MCAS High School Introductory Physics Constructed Response

Welcome to our presentation on the MCAS High School Introductory Physics Constructed Response Items. My name is Isadel Saunter, 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 9th and 10th grade physics educators and science curriculum coordinators in Massachusetts. The presentation was originally shared as a webinar on January 31, 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 and overview of the MCAS test development and scoring process, analyze student work samples from two released introductory physics 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 introductory physics items, we are going to take a 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.

I'll start with an overview of the setup of the slide. Each box in the graphic shows one step of the development process.

All items get FT before they count toward students' scores. The graphic shows the steps of the item development process that take place before field testing as boxes with dashed lines around them. This is when items can be edited. There are solid lines around the steps in the process that take place after items are field tested when they can no longer be edited.

You'll notice there are many layers of review.

We have two educator committees. The Assessment Development Committee is shown in purple. There is one assessment development committee per test, which is primarily made up of educators teaching the subject and includes some science coaches and curriculum coordinators. We also have a Bias and Sensitivity Committee shown in blue.

Both committees see all items twice, before and after the items are field tested.

We're now going to move along the graphic in the order that the processes take place.

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

Once educator committees have completed first review, if the items are accepted or accepted with edits, they go into a FT eligible pool, get reviewed by two content "experts" to ensure accuracy, get edited by a publications team to improve clarity and grammar, and then field tested.

After field testing, the items are scored. If they're CR items, they 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 both see the final version of the items and the field test data associated with them at their data review meetings. At this point, items can no longer be edited so educators either accept or reject each item.

If accepted, items go into the CM eligible pool and could appear on any future administration and count toward student scores.

We will spend the rest of time focusing on constructed-response items.

In CR items, students write a response to a multi-part question.

HS Science has 4-point discrete CRs and 3-point items that are part of a module, which is a group of questions all about the same scenario or phenomenon.

CR questions often ask students to explain their reasoning, use evidence from data, or show their work => all of which assess students' application of the science and engineering practices

Items are scored holistically with partial credit given if partial knowledge is demonstrated.

This holistic scoring is often achieved through scoring rules. We are showing one example of a scoring rule today in the first item, but many items have several scoring rules that give students credit at lower scores for showing partial knowledge.

On this slide we show two separate items to illustrate the different ways students respond to CR items on the CBT and PBT versions of the test. The CBT screenshot on the left shows an equation editor box and a text response box. The PBT item shows how in addition to having a blank response space we can add-in background art to keep items more similar to the CBT 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.

For questions that ask for a calculation

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

Lastly, significant figures are not considered on the MCAS STE tests, but when students round, they are expected to do so correctly.

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. We'll see an example of this in the first item we look at today.

However, we often explicitly and clearly ask students to integrate science and engineering practices in one or more parts of a constructed-response question.

For example, we might ask them to "explain your reasoning." Explaining scientific reasoning can be aligned with the SEPs in what we put in Practice Category C for the MCAS. This includes, "constructing explanations and designing solutions" and "developing and using models." Clearly explaining their reasoning is an SEP that many students have trouble with.

We also might ask students to "use data from table" or "from a graph" OR "show your calculations" are aligned with what we put in Practice Category B for the MCAS. This includes "analyzing and interpreting data" and "using mathematics and computational thinking." When students are asked to show their calculations, they are also asked to include units in their answer, except for efficiency calculations. Using incorrect units or not having units in an answer keeps students from earning the highest score point and then does not impact credit at lower scores,

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

Test developers from DESE and our contractor meet along with scoring staff for long all-day meetings to review many student responses for each field-tested item and to discuss how the items should be scored. The scoring staff includes a scoring leader for each item and a lead science scorer who meets about every STE item. We go into benchmarking with draft score notes that were informed by feedback from our educator committees, but students almost always find unique ways to answer questions that are not included in the original score notes.

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

We have discussions about whether students should be able to earn a point for an incomplete response to a part if that would result in their total score being a 1. If so, this would result in a holistic scoring rule.

We have discussions about science content errors students make.

And we have discussion about whether less detailed reasoning to a part should be credible 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. We want to make sure the scoring for an item matches the general guidelines provided in the scoring rubric we release with each item.

