

## MCAS Grade 8 Science and Technology/Engineering Constructed Response

Welcome to our presentation on the MCAS Grade 8 Science and Technology/Engineering Constructed Response Items. My name is Sarah Boehm, and I am a member of the STE test development team at the Massachusetts Department of Elementary and Secondary Education. This presentation is geared toward 6<sup>th</sup>, 7<sup>th</sup>, and 8<sup>th</sup> grade educators as well as science curriculum coordinators in Massachusetts. The presentation was originally shared as a webinar on February 15, 2024 by members of the STE test development team.

In addition to the PowerPoint presentation, you will need to access the participant packet, which includes the items with their associated rubrics, score notes, anchor papers, and responses for you to score. This packet will be referred to throughout the presentation.

During today's session we will provide an overview of the MCAS test development and scoring process, analyze student work samples from two released items, individually score student responses, and review additional resources that are available on the Department's website.

Before we look at student responses from two released items, we are going to look at how questions, which we refer to as items, end up on an operational test. The process outlined on this slide takes about two years.

This process has many layers of review, including two educator committees.

The Assessment Development Committee, shown in purple, includes educators teaching science in this grand band as well as science coaches and curriculum coordinators.

The Bias and Sensitivity Committee shown in blue, includes a wider range of people involved in education in Massachusetts.

We're now going to move along the graphic starting in the top left corner when the items are initially written.

When items are first reviewed by educator committees, the educators provide edits to the graphics, wording, context, and layout to ensure they align with the standards and that both the language and presentation will be accessible and fair to students.

The items that are accepted or accepted with edits continue on to the FT eligible pool, then they are reviewed by two content "experts," and are edited by a publications team to improve clarity and grammar.

Items are field tested on the Spring MCAS test and then are scored. CR items will be benchmarked. Benchmarking is a meeting where we finalize the scoring notes and select anchor papers to help with scoring. You'll learn more about this process in this presentation.

After scoring, the educator committees see the final version of the items along with the field test data showing how students performed on each item. At this point, items can no longer be edited so educators accept or reject each item.

If accepted, items go into the operational eligible pool and could appear on a future test as an operational item that counts toward the students score.

We are going to focus on constructed response items for the rest of this presentation.

The G5 & G8 tests have 2 and 3-point CRs, some of the 3-pt items are embedded into modules, which are a group of questions all about the same scenario.

All constructed responses items are scored holistically with partial credit given if partial knowledge is demonstrated. We have a few examples of scoring rules in the items we will look at in this presentation.

In constructed responses, students are frequently asked to explain their reasoning, use evidence from data, or show their work; all of which assesses students' application of the science and engineering practices.

The images here show a computer-based CR with boxes for students to type their answers on the left and on the right there is a different item in the paper-based format. We can add graphics to the response box on the paper version to make it more similar to the computer-based version.

Because we are trying to assess students' understanding of the science standards and application of the science and engineering practices, there are errors students can and do make that do not affect their score.

This includes errors in spelling, grammar, and punctuation, as long as we can understand the student's intent.

Students also sometimes correctly answer an item but include an extra incorrect statement that is above what is expected in the standards. As long as their grade-level knowledge is clearly demonstrated, they will still earn credit.

Additionally, credit is not impacted if students include extra information that is true and does not contradict their correct answer.

This last bullet mainly applies to G8 and High School tests when calculations are expected.

If they are using an answer from a previous part that they solved incorrectly, then they **typically** can still earn credit as long as they show their work and it is correct.

Students who earn full credit demonstrate science content knowledge and application of the science and engineering practices by answering all parts of the question clearly, as scorers can only score what is stated in the response.

We sometimes only ask for an identification in one part to scaffold a question. When this is the case students are only expected to give the ID, they do NOT need to give an explanation.

However, we often ask students to explain their reasoning, or to use data from a table or graph to support their answer. Sometimes we ask students to complete a model, or ID an error in a model and how to correct it. To get full credit a student needs to follow these instructions for what to include in their response.

Now I'm going to describe the benchmarking process that we complete each summer after CR items are field tested.

Test developers from DESE and our contractor as well as scoring staff spend days together during the benchmarking meetings where we review lots of student responses and discuss how the items should be scored. We go into benchmarking with draft score notes, but students almost always find unique ways to answer questions that we had not anticipated.

