# What Happens When Large-Scale Items Actually Use the Computer's Capabilities?

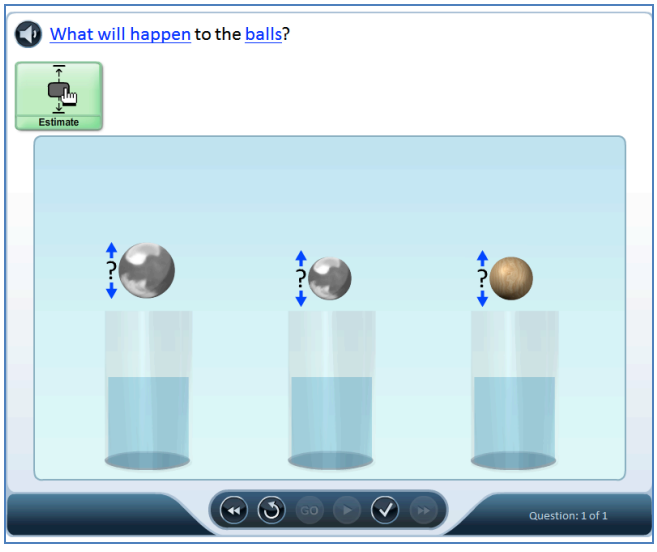
## *Exploring Issues and Redefining Challenges*

Rebecca Kopriva, University of Wisconsin
David Gabel, Center for Applied Linguistics
James Bauman, Center for Applied Linguistics

**Issue 1:** Using computer capabilities gives us ways of getting at what students know that are different from what we can know from multiple choice and even to some extent from constructed response test items. The quality of information may be more on a par with what we derive from live performance assessment, though on a more reduced scale. The following item pairs will exemplify this claim. Note that the interactive item in each pair is designed to take approximately the same amount of time as would a simple or somewhat complex constructed response item.

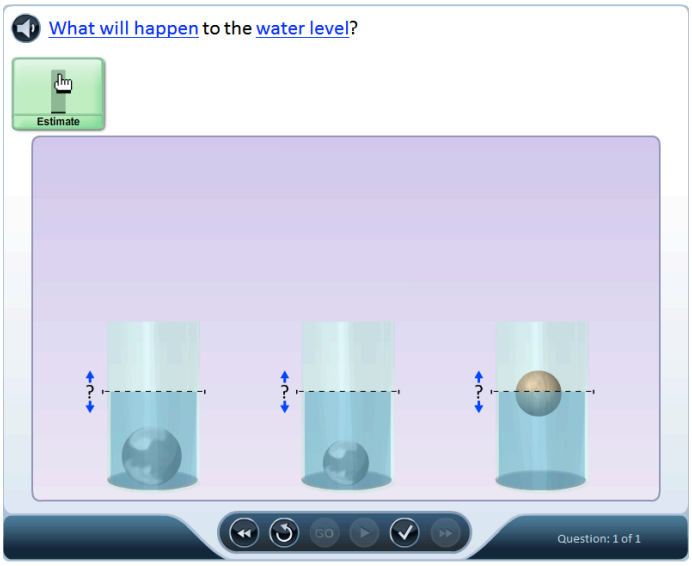
### *ONPAR Item: Buoyancy*

**Target**: Determine the relative position of objects in water and the resulting water displacement based on the object's density and volume.




This screen is preceded by an animation showing the three balls placed on a platform suspended over the beakers. The platform is then removed. In the first of two question and response scenes shown above, students roll over the balls to determine that the metal balls are solid and the wood ball is hollow. They move the balls up and down to show their relative position in the water. In the second question and response scene (shown below) students drag the water level

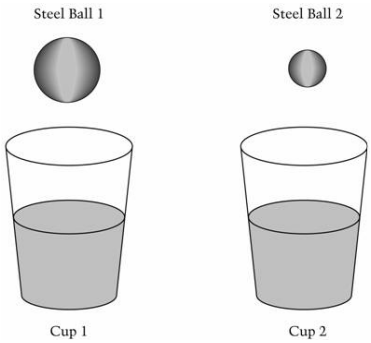
to a position that reflects the position of the balls as they placed them in the previous scene.