During the meeting we adjust the score notes based on our discussion and to make sure they include:

Expectations of what students could write to receive credit.

How points are assigned for each part of the item

And scoring rules for holistic scoring

As a reminder, along with the items, our educator committees review the final score notes again during the fall data review meeting when they see the final version of the item and its associated field test data.

Score notes are not enough information on their own to explain how an item will be scored in all circumstances though.

During benchmarking we also work on creating a score pack for each item, which includes the score notes and also:

Three "anchor" papers for each score level to show the full range of responses at that score.

At least one practice papers for each score level—which show scoring rules and different ways students may earn credit. This sometimes includes less common ways that are not in the score notes.

All of these responses in the score pack also get annotated to explain how each response was scored.

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

Before the packs are used to score all the student responses, scorers are trained. Their training is much longer and more in depth than what we will demonstrate today! 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 entire process. There are checks in place such as embedded responses that have known scores, 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 PowerPoint 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 you will independently score a set of student responses.

Our first item was released in 2022 and is about a roller coaster.

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

We align each item to one standard and no more than one practice category. This item is aligned to the energy conservation standard, HS-PS3-1, however it also overlaps with standard HS-PS3-2. We aligned this item to practice category B, however students are also asked to explain their reasoning which overlaps with practice category C.

Now we're on page 3 of your packet. You've probably seen this score guide as we release them for every item.

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

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

The score guide makes up one of several tools we give scorers, along with the score notes, anchor papers, and practice papers. The packet we put together for you is a miniature example of that.

These are the score notes for the item.

In Part A, we only asked students to identify the point where the gravitational potential energy was greatest. Therefore, only "W" was expected, but it was of course fine if students provided reasoning as well.

In Part B, students were expected to calculate gravitational potential energy using GPE = mgh. They could use g from the reference sheet = 10 m/s^2 or 9.8 m/s^2. When you see capital "OR" in the SN, this means a response would receive credit if they responded either way. Note that typically on the introductory physics MCAS an incorrectly calculated value from an earlier part could be used in a later part and receive credit, so you could carry down a wrong value from Part A to Part B, use it correctly in Part B, and earn credit in Part B. However, in this case a correct calculation of the GPE for the wrong point identified in Part A only received credit for scores going from 0->1. This is noted in the Part B score notes. This was because not understanding that height affected GPE, even after applying the equation that includes height, was judged to be a minimal understanding of the content.

Students were expected to have a correct comparison, like that the kinetic energy was less at point Y than point Z. It would be credible for students to phrase it the opposite way, that the kinetic energy was greater at point Z than point Y. Students were also expected to explain their reasoning either based on the relative height of the points or the relative amounts of GPE.

Students could ID the height at which the KE and GPE were equal either using the labeled point of X or the height of 25 m. There were generally three acceptable explanations, which you can see are bulleted, depending on whether students reasoning included the height or went to GPE converting to KE. In the third bullet we have a backslash between two synonyms. When we see a synonym a lot at benchmarking we may put it in the score notes this way, indicating either is acceptable. In general, synonyms are always acceptable even if they do not directly appear in the score notes, for example the students could talk about energy being converted or transformed in the second bullet. You'll also notice in the third bullet that we have the heights of the lowest and highest points bracketed. The brackets indicate that the information does not need to be included for credit but is included to help the scorers.

At the bottom you see that this item was scored as each part being worth 1 point. Sometimes we use a holistic scoring rubric with more "little points" and a conversion chart over to what score different numbers of little points equal. Unfortunately, we don't have an example of this today. I just want to point out that how scoring notes are set up is very item dependent and influenced by both feedback from our educator committees and the student responses we see at benchmarking.

Lastly, if a response has missing or incorrect units in the answers or if the work is not shown, it cannot receive a score of 4, even if the numerical part of the answer is correct. This is typical of introductory physics MCAS items because an answer is not fully complete without a unit and students are prompted to show their work and include units in their answer.

Now we're on pages 5 and 6 of your pack to review our anchor score 4.

Part A: Correct ID of Point W. Also explains why, which is not required.

Part B: Correctly multiply the mass by the acceleration due to gravity by the height at point W to get the gravitational potential energy. The units are also correct in the answer, which is required for students to earn the highest score, in this case a 4-score. Note that the response does not include units in the work they showed, which does not impact their score.