As we review responses, we look for ideas students had that are valid but not reflected in the score notes.

We have discussions about science content errors students make and if that error is within grade level expectations or above.

We have discussions about whether an incomplete response is enough to show minimal understanding and a score of 1.

And we have discussion about whether less detailed reasoning to a part should be creditable at lower scores, like maybe at a 2 score but not a 3 score.

Our meetings are often full of debates because we're trying to ensure the scoring is fair and consistent even when student responses are not clear-cut.

We want to make sure the scoring for an item matches the general guidelines provided in the scoring rubric we release with each item.

So, during BM we revise Scoring Notes and make sure they include:

Expectations of what students can write to receive credit.

How points are assigned for each part.

Scoring rules for holistic scoring.

We also annotate the student responses to explain why certain responses are getting certain scores.

Training packs are put together that include:

Score notes (that I just described) and the Score Guide, which is the rubric that is released when an item is released.

Three "anchor" papers for each score level to show the full range of responses at that score--remember we cannot have a 1.5 score, all scores must be whole numbers, so we have a range of responses within one score.

Practice papers for each score level—which may exemplify a certain scoring rule or show an alternative way of answering.

After we benchmark an item, the scoring packs are given to scoring leaders who have content expertise.

Discuss qualifications of scorers

The scorers go over all the anchor papers and practice papers and discuss why each response was scored in a specific way. After the training, the scorers take a test, called a qualification set, to ensure they are scoring properly. If they do not earn a satisfactory score, then they are re-trained and take a different qualification set. If they do not pass the second time, then they do not score the item.

In addition to this initial qualification, scorers must continue to score accurately throughout the whole process. There are checks in place such as embedded responses and also “read behinds” by scoring leaders; if at any time a scorer is not scoring accurately, they are removed from the process and all the responses they scored for that item are rescored.

As I mentioned before, we have two items that we are going to review today. These are also in the packet posted as a resource with this presentation.

We will read through each question, score guide, and scoring notes. Then we will review an anchor paper at each score point. This is just a small portion of responses that the actual scorers would receive for their training.

Then each of you will independently score a set of student responses.

Our first item is a physical science CR. This is a 2-point item.

Please pause this presentation and read through the item, which is also on page 2 of the packet.

This item goes to a 7th grade Physical Science standard (Construct and interpret data and graphs to describe the relationships among kinetic energy, mass, and speed of an object.) Within the clarification statement I'd like to point out that at this grade level students are expected to "Consider relationships between kinetic energy vs. mass and kinetic energy vs. speed separate from each other" and that the formula for KE is not expected.

This item is aligned to Practice category B: Math and Data. Students are working with data including mass and speed and they are using mathematical thinking to interpret the graphs.

Many items are truly aligned to multiple practices (and CRs sometimes to multiple content standards). For reporting purposes, we assign one standard and one practice to each item.

This item also contains some practice category C with constructing explanations and communicating information.

This score guide is released along with the item. It is on page 3 of your packet.

Each score represents a different level of understanding: a score of 2 shows a thorough understanding and a score of 1 shows a partial understanding.

A response given a zero score does not demonstrate knowledge of the content or skills being asked in the question.

These score notes are used by the scoring team to determine the score for each student response.

Score notes are on page 3 of the packet.

The score notes reflect different ways or wording the student can use that are acceptable for credit.

Information in brackets is there to help the scorers but is not required for students to earn full credit.

So, in Part A, we are looking for students to write: Student R has a greater amount of kinetic energy, because (and an explanation). The response may include the bracketed part: [the students are riding at the same speed] which would be part of a full and complete answer but is not required to earn the point.

The bullets in Part A show different ways students can explain their reasoning and get credit – they can state that student R has more mass or use data to show that. The third bullet shows how a student could use the formula to calculate KE. This is above grade-level expectations, but we include it here to make sure credit is given if a student gives this explanation.

In Part B we are looking for students to identify Graph 3 and explain that as kinetic energy of an object increases its speed increases.

Now let's look at some anchor papers for this item. Anchor papers start on page 5 of the packet.

Take a moment to read this student's response. Pause the recording as needed.