**ONPAR approach:** In the first scene students must compare the properties of each of the balls, and must determine that it is density and not size that determines where the balls will go in the water. A comparison must be made between wood and metal of the same size and between metal of different sizes. In the second scene students demonstrate their knowledge about the resulting water displacement in each container. The answers from the first scene are carried over to the second scene. Students must compare relative water displacement.

## *Traditional Item: Buoyancy*

As shown in the picture below, Christina has two identical cups that are filled to the same level with water. She also has two solid steel balls.




1a) Christina puts ball 1 in cup 1 and ball 2 in cup 2. In which cup will the water level rise the most?

Tell why you think so.

1b) Christina has another ball that is the same size as ball 2, but this ball is made of wood and is hollow. If she put this hollow ball in one of the cups, do you think the water level would rise more or less than it would if ball 2 were put in the cup?

Tell why you think so.

**Traditional item approach:** In question 1a) students compare two steel balls of different sizes and indicate which water level will be the highest and then provide an explanation. In question 1b) students compare a wood ball and steel ball of the same size and then indicate which water level would be the highest and provide an explanation.

## *Differences*

The ONPAR item asks students to compare three different balls as they show their position and the water level displacement. However, the traditional item asks students to compare only in pairs of two.



Kopriva, R.J, Gabel, D., & Bauman, J. (2009)

Whereas the traditional item asks students to explain why the water levels change, the ONPAR item asks students to show why the water levels would be different.

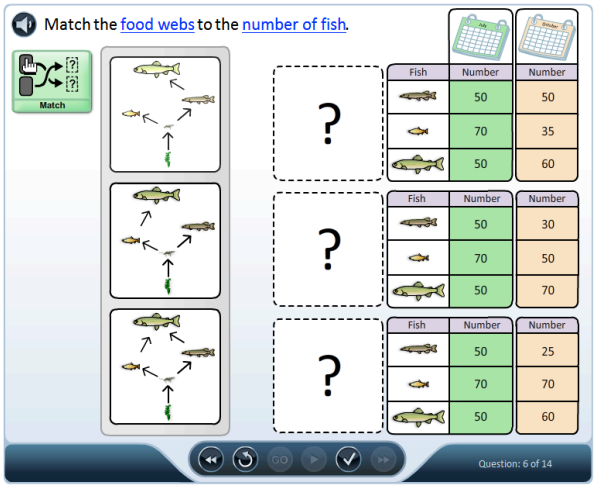
### *Justification*

The ONPAR item asks students to interact with the screen elements and engage in the experiment, as compared to their more indirect relationship with the content in the traditional item. Additionally, in ONPAR the students are *demonstrating* their conceptual mastery, maintaining a depth of knowledge of two for the subtle comparisons based on several factors and demonstrated knowledge of cause and effect relationships. The traditional item asks students to explain but, depending on their meta-cognitive abilities and their proficiency with language, their responses may or may not represent the true sophistication of their knowledge. Both items appear to be getting at a similar level of science complexity, albeit in different ways but, it can be argued, in more direct and less direct ways.

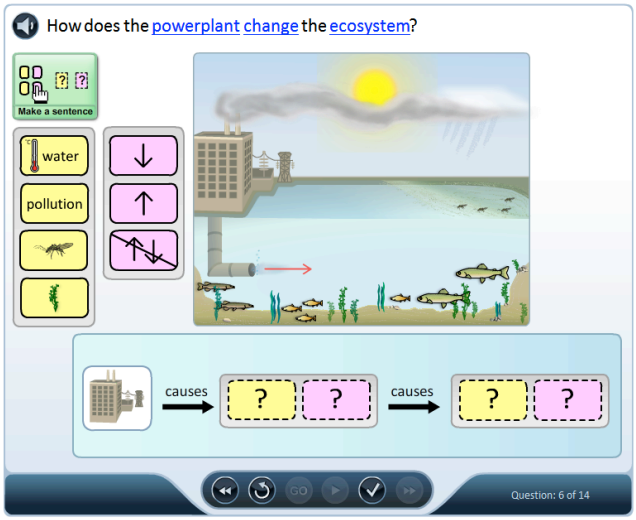
## *ONPAR Item: Pond Ecosystem*

**Target**: Explain how changes in the environment can cause ecosystems to change over time.

The pond ecosystem item includes several screens, including an initial animation that shows the pond and the surrounding environment, and subsequently a man adding a large bucket of big fish into the pond. The next animation shows three months elapsing. The two screenshots below show the third and fourth screens in the item. These are where the target questions are asked and students interact with the item to respond.