Part C: Correctly state the car and its passengers have more kinetic energy at point Z. Note this is stated opposite from the score notes, which is fine. The response also has a couple valid explanations. One is that there is more potential energy at point Y than point Z, so there is more kinetic energy at Z than Y. The student also correctly reasons that point Y is higher off the ground. Either of these explanations would have been considered valid alone.

Part D: Correctly IDs both point X and the 25 m height. Either would have been acceptable. There are two correct explanations. One is that the car has moved down halfway from its original point, which clearly means the car is at half its original height. The other is that the potential energy is half as much and the other half was converted into kinetic energy.

The anchor score 3 starts on page 7 of your pack.

Part A: Correct ID of Point W. Also explains why, which is not required.

Part B: Correct calculation with work shown and units.

Part C: Correctly states the car and its passengers would have less kinetic energy at point Y and explains this is because point Y is higher.

Part D: Identifies two heights and is vague about what is equal, perhaps this means the mechanical energy is equal? We must score what is actually written down by the student, so we can't interpret this sentence for them. The response does correctly state that PE converts to KE, but this is not enough for credit as the ID is not present.

Now we're on page 9 & 10 for the anchor 2 score.

Part A: Correct ID of Point W.

Part B: Has the correct answer with correct units and work shown.

Part C: Incorrectly states the kinetic energy is the same at point Y and point Z. Arrives at this solution by using 10 as the velocity at each point. Note that we do typically accept calculations as reasoning on the Introductory Physics MCAS, as long as they are correct. For Part C the students would have had to calculate the kinetic energy at point Y by subtracting the gravitational potential energy at point Y from the gravitational potential energy at point W. We didn't see this often during benchmarking, which is why it is not included in the score notes. However, we did include a paper that both did the calculation and included other valid reasoning in the full pack for the scorers so it would be trained and students would receive credit. This is an example of how the score notes are not complete without the anchors and practice papers, all of them together represent the full range of student responses.

Part D: Incorrect ID with no explanation.

This anchor score 1 starts on page 11 of your pack.

Part A: Correct ID of Point W

Part B: Incorrectly divided the mass by the height at point W.

Part C: Incorrectly states there was more kinetic energy at point Y, with a misconception that there is more kinetic energy when the car is at a greater height.

Part D: Not responsive. Seems to just be discussing that the height of point Y is closest out of the other points to point W. Also repeats the knowledge shared in Part A, that point W has the greatest potential energy.

This zero anchor starts on page 13 of your pack.

Part A: Incorrectly IDs point Y.

Part B: Incorrect GPE, multiplied 5 x 8

Part C: Incorrect comparison, divides the 40 m calculated in Part B by 8 and comes back out with the 5.

Part D: Incorrectly identifies point W and then divides the mass by the height at W to justify this.

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.

This response is on pages 16-17 of the mini training pack. Please pause this presentation and take the time you need to score the response.

This response was given a score of zero.

Part A: Wrong point identified.

Part B: In Part B the student attempts to calculate the gravitational potential energy for Point Y, which they identified in Part A. However, the student does not multiply by 10 m/s^2 as is required. Therefore, the student does not earn the potential 0-> 1 credit for a correct calculation in Part B of the GPE of the point identified in Part A.

Part C: Incorrectly uses the heights in place of velocity in the KE equation.

Part D: Incorrect ID and reasoning.

This response is on pages 18-19. Again, please pause the presentation to score the response.

This response was given a score of 3.

Part A: Correct ID of GPE being greatest at point W

Part B: Correct calculation of the potential energy at point W and with the correct final unit. There is a unit error when the givens are being listed, but this does not affect the student's score and would not at any score point as the unit in the answer is correct.

Part C: Correctly compares the kinetic energy as less at point Y with reasoning that the height of point Z is lower.

Additionally, the student states that point Y "has decreased in speed." The interval over which the speed decreases is not provided so we decided this statement was vague and we ignored it.

Part D: Does not answer the question by comparing the KE at point Z with the PE at point W. Then goes on to make an untrue statement about how KE is high when PE is high.

This response is on pages 20-21. Again, please pause the presentation to score the response.

This response was given a score of 1.

