are tested: one on the left from the computer-based test, labeled CBT and one on the right from the Paper-based test, labeled PBT.

The high school Biology test contains both 3-point and 4-point CR items; the 3-point items are embedded into modules, and 4-point items are standalone items.

Even though CRs are organized into parts, all responses are scored holistically, meaning that partial credit may be awarded even if a response for one or more of the parts is incomplete.

Some things that are NOT scored on science tests include:

Spelling, grammar, or punctuation, as long as scorers can understand what is being said

Incorrect statements that are above what is expected at that grade level

Extra information that is true but that does not answer the question. However, students should be instructed to minimize adding additional information, since very long or rambling responses often contain errors or information that is not correct.

For questions that ask for a calculation

If they are using an answer from a previous part of the item that they solved incorrectly, then students typically can still earn credit by applying that information to subsequent parts of the item, as long as they show their work, and subsequent calculations are correct

For students to earn full credit, they must answer all of the parts of the question, including explaining their reasoning. In addition to explaining their reasoning, constructed response items may ask students to analyze a graph or data table or to identify an error in a model and explain how to correct it.

Scorers can only score what is stated in the response, so the response must be complete to earn full credit.

Students often provide only an identification or a partial response. If only an ID is asked for, they do NOT need to give an explanation, but if the item asks for an explanation, then it must be included for full credit.

Each summer, DESE & Cognia test developers and scoring staff hold “benchmarking” meetings where we review and discuss a wide range of student responses to identify responses that provide clear examples of each score level [called anchor papers] as well as responses that provide novel or alternative ways to answer the questions [called practice papers]. Notes are added to each of

these responses to show how each part of the item was scored, so these annotated papers can be used during scorer training.

In addition, we revise the Scoring Notes, which were created during the item review process, to make sure that the notes clearly describe: the different types of responses can receive credit for each part of the item, the number of points that are assigned to each part of the item, and any scoring rules for holistic scoring.

Scorer training packs include:

Three "anchor" papers that show the range of responses within each score level.

One or more "practice" papers for each score level that show different ways students may earn credit and/or exemplify certain rules for earning credit.

Scoring notes and the Scoring Guide, which is the rubric that is released with the item.

Scoring leaders and item scorers are selected based on their content expertise. The scoring leaders also attend benchmarking discussions, so they are familiar with the scorer training materials and the rationale for scoring decisions.

Each scorer attends an in-depth training session and must pass a qualification test where they score actual student responses on the item. If they fail to qualify, they are not allowed to score that item.

In addition to the initial qualification, the accuracy of each scorer is assessed throughout the scoring process using checks built into the scoring process including comparing each scorer's assigned scores: to other scorers who scored the same response, to previously-scored responses that are embedded into their scoring set, and to the scores assigned by the scoring leaders.

If any of these checks show that a scorer is not scoring accurately, that scorer is removed from the process and all the responses they scored for that item are rescored.

During this session, we will review the question, the scoring guide, and the scoring notes. We will also review one anchor paper representing each of the five score levels.

Then you will independently score five student responses.

Keep in mind that, due to the time constraints of the presentation, you will see a much smaller sample of responses than the actual scorers would receive during their scorer training.

The first item assesses students' knowledge of evolution, and more specifically of natural selection. It also touches on the overlap between natural selection and genetics.

Take a moment to review the item. You may want to use the printed item on page 2 of the participant packet for taking notes.

"This item is scaffolded going from an easier identification in Part A to a more difficult explanation in Part C. Green underlines have been added to the slide to highlight the key elements of each part of the item."

In part A, students are expected to apply the inheritance information to identify the genotype of a chicken with a pea comb.

In Part B, students are expected to apply the information about heat loss to compare the allele frequencies for chickens in a cold climate.

In Part C, students are expected to apply their knowledge of natural selection to explain how the change in allele frequencies could occur.

As you look over the item be sure to notice that:

The alleles, the inheritance pattern, and the phenotypes are all provided in the background information and in Part A.

Students also learn in Part B that chickens with pea combs have an advantage in cold climates because the pea comb reduces heat loss.

It is tempting to want to credit a student for providing this information, but this information must be used as part of a complete explanation to be scorable. Simply repeating the prompt is not creditable.

Constructed response items can often be aligned to more than one content standard or science practice. Discussing how to best align an item is part of the work of the content ADC.