In this scene students must drag the different food webs that could produce the data shown to positions that correctly match the fish population changes taking place three months after the increase in the big fish population.

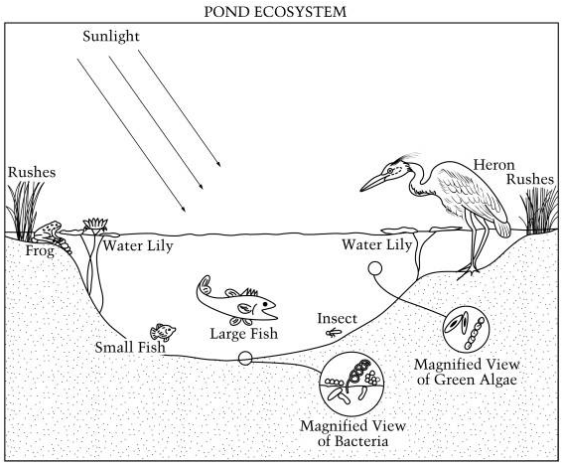



In this scene students create a "causal chain" by dragging yellow (light background)response options on the left to the yellow targets, and pink response options (shaded background) on the left to pink targets to put together a well formed proposition. Rollovers help clarify response meanings.

**ONPAR approach:** In the first scene students must analyze different plausible food webs and predict how the populations will change by analyzing trends in a data table. In the second scene, students create causal chains to show how the ecosystem changes based on the introduction of the power plant.

## *Traditional item: Pond Ecosystem*

The picture below shows a pond ecosystem. Use this picture and what you know about the things in it to answer the questions in this section.




1a. Suppose that one spring a new type of large fish was put into the pond. So many were put in that there were twice as many fish as before. By the end of the summer, what would happen to the large fish that were already in the pond?

Explain why you think these new large fish would have this effect.

1b. If a rainstorm washed some fertilizer from a nearby field into the pond, what would happen to the algae in the pond system after one month? Why do you think the fertilizer would affect the algae this way?

**Traditional item approach:** In the first question, students must be able to state and explain the effect of adding an additional and potentially competing fish to the ecosystem. In the second question, students must explain how and why an introduced pollutant would affect the algae in the ecosystem.

## *Differences*

In the first of the two question screens the ONPAR item asks students to analyze several related food webs and predict population outcomes. For students to respond to this screen successfully over four different scenarios, cumulative correct responses would seem to indicate they understand the concept of how variations in food chains cause the increase or decrease of related populations in the ecosystem. This screen can be scored for partial knowledge as well as for the most sophisticated understanding.

While the traditional item does not state that the populations would be affected, it is an obvious choice. The explanation response is scored for students' causal knowledge in this situation, but unless the answer is rich in detail, it may be less clear when students have a partial understanding of this kind of relationship.

The second question (traditional) asks students to determine how and why fertilizer affects algae. The response to this question would continue to confirm or question the student's grasp of the relational connections. The next question screen in the ONPAR item asks students to make a two step causal inference. First they must state how the power plant introduces a possible pollutant and then state how that change in pollution level will affect an organism in the ecosystem. Students have the ability to both select the pollutant and choose the organism it affects. In this way the second question screen for ONPAR goes beyond the traditional item. However, item writers felt that because the ONPAR item constrains the choices here, it is generally getting at a similar level of cognitive complexity.

