

Grade 5 Science and Technology/Engineering Constructed Response Items

Welcome to our presentation on the MCAS Grade 5 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 3rd, 4th, and 5th grade educators and science curriculum coordinators in Massachusetts. The presentation was originally shared as a webinar on February 13, 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 to 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 benchmarking 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 life science CR that was released in 219. 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 4th grade Life Science standard about animal and plant structures and their functions.

This item is aligned to Practice category C. Evidence, Reasoning, and Modeling.

Students are using the picture and intro text as evidence to explain how lynx structures help it to survive.

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 also on page 3 of the packet.

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

In Part A we see the different options around fur color or pattern and different ways to describe the forest habitat.

In Part B, the bullets show different acceptable descriptions for the function of claws.

Information in brackets is there to help the scorers but is not required for students to earn full credit. In the second bullet of Part B the bracketed wording [by predators] would be part of a full and complete answer but is not required to earn the point.

The note in Part A describes how a response could show a partial understanding without fully answering the prompt. 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, if a student misses part B and has a partial answer to A, they can earn a 1 score. We will see an example of this in the student responses.

Anchor papers start on page 5 of the packet.

Take a moment to read this student's response.

Green = creditable student response

Red = wrong

Green = creditable

Part A – does not name a trait of the lynx and does not show understanding of a trait or camouflage.

Part B – earn 1 point for a correct part B

0-Score Page 7 Booklet 1283440

This is a zero paper – the student repeated the prompt in Part A without answering the question. In part B the response does not show an understanding of how a structure of an animal, in this case claws, can help with its survival. (Repeats part of prompt from Part A with “camouflages”.)

If you're not already using the mini training pack, then I encourage you to take it out now.

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 pages 9 of mini training pack. Please pause this presentation and take the time you need to score the response.

Response A earned a score of 1.

Part A: correct ID, but no explanation

Part B: gets credit for a description of catching prey

Response B is on page 10. Again, please pause the presentation to score the response.

Response B earned a score of 2.

Both parts are correct. Notice we don't mind a lack of capital letters and punctuation.

Response C is on pages 11. Again, please pause the presentation to score the response.

Response C earns a score of 1 using the 0-1 Rule.

Part A only has the explanation, without an ID of a trait.

Part B does not clearly get to survival.

The response gets one point to bring the paper from a 0 to a 1 for the incomplete answer in Part A.

On to Response D. Pause as needed.

Response D is a 0-score paper.

Part A – no understanding of traits

Part B – does not get to how the claws are used.

Response E is our final one for this item.

This is another example of a 2-Score response.

Part A – The score notes only mention “blending in”, and that was a very common way students explained camouflage, but other ways to say the same thing are acceptable. Here the student uses “hiding” and earns credit.

Part B – another way to say the claw is used in defense.

Our second constructed response is a 3-point physical science item. It was released in 2022.

The toy boat item is also on page 15 and 16 of your packet. The item takes up too much space to fit on one slide, so it may be easier to read your copy.

Please pause this presentation and read through the item.

Here is part B and C of the item.

This is the PBT version of this item. The drag and drop in Part A is replaced here with a diagram of the boat and directions for students to label each box either N or S.

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

This item touches on a few standards.

Part A is all about magnets, forces and orientation from standard 3-PS2-3. This is the standard we used for reporting.

Part B gets to 5-PS1-3 and properties of materials.

Part C gets to 4-PS 3-4 and forms of energy.

This item is aligned to practice category C. Students are building a model in part A and constructing explanations in part B and C.

The score guide is always released with a released constructed response. The score guides show how each score represents a different level of understanding, which are bolded in this example.

It is on page 17 of your packet.

Here is how the item was scored.

Part A has two correct ways to show opposite poles facing each other. TEIs (or Technology Enhanced Items) like this can be machine scored or human scored, depending on how we decide to use holistic scoring on the item. This part was machine scored when the item was operational, but today you will be human scoring Part A.

A response gets credit in Part B for including both the description of the boats moving toward each other and an explanation. There are several ways students can explain why.

A response gets credit in Part C for identifying kinetic or mechanical energy along with an explanation about the boats moving. Motion energy was also accepted at the elementary level. At upper levels, this would not be acceptable.

There is a 0-1 rule for this item as well. If a response does not get credit in A but has partial answers in B and C it can earn a score of 1.

Anchors are on pages 19-22 of your packet.

Note that only the Parts from the questions are shown in the slides for spacing reasons.

This response gets full credit which is a score of 3.

Part A shows opposite poles facing each other.

Part B – note that students can say both boats move or one boat moves as long as the direction is toward each other.

Part C identifies kinetic energy and explains that the boats start moving.

On our Anchor 2 paper

Part A has opposite poles facing each other.

Part B has the boats move closer and that magnets attract to iron.

Part C – does not name another form of energy and does not show an understanding of energy conversion. So, no credit is earned here.

The response earns two points for part A and B

On the 1 Anchor paper

Part A incorrectly has like poles facing each other.

Part B is clear and correct.

Part C incorrect ID with no explanation.

This response earns one point from Part B.

The 0 Anchor response does not show an understanding of how magnets work.

Part A shows like poles facing each other, so the boats would not move toward each other.

Part B says the boats would stay in place and not connect, which does not show an understanding of properties of iron.

Part C does not identify another form of energy.

Now it's your turn to score some student responses using the score notes, anchor papers, and score guide we shared with you.

Response A is on pages 24 of packet. Please pause this presentation and take the time you need to score the response.

In part A we see opposite poles facing each other

In part B we see a description of the boats moving toward each other because iron is magnetic.

Part C has an incorrect ID of electrical energy.

This response gets a score of 2.

Response B is on pages 25 of packet. Again, please pause the presentation to score the response.

Part A is correct.

Part B is clear and correct.

Part C A clear ID in the first sentence, but then the second sentence there seems to be a mistake when the student writes "magnetic energy turns into magnetic energy," but then they go back to mechanical energy. So, we just read over that, since there is enough before and after to earn credit and this phrasing does not contradict what is said elsewhere. Also note that at 5th grade, we accept mechanical energy as a substitute for kinetic energy. In high school they will learn that mechanical energy includes both kinetic and potential energy, but that distinction is considered above grade level here.

This is a 3-Score response.

On to Response C. Pause as needed.

Part A earns 1 point.

Part B is incorrect – the student does not know iron is attracted to magnets.

Part C includes a correct ID, but also an incorrect form of energy and no explanation. This does not earn a point.

That means that this is a 1- Score.

Let's look at Response D. Pause as needed.

This response did not earn a point in any of the parts.

Note that saying the boat would sink was a common error we saw in Part B. But a block of iron the same size as a bar magnet is not necessarily heavier than the magnet, so the boat is unlikely to sink.

This is a 0-score response.

Let's look at Response E. Pause as needed to read and score this response.

Part A is incorrect.

Part B and C are both incomplete as they lack the explanation. However, this student is showing some understanding of magnets. For this toy boat item, we have a score note to look holistically at a response to see if it meets the "minimal understanding" in the score rubric. If a student does not earn a point in Part A and has a correct description in B and identification in C, they earn a point which brings the score from a 0 to a 1.

So, using the 0-1 Rule for this item, this response earns a score of 1.

This is our last example of student work. Go ahead and score this one.

Part A and B earn points.

Part C is incomplete as it does not have an explanation, so it does not earn a point here.

This is the type of response you might put a 2+ on as a classroom teacher, but on MCAS we can only give whole number scores. So, this is a 2-Score paper.

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 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 5 STE MCAS constructed-response items.