Part A: Correct ID of point W. Also correctly explains this is due to it being the greatest height, though an explanation was not asked for.

Part B: Incorrectly leaves off the acceleration due to gravity from the equation, so gets and answer that is a factor of 10 too small. Also has the wrong units, but this would only keep students from earning a 4 score.

Part C: Not responsive to the question and includes a misconception that the acceleration depends on the height.

Part D: Doesn't answer what is asked.

This response is on pages 22-23. Again, please pause the presentation to score the response.

This response was given a score of 2.

Part A: Y is incorrect because it does not have the greatest height and GPE depends on height.

Part B: The student correctly calculates the GPE of point Y, but this can only get credit going up to a 1-score. Note that the units of FN are not correct either, but Newton meters would of course be acceptable instead of joules.

Part C: Correctly said the kinetic energy at point Y was lower. And provided two different correct reasonings, because it's at a greater height and because most of the energy is GPE.

Part D: Correct ID of 25 m height with reasoning that it's halfway down from the highest point.

This response is on pages 24-25. Again, please pause the presentation to score the response.

This response was given a score of 4.

Part A: Correctly ID's point W.

Part B: Nicely shows their work, including writing down the givens, and has the correct answer and unit.

Part C: Correct comparison with acceptable reasoning of point Y has a bigger height.

Part D: Correctly identifies point X, with acceptable reasoning of it being the half of the original height.

Our second item was also released in 2022 and is about a series circuit.

Please pause this presentation and read through the item, which is also on page 27 of the pack.

This item is aligned to standard HS-PS2-9, which is our circuits standard.

It is also aligned with practice category B, though students again use scientific reasoning and interpret the model of a circuit diagram, so there is some alignment with practice category C.

The score guide is on page 28 along with the score notes. Again, this was released. We always release a score guide with each released CR. The score guides show how each score represents a different level of understanding, which are bolded in this example.

These are the score notes for the item.

Part A: In Part A students were asked to identify component K and explain its function. A correct response identified component K as a switch and included at least one of the bulleted explanations included in the score notes.

Part B: In Part B, students were expected to calculate the total resistance of the series circuit by adding the resistors together. We show the calculation here, but it would be equivalent for students to explain in words that they added each resistance and to give the 30-ohm answer. We showed units in our work, but students are only required to include units in their answer to receive full credit.

Part C: In Part C, students needed to identify that the current flowing was the same through R1 and R2 and explain their reasoning. Different explanations are again shown with bullets.

Part D: In Part D, students needed to calculate the voltage drop across R1 as 2 V. They could do this by applying Ohm's law to find the current through the circuit and then applying Ohm's law again to find the voltage drop across R1.

Or they could use proportional reasoning, R1 has 1/6 of the total resistance so the voltage drop is 1/6 of the total voltage. Of course, it would be fine if students discussed Kirchoff's loop rule here, but that is above grade level and not expected.

This item also has each part being worth 1 point, but I'll remind you that different items can be scored differently.

And we see the typical Introductory Physics note, that students can only earn the highest score if all calculations and units are included and correct.

Now we're on pages 30-31 of the pack for anchor score 4.

Part A: Correct ID of switch. Correct explanation of a switch opening and closing the circuit. Also discusses how current will stop flowing when the circuit is open.

Part B: Correctly added the resistances of each resistor in a series circuit and included correct units.

Part C: Correct ID of the current being the same with explanation that it's a series circuit.

Part D: Correct use of Ohm's law to first find the current and then calculate the voltage drop across R1.

Anchor 3 is on pages 32-33 of your pack.

Part A: Correct ID of switch. Two correct explanations given, to turn the circuit on and off and to control current flow.

Part B: Correctly added the resistances of each resistor in a series circuit.

Part C: Incorrect ID of current not being the same in a series circuit. Confusion between current and voltage in a series circuit.

Part D: Correct solution using proportional reasoning. The voltage drop across R1 is 1/6 of the total voltage because R1 has 1/6 of the resistance.

On pages 34-35 of the mini training pack

Part A: Correct ID of switch with correct reasoning that it can open or close the circuit.

Part B: Correctly adds the resistance of each resistor.