### *Justification*

In the first scene the ONPAR item maintains the cognitive difficulty of the traditional item by requiring students to analyze three distinct food webs and to predict the population effects over time. Students must compare and analyze the population tables to notice trends. While the traditional item asks for the effect of the new large fish on the old large fish, the ONPAR item asks how both the large fish and the small fish populations will change over time. In this way the ONPAR item asks students to demonstrate several connected variables, and demonstrate the interconnectedness of food webs within the ecosystem. The second scene (ONPAR) and the second question (traditional) both ask about how external factors change the ecosystem. While the traditional item is open-ended, it focuses students on the relationship between fertilizer and algae. The ONPAR item maintains the complexity by allowing students to choose how changes producing a set of intermediary effects themselves cause changes leading to tertiary effects.

**Issue 2:** Now what?

a) What are the implications of collecting different kinds of observations for different students?

b) Over time, what are the implications in the general assessment of changing item types or different approaches to measuring the same latent constructs?
e.g. Mislevy's argument in ECD—seems plausible, BUT right evidence must be collected….(and what is THAT?)

**Issue 3:** What are the implications associated with the cognitive demands that are being activated when using different approaches to measuring the same latent construct? There appear to be at least three:

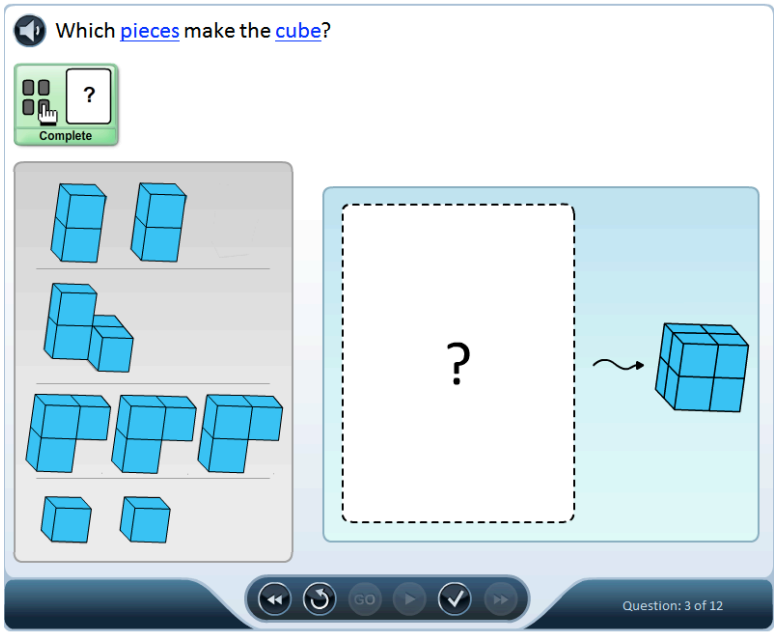
1. **Directness to latent construct**

   Directness indicates how close the task demand is to the target. For example, if the target includes the ability to summarize data in a chart, an indirect task would be to select a chart from

four choices, but a more direct task would be to actually fill in the chart.

## *ONPAR Example: Build a Cube*

**Target**: Predict and validate the results of partitioning a cube.




This screen is preceded by an animation which shows a 2x2x2 cube being built from 1x1x1 cubes. The purpose of this is to both activate the schema of constructing a cube as well as to be explicit about the construction of the object (i.e. so there are no gaps). In the scene shown above, students drag pieces of the cube into the response space. They can drag up to six objects into the response space. There is intentionally no information given about how many pieces are needed to make a cube. This increases the opportunity for students to create a possible outcome. There are several correct answers to this question.

In this item students have to think about which blocks would form a cube. They can devise one of several strategies depending on how much background knowledge they have and how sophisticated they are in representing that knowledge, including counting blocks and mentally manipulating the blocks. Then students can verify their answer by combining the blocks with some mental manipulation required. While students can "hold" one piece over another to get a sense of whether it will fit, this still requires some manipulation to verify the fit. Students will have to devise a strategy to verify the fit of the 3 to 4 pieces required to build the cube. This item is being scored by not only looking at the final result but also tracking the students' approach to creating their final model and making assumptions about the students' cognitive strategies based on how they go about solving the problem.