This item was aligned to the natural selection standard since most of the points assessed this standard, but Part A could also have been aligned to the genetics standard (HS.LS.3-3).

This item was aligned to practice C, Evidence Reasoning and Modeling, because students are expected to construct an explanation in Part C and use reasoning in Parts A and B.

The Scoring guide is released with each item. It provides an overview of what knowledge and skills are expected in a complete response.

Each score represents a different level of understanding: thorough, general, limited, or minimal.

A zero-score response does not demonstrate the knowledge and skills being assessed by the item.

The scoring notes are used in training scorers. The scoring notes highlight key elements of each part of the item.

The Part A response only needs to identify the correct genotype for a chicken with a pea comb.

Part B asked for a comparison of the allele frequencies. The brackets in Part B show information that is not required to earn credit for that part of the item. Responses do not need to explicitly mention “the H allele” as long as the comparison is clearly implied.

Part C asks for a full explanation of natural selection occurring in the population. Note that the “OR” between responses in Part C shows that there is more than one way to correctly answer the item [can address advantage of h allele or disadvantage of H allele as long as the comparison (h>H or H<h) is correct].

The underline in notes is used to emphasize a key element of the response (i.e., to achieve a 4 score, the response must address passing the trait to offspring).

“This is the 4-score anchor paper, so all parts of the item are completed correctly.”

Part A correctly identifies the hh genotype.

Part B correctly describes that the h allele will be more frequent in chickens in a cold environment.

The response does not need to explicitly refer to the H allele because the comparison is implied by the “more frequent” part of the response

Part C response includes increased survival, reproduction, and passing trait to offspring [You can see that we accept synonymous terms, such as using “next generation” to refer to offspring]

“This is the 3-score anchor paper, so it contains most of the information that the item was designed to assess. However, most 3-score responses either contain an error or are missing a key element from the scoring notes.”

Part A correctly identifies the hh genotype for 1 point.

Part B correctly describes the h allele having greater frequency than the H allele in a cold environment for 1 point. [the response a nice explicit comparison of allele frequencies]

Part C: Response receives 1 point for explaining that chickens with a pea comb would experience increased survival, but the response is missing passing the trait to offspring. [see scoring note “For scores of 4, response must explicitly address passing on of genes/traits.”]

“This is the 2-score anchor paper.”

Part A correctly identifies the hh genotype. (1 point)

Part B correctly describes that the h allele will be much higher than the H allele in a cold environment. (1 point)

Part C: Response is missing the explanation of how natural selection could produce a change in allele frequency. The response in Part C is primarily a repeat of information provided in the item without adding any explanation. (0 points)

“This is the 1-score anchor paper. Generally, a 1 score response shows only minimal understanding of the concepts being assessed: in this case, only the genotype ID.”

This response earns 1 point in Part A for correctly identifying the genotype.

Part B is simply a repeat of the information provided without any comparison of allele frequencies, and

Part C does not explain how natural selection could result in more chickens with pea combs. [The statement “helps them reduce the amount of heat loss” is just a repeat of prompt without any explanation]

“This is the 0-score anchor paper, so it will either contain multiple errors or have no creditable responses.”

In Part A, the response provides two genotypes, only one of which would have a pea comb, so it is not creditable.

Part B fails to compare allele frequencies, which is required by the scoring notes.

Part C fails to explain natural selection at all.

Now it is your turn to score some actual student responses using the information from the scoring notes, anchor papers, and scoring guide we shared with you.

As a reminder, actual scorers see many more anchor and practice papers, along with having to correctly score a qualification set of responses.

As you read the student responses, you may find it helpful to mark up the papers to indicate where a student earns points.

If you want to practice your own scoring, it would be best to score all five of the practice papers before resuming the presentation. I will be providing a detailed scoring summary of each of the practice responses in the next set of slides.

Response A received a score of 2.

Parts A correctly identifies the hh genotype for 1 point.

Part B correctly describes the h allele as being more frequent than the H allele in a cold environment for 1 point.

Part C fails to explain how natural selection would result in the h allele becoming more frequent in the population (i.e., no survival advantage or reproducing and passing of traits to offspring).

Also, the “produce offspring with the pea comb” statement in Part B sounds Lamarkian (i.e., as if chickens could just choose to produce offspring with a specific trait with the intent of preventing heat loss).

Response B received a score of 4.