Parts C & D: ID that current is not the same in a series circuit is incorrect in Part C. This response shows a misconception that the voltage drop is 12 V across each resistor, which is stated in Part D.

The student applies this reasoning to Part C with Ohm's law to reason that the current is not the same.

On pages 36-37 of the mini training pack

Part A: Incorrectly identifies component K as a resistor and shows a misconception that resistors stop the flow of current.

Part B: Correctly adds the resistances together to get the total resistance.

Part C: Incorrect ID of more current will flow through R2, with the misconception that the resistance is the current.

Part D: Attempts to use Ohm's Law but uses the total resistance as the current.

On pages 38-39 of the mini training pack

Part A: Incorrectly identifies component K as a short circuit.

Part B: Adds all resistors in a series circuit, but then divides by 3, as if they are trying to find the average resistance. This is incorrect.

Part C: Incorrectly tries to apply Ohm's law and incorrectly concludes that resistors with different resistances cannot have the same current flowing through them.

Part D: Seems to correctly understand the voltage drop across all three resistors will be 12 V, but incorrectly splits the voltage drop evenly across the resistors even though they have different resistances.

It is again your turn to score some student responses using the score notes, anchor papers, and score guide we shared with you. Don't worry if you score some papers differently than we did. Remember that actual scorers receive a much more in-depth training, have three times as many anchors to reference, have practice papers to reference, and are monitored to ensure they are scoring responses correctly.

This response is on pages 41-42 of the pack that goes with this presentation. Please pause this presentation and take the time you need to score the response.

This response was assigned a score of 2.

Part A: Correctly IDs component K as a switch and has two correct reasonings, that a switch turns the circuit on and off and that a switch allows the current to flow. Notice that there was a missing word in the first reasoning and a spelling error in the second, both of which were ignored.

Part B: Correctly adds the resistance of each resistor. Note that the units are included up top, even though they are not at the end of their calculation. So, Part B would be credible at any score.

Part C: Incorrectly IDs not the same current through each resistor. Has a misconception that current decreases with each resistor.

Part D: Incorrectly divides the total voltage by the number of resistors. Then dives the total voltage by the resistance of R1. Then multiplies both values.

This response is on pages 43-44. Again, please pause the presentation to score the response.

This response was assigned a score of 4.

Part A: Correctly IDs switch and explains it opens and closes the circuit.

Part B: Correctly adds the resistance of each resistor.

Part C: Correct ID and explanation that it's a series circuit.

Part D: Correct calculation using Ohm's law. First finds current through circuit and then calculates voltage drop across resistor 1.

This response is on pages 45-46. Again, please pause the presentation to score the response.

This response was assigned a score of 0.

Part A: incorrectly states component K lets excess energy out

Part B: incorrectly divides the total resistance by 3

Part C: incorrectly IDs that the current is not constant because the resistances are different

Part D: incorrectly multiplies the voltage by the resistance of R1

This response is on pages 47-48. Again, please pause the presentation to score the response.

This response was assigned a score of 1.

Part A: Incorrect, K is not a motor.

Part B: While this response is a bit sloppy with the equal signs, it is clear they are adding up the resistances of each resistor to get the total resistance, so with the inclusion of units we would have given it 1 pt. at any score. However, since this is 1-score they also could have only given us the numerical answer and received credit.

Part C: Wrong ID, not credible.

Part D: Incorrectly divides the voltage by the resistance of R1.

This response is on pages 49-50. Again, please pause the presentation to score the response.

This response was assigned a score of 3.

Part A: Identifies component K as a switch and explains the function as turning the circuit on and off.

Part B: Calculates the total resistance as 30 ohms and both shows the correct calculation and additionally explains the calculation in words.

Part C: Incorrectly says the current changes in the series circuit and justifies with ohm's law, with the misconception the voltage drop across each resistor is the same.

Part D: However then correctly calculates the voltage drop across R1 using ohm's law. This is an acceptable amount of work in this item. Even though the total current calculation is not shown, Ohm’s law is written down and the total current is clearly labeled and then used correctly when Ohm’s law is applied again.

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 link to from our main STE test development page under additional resources.

Practice tests for each test and tutorials.

The MCAS training page, which has registrations for all trainings offered, including this webinar before it was given.

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 Introductory Physics MCAS constructed-response items.