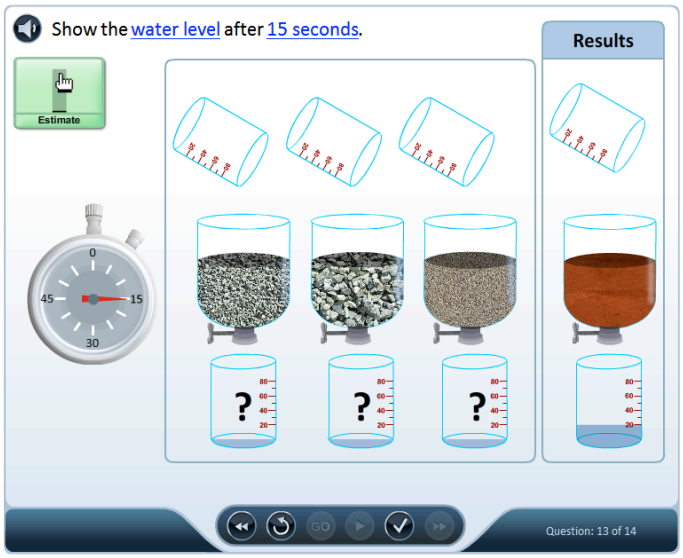
## 2. Cognitive density

As defined here, cognitive density is the amount of cognitive activity that students have to deal with to understand the question, solve the problem, and display their answer. The number and type of cognitive connections need to be considered as well as the complexity of the elements that the student has to connect. The richness of the item environment adds to the density of the relevant—that is, performance items would tend to be more 'dense' than multiple choice items. Density can be good or bad and has relevant (pertaining to the science) and irrelevant cognitive aspects. The goal would be to minimize the dissonant density aspects (construct irrelevant aspects that are not acting as facilitators of the item demands) and modulate the relevant aspects to reflect the targeted cognitive complexity the item is designed to elicit.

### *ONPAR Example: Earth Materials*

**Target**: Predict results from percolation experiments conducted on different Earth materials.

In this item students must follow an experiment. Because of the complexity of the study the item writers were careful to provide enough context and clearly focus the students on the primary objects associated with the target, while not overwhelming them with non-target elements that could add to the cognitive density. The first set of screens in the *Earth Materials* item shows animations associated with identifying the substances and demonstrating the porosity of the substances as the same amount of water moves through the materials in different amounts of time. Then students are asked some questions related to this experiment. For the second section of the item, another experiment is conducted with the same materials. The first screen shows an animation of water pouring through sand for 15 seconds and then shows the flow stopping, at which time the experimenter sees how much water has passed through. The next screen, shown below, shows the results of the previous screen at the far right and then asks students to estimate the amount of water that would go through the other three substances. Detail about the texture of the materials is provided at the beginning of the item, and both experiments follow conventional experimental procedures.