Part A earns credit for student R has a larger amount of KE because they have a higher mass.

Part B earns credit for identifying Graph 3 and describing the relationship between speed and KE.

Ignore "gradually" in Part B, note that KE increases faster than speed due to "exponential relationship" as discussed in standard 7.MS-PS3-1, which is why the graph is curved. This response earns full credit.

Please read the 1-Score Anchor paper.

Part A – the student shows a misconception that speed is the only factor in KE. Part A does not earn a point.

Part B – earns 1 point for a correct part B.

Please read the 0-Score Anchor paper.

In Part A incorrectly compares the KE of the students.

In Part B the wrong graph is chosen.

This is a zero paper – the student made an effort but has some misconceptions.

If you're not already using the mini training pack, then I encourage you to take it out now. This will let you flip between responses and the score notes as well as mark up the responses as we go.

It's your turn to score some student responses using the score notes, anchor papers, and score guide we shared with you. As a reminder, actual scorers get many more anchors and practice papers, along with a much longer training.

Response A is on page 9 of the mini training pack. Please pause this presentation and take the time you need to score the response.

Part A shows a misconception that ONLY speed and not mass affects kinetic energy.

The answer in Part B is clear and correct.

This response earns a score of 1.

Read and Score Response B. Pause the recording as needed.

Both parts are clear and correct

This earns a score of 2.

Next up is response C. Please read and score this response.

In Part A the student uses the data to show the Student R has a larger mass.

Part B has a longer answer that goes into more detail than the other responses we have seen. This student correctly names graph 3 and provides an explanation.

This is another response that earns 2 points.

Please read and score Response D.

In Part A the response incorrectly says more weight will slow the bike down.

In Part B the wrong graph is chosen, with reasoning that the graph should be a flat line since the bike path is level.

This response does not receive a point and has a 0-score.

Please score response E now. This is the last of the responses for this CR.

This response has a full and complete part A that uses comparative language and the data and includes that the speed is the same.

In Part B the wrong graph is chosen and shows confusion about KE when at rest.

This response earns a score of 1.

Our second CR is a 3-point life science item about Cownose Rays.

This is a module item, so it was within a set of 4 items with a common stimulus. The stimulus is on the left side of this slide and the question is on the right side. In the packet, these appear on pages 16 and 17.

Please pause the recording to read the stimulus and item.

This item is aligned to the 7<sup>th</sup> grade life science standard about how periods of abundant and scarce resources affect the growth of organisms and the size of populations.

Part A of this item also goes to 7.LS2-2 with the relationships between organisms in an ecosystem. For reporting purposes, we can only tag the item to one standard.

This item is aligned to practice category C.

To complete the item, students must use the food web, which is a model, to examine relationships among organisms and use that information to make and support a claim about how a change in one population would affect other populations in the ecosystem, which is constructing explanations.

This score guide is released along with the item. It is on page 18 of your packet.

A response that gets a score of 3 has demonstrated a thorough understanding of the effect of scarce resources on the size of populations in an ecosystem. A response that earns 2 points has shown partial understanding, and a 1-score has shown minimal understanding.

The scoring notes highlight some of the most common ways that students answered the question and earned credit.

Score notes on pages 18-19 of their packet

In Part A: We are looking for an ID of predatory with an explanation. The capital “OR” means students can say one or the other.

In Part B: We are looking for an ID of decrease with an explanation. The bullets show 2 different ways to explain. Remember the information in the brackets may be included in a full answer but is not needed to earn credit.

In Part C: There are several ways students can describe what would happen to the hammerhead shark. The slash marks in the indicate that there are multiple ways to explain the relationship (only/biggest/main food source). We try to capture the most common ways students answer the question in our score notes, but Scorers also accept synonyms to the words here.

At the bottom of the score notes there is a 0-1 rule. Many items have 0-1 rules that kick in when a response did not earn a point in any part, so would score a 0, however some understanding is shown. On this item a student can earn one point for correct IDs in each part without any explanation. So, a student who wrote predatory in Part A, decrease in part B, and decrease in Part C would earn one point. We will see an example of this in the student work.

Now let’s read the anchor papers for our Cownose Ray item. Anchors are on pages 21-24 of your packet.