Part A correctly identifies the hh genotype. (1 point)

Part B provides an example of an alternative way to answer the question by focusing on the frequency of the H allele being less than the frequency of the h allele. (1 point)

Part C provides a very thorough coverage of survival advantage, producing more offspring, and passing on genes. (2 points)

Response C received a score of 1.

Parts A correctly identifies the hh genotype for 1 point.

Part B is not creditable because it does not compare the allele frequencies, only that there would be more chickens with pea combs, which was provided in the prompt.

Part C fails to address how natural selection would lead to more chickens with pea combs.

Response D received a score of 3.

Parts A correctly identifies the hh genotype for 1 point.

Part B correctly describes the H allele frequency being less than the h allele frequency in a cold environment for 1 point.

Part C addresses the survival advantage for 1 point, but it does not fully explain natural selection (i.e., no passing trait to offspring).

Response E received a score of 2.

The response earns 1 point in Part A for correctly identifying the genotype and 1 point in Part B for describing the h allele as more frequent.

Part C fails to explain how natural selection could lead to the higher frequency of the h allele. (0 points)

Instead, the response "would want to adapt" suggests a Lamarckian view of natural selection (i.e., that chickens would chose the h allele because they want to adapt to their environment), which represents an error.

The second item asks students to distinguish between two cellular processes in plants and insects that cycle carbon dioxide and oxygen.

Take a moment to review the item. You may want to use the printed item on page 18 of the participant packet for taking notes.

This item is scaffolded going from an easier identification of processes in Parts A and B to a more difficult analysis and explanation in Part C. Green underlines have been added to highlight the key elements of each part of the item.

Part A asks students to identify the process that is only performed by plants, which is photosynthesis.

Part B asks students to identify the process that is performed by both plants and insects, which is cellular respiration.

Part C asks students to analyze a data table to determine the contents of each flask based on changes in the concentration of gases (O_2 & CO_2) in the flask.

For a complete response, students are expected to use the data (i.e., actual numbers OR increases/decreases in gases) to support their claims about the contents of each flask.

As we discussed earlier, constructed response items can often be aligned to more than one content standard or science practice.

This ADC aligned this item to the photosynthesis standard, though it could also have been aligned to the cellular respiration HS.LS.1-7 standard.

Since the item asks about the concentrations of both O_2 & CO_2 it was considered too broad for the carbon cycle standard.

Although the ADC decided to align this item to Practice B (Math & Data), it could also have been aligned to Practice A or Practice C.

Practice A: Investigations and Questioning; since Part C asks for information about a scientific investigation.

Practice C: Evidence, Reasoning, and Modeling; since Part C also asks students to explain their reasoning.

The Scoring guide is released with each released item.

Each score represents a different level of understanding: thorough, general, limited, and minimal.

A zero-score response does not demonstrate the knowledge and skills being assessed by the item.

These scoring notes are used for training scorers. The scoring notes highlight key scoring elements of each part of the item.

Part A asks students to identify the cellular process that is only performed by plants, which is photosynthesis.

Part B asks students to identify the cellular process that is performed by both plants and insects, which is cellular respiration.

Part C asks students to analyze a data table to determine the contents of each flask based on changes in the concentration of gases (O_2 & CO_2) in the flask.

The flask in row 1 contained no organisms because the gases had little or no change in concentration.

The flask in row 2 contained one insect because cellular respiration caused O_2 to decrease and CO_2 to increase.

The flask in row 3 contained two plants and one insect because the plants would perform photosynthesis causing O_2 to increase and CO_2 to decrease.

For a score of 4, students are expected to support their claims about the contents of each flask using the data table (i.e., actual numbers OR increases/decreases in gases) AND identify the correct cellular process in each flask.

For scores of 3 and below, if Parts A and B correctly identify the cellular processes, then responses in Part C do not need to include processes in the explanation.

For scores of 2 and below, responses can earn 1 point for correctly identifying the contents of the flasks without an explanation.

Since 5 score points are possible for this item, there is a conversion where 5 points = 4 score, 4 points = 3 score, 2-3 points = 2 score, and 1 point = 1 score.

This is the 4-score anchor paper, so all parts of the item are completed correctly.

Part A correctly identifies photosynthesis for 1 point.

Part B correctly identifies cellular respiration for 1 point.

Part C correctly identifies the contents of each flask, provides the changes in oxygen and carbon dioxide from the table that were used to support the responses, and identifies both cellular processes. (3 points)

5 points = 4 score

This is the 3-score anchor paper, so it contains most of the information that the item was designed to assess.