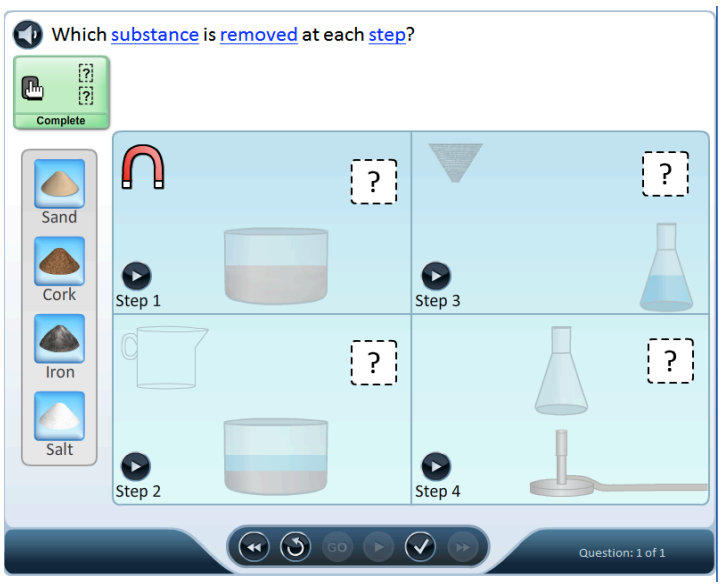



### 3. How target cognitive schemas are engaged

Cognitive schemas can be engaged with wet labs, animations, static images and text. It needs to be thought through under what circumstances one choice is better than another. How the test taker engages the schema can impact the demand in intentional and unintentional ways. As such, item writers have to intentionally select the proper approach that best reflects the target concept, and as appropriate, focus the test taker by intentionally placing some elements in the background or foreground as they relate to the construct. Always using the richer environment isn't always better; what is desired is that the item should match the intended complexity of the targeted construct and the cognitive activity of the student it is expecting to elicit.

#### *ONPAR Example:* **Mystery Substances**

**Target**: Use experimental data about characteristic properties of matter to compare and identify different substances.




This screen, at the end of an experiment, asks students to identify the substances manipulated during an animated experimental procedure. Students can replay each of four different separation techniques portrayed and must correctly identify which substance was removed at each step of the procedure.

In contrast, the traditional item of this pair explains the experiment to students, rather than having them engage in it. It also uses an abstract text-based description explaining how a process is used to identify substances X, Y, Z and W. There are no graphic depictions of the experiment or the materials.

To address this target construct, the item writers felt that the ONPAR item more directly engages the student and gets at the latent conceptual knowledge of the student, whereas the traditional item seems to use abstract depictions which may alienate the test taker. The reliance on written text to depict a rich environment may mask how well the student is able to distinguish materials with different properties. However, another item focused specifically on identifying substances from a simple memorized list of properties might find a text-based approach superior to an animated approach.

**Issue 4:** What are the implications when probability of correct response changes dramatically?

a) When is this a problem and when is it not?

b) When is equating across approaches a viable approach and when is it not?

c) Examples:

(1) More constrained opportunities for response in traditional multiple choice items with a set number of options, as compared to interactive item types

(2) Less constrained opportunity for response in some constructed responses items as compared to some interactive item types that choose from selections to show causality,

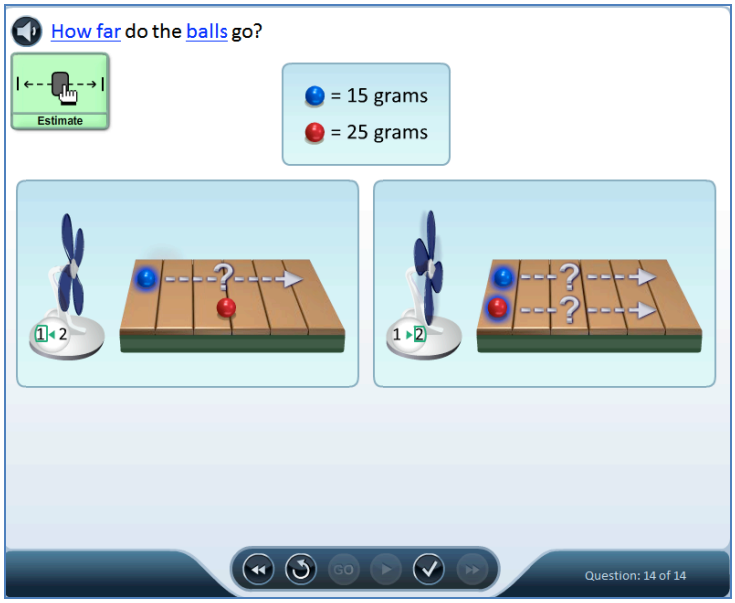
form logic sequences, assemble components, etc.

(3) There can be cases where the interactive type may be more similar to the open-ended constructed response approach.

The ONPAR items below are examples that are more constrained than the traditional constructed response items. The first ONPAR example, Rolling Balls, is also more constrained than the second ONPAR example, Power Plant.

## *ONPAR Example: Rolling Balls*

**Target**: Use data to predict how the different magnitudes of force affect the position of objects with various masses.