Note that only the graphic in the stimulus and the Parts from the questions are shown in the slides for spacing reasons. Some responses still need two slides to fit everything.

Here on Anchor Score 3, we have predatory and cownose rays prey on oysters for a point in Part A. The identification part of the answer has a double underline and the explain part has a single underline in these annotations.

Part B: Correctly describes the population as decreasing because phytoplankton is a food source. We ignore the word “almost” in “food source to almost all of the Cownose’s food sources”. The food web shows phytoplankton as a food source to all those organisms, but this is acceptable wording for credit.

Part C: the student earns credit for the description and explanation here. “falter” is an acceptable synonym for decrease.

3 correct IDs with explanations is worth 3 points.

Please read the Anchor Score 2 response.

This response earns credit in Part A with a clean and clear answer.

This response provided the incorrect description of the population of rays in Part B, so it does not earn a point even though the relationship to phytoplankton is correct.

In part C the response earns credit for saying the sharks would “die off” with an explanation from the food web.

This paper earns 2 points.

Please read the Anchor Score 1 response.

This response earns credit in Part A.

In Part B, the student does not show an understanding of how changes to one population would affect other organisms in the ecosystem.

The response does not address “what would happen to the hammerhead shark population” in Part C and the description given is not supported by the food web shown.

So, this response earns only 1 point for the correct answer in Part A.

Now let's look at the Anchor paper for a Score of 0.

This response contains 3 incorrect IDs, and also misinterprets the arrows in the food web. This response does not show an understanding of how changes to resources affect population sizes, so it receives 0 points.

Now it's your turn to score some student responses.

We will work through the responses one at a time before moving on to the next response. Take out your packet so you have access to the item, score notes and anchor papers.

Read through response A and give it a score. This screen only shows Part A and B, but you will find the entire response on page 26 of your packet. Pause this recording as needed.

Part A earns 1 point for naming predatory as the type of relationship and explaining that oysters are a food source for the rays.

No credit in part B. Stating the population will stay the same is incorrect.

In Part C a correct response about sharks dying off with an explanation about the food source changing earns 1 point.

Response A would score a 2 with credit in Parts A and C.

Now score Response B, pausing as needed.



Response B has three incorrect IDs with faulty explanations, so this response earns a 0.

Next up is Response C.

Response C has 3 correct IDs without any explanations.

Sometimes we look holistically at a response that does not earn full credit in any one part. We call this a “0-1” rule that will bring a response from a 0 score to a 1 score. On this item we decided that three correct IDs but no explanations (or vague explanations) show a “minimal” understanding about population sizes.

So, Response C earns a score of 1.

Response D takes two slides. Please refer to the version in your packet to score this one.

Part A earns 1 pt for a correct relationship and explanation.

Part B earns 1 point. The response describes how all of the cownose ray’s prey eat phytoplankton, so a decrease in phytoplankton will decrease the cownose ray’s food, leading to starvation.

Part C does not get credit, the response does not explain what would happen to the hammerhead shark population. Even though their logic is sound, it does not answer the question that was asked.

Response D earns a point in Parts A & B, so this response would receive a 2 score.

Response E is the last of our student work samples. Please take a moment to read and score this one.

Parts A and B: each earn a point for correct ID with explanations.

Part C: 1 pt here. We accept the misspelling of “extinct” since the meaning is clear with the explanation about sharks starving without food.

So, Response E received a score of 3.

On the department’s website, you can find:

MCAS headlines and links to the MCAS site

Our main STE test development and design webpage, which has links to many resources.

Student work samples, which are released for each released constructed-response item.

Released questions - This link is to the computer-based release of the questions, but we also release a paper “released item document,” which is linked to from our main STE test development page under additional resources.

Practice tests and tutorials.

MCAS Training page which includes links to register for other trainings run by our team.

And a link to information about the pilot for a new test design for the G5 and G8 STE assessment

If you have policy questions, like about the test design or accommodations, please reach out to DESE via email or phone.

Please contact the MCAS Service Center for questions about logistics, like technology support on the testing platform, ordering materials, and reporting.

On behalf of the Massachusetts Department of Elementary and Secondary Education, I would like to thank you for viewing and participating in our presentation on Grade 8 STE MCAS constructed-response items.