Part A correctly identifies photosynthesis for 1 point.

Part B correctly identifies cellular respiration for 1 point.

Part C correctly identifies the contents of each flask and provides the changes in oxygen and carbon dioxide from the table that were used to support their responses for 2 points.

Although the response is nicely written, the cellular processes are missing from the explanation. Therefore, this response cannot score higher than a 3 [see the first scoring note on p.19].

4 points = 3 score

This is the 2-score anchor paper, so it will demonstrate a limited understanding of photosynthesis and cellular respiration.

Part A correctly identifies photosynthesis for 1 point.

Part B correctly identifies cellular respiration for 1 point.

Part C the response misinterprets the data by arguing that oxygen and carbon dioxide are being produced, even though the level of oxygen is declining in row 2, and the level of carbon dioxide is declining in row 3. (0 points)

2 points = 2 score

This is the 1-score anchor paper. Generally, a 1 score response shows only minimal understanding of the concepts being assessed, in this case correctly identifying only one of the processes.

Part A correctly identifies photosynthesis for 1 point.

Part B incorrectly identifies photosynthesis. (0 points)

Part C nothing is creditable because the contents of the flasks are not correctly labeled. (0 points)

1 point = 1 score

This is the 0-score anchor paper, so it will either contain multiple errors or have no creditable responses.

Nothing creditable. The cellular processes in Parts A & B are misattributed to the wrong sources (i.e., insects do not perform photosynthesis).

The Part C response mislabeled the contents of the flasks w/o explanations.

Reminder that this is a fairly common 0 score (i.e., all parts of the item were attempted, but nothing is creditable).

0 points = 0 score

Once again, it is your turn to score some actual student responses using the information from the scoring notes, anchor papers, and scoring guide we shared with you.

As a reminder, actual scorers see many more anchor and practice papers, along with having to correctly score a qualification set of responses.

As you read the student responses, you may find it helpful to mark up the papers to indicate where a student earns points. Remember to apply the scoring conversion found at the bottom of the scoring notes on page 19 of the participant packet.

If you want to practice your own scoring, it would be best to score all five of the practice papers before resuming the presentation, since I will be providing a detailed scoring summary of each of the practice responses in the next set of slides.

Response A received a score of 2.

Part A correctly identified photosynthesis for 1 point.

Part B correctly identified cellular respiration for 1 point.

Part C correctly identified the contents of each flask with no explanations.

Since this response already earned a 2-score based on Parts A & B, it cannot receive credit for correctly identifying the contents of each flask. Therefore, the overall score remains a 2.

2 points = 2 score

Response B received a score of 3.

Part A correctly identified photosynthesis for 1 point.

Part B correctly identified cellular respiration for 1 point.

Part C correctly identifies the content of each flask (for 1 pt) and provides data to support responses (for 1 pt), However, the response does not include both cellular processes in the explanation, so it cannot be a 4 score.

4 points = 3 score

Response C received a score of 2.

Part A correctly identified photosynthesis for 1 point.

Part B correctly identified cellular respiration for 1 point.

Part C correctly identifies the contents of the three flasks for 1 point, but the explanations contained errors [no oxygen in flask 1] or were incomplete [only mentioning one of the two gases and lacking the cellular processes].

3 points = 2 score

Response D received a score of 4. All parts are creditable.

Part A correctly identified photosynthesis for 1 point.

Part B correctly identified cellular respiration for 1 point.

Part C correctly identifies the contents of each flask and provides an explanation that includes changes in gases and both cellular processes for 3 points.

5 points = 4 score

Response E received a score of 1.

Parts A and B are both incorrect.

However, in Part C the response correctly identified the contents of the three flasks for a "minimal" 1 score. [see scoring note about 3 correct IDs]

1 point = 1 score

Thank you for participating in this scoring practice presentation. We hope that you learned something about the scoring process that will be helpful to you and your students.

This slide has a list of links to our website that you may find useful.

If you should have any additional policy questions such as test designs or accommodation or general questions, please reach out to the Department by email or phone.

If you need logistical support, including tech support on the testing platform, please contact the MCAS Service Center.

On behalf of the Department of Elementary and Secondary Education, we thank you for viewing this presentation on Scoring Constructed Response items for the Biology MCAS.