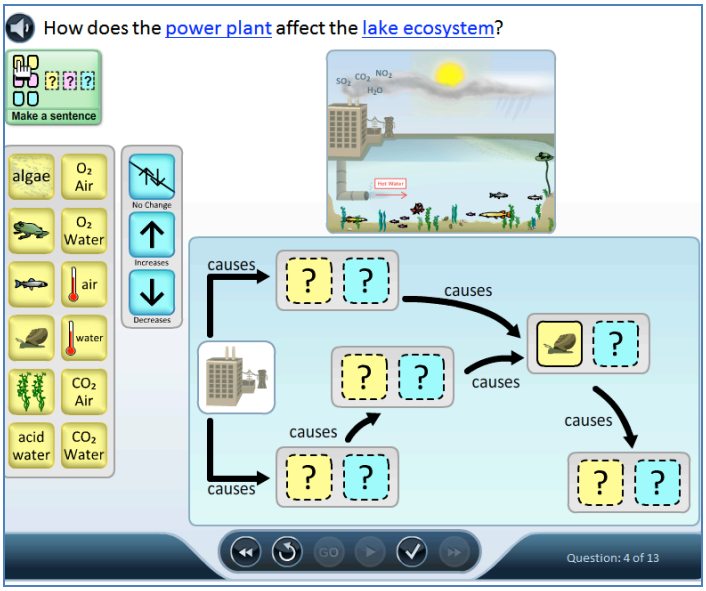



This scene begins with an animation of the red ball on the left being blown by the fan with a speed of 1 to the middle of the platform. This animated scene has the fan on speed 1 turning slower than the fan on speed 2. Students must drag the highlighted balls in a straight line on the platform to indicate how the fan speed and mass of the ball determine its final resting position.

In this item students must consider three critical factors: the relative speed of the fans, the two different masses of the balls, and the relative positions of four balls. While success on this item shows good evidence for the understanding of how force and mass are connected and the item itself allows a large number of possible answers, the item is still more constrained than its traditional item counterpart. The traditional item is a constructed response which asks two questions. The first asks students to compare how a fan at one speed affects the position of balls with two different masses, and then asks students to explain their reasoning. The second question asks students how a fan at different speeds affects a ball of the same mass, and then to explain their reasoning. While the ONPAR item asks for more fine-tuned comparisons, the traditional item is more open-ended.

## *ONPAR Example: Power Plant*

**Target**: Predict outcomes when abiotic and biotic factors are changed in an ecosystem.




This item shows a series of animations prior to the question and response screen shown above. In this scene students are asked to form causal chains by dragging blue and yellow icons on the left into the response spaces indicated with dotted boxes and question marks. Students must match the dragger to the correct color, and draggers can be used multiple times.

Although the traditional item was an open-ended constructed response asking how a power plant might affect the ecosystem, the above item has 5,038,848 ($12^4*3^5$) possible choices and hundreds of correct answers.

**Issue 5:** What are the tradeoffs within and across different item approaches in how communications devices are used for two purposes: the test asking the target question of the test taker, and the test taker providing a response back to the test? This is a concern about the centrality of different communication devices used in different items or in different aspects of items. When is this an issue and when is it not?

Sometimes it is important to use language for conveying the high precision required of an involved and elaborated response. Pictures are useful, but can't often match the precision of words in such situations. But the question is whether the maximum precision attainable is always most desirable. In the implicit communication taking place in a test item, it is important to be aware of the functional purposes that underlie the different means of representing content, but also the overall cognitive processing load that these means impose in combination and relation to one another. It seems, in other words, that the degree of precision required needs to be assessed in regard to the abilities of the targeted audience to process it successfully,

hence the importance of considering the two following points.

i) The function of language, static or dynamic visuals, and interactive components in establishing context, asking questions and in structuring test takers' response mechanisms. Example (if needed): Mystery substances. Original item has lots of confusing text and an abstract presentation. ONPAR uses images and animations.

ii) The function of language, static or dynamic visuals, and interactive components in allowing students to demonstrate what they know as compared to responding to a multiple choice item or explaining on a constructed response.

It is important to be aware of how different communication devices are useful in different places in a problem. For example doing real labs, interacting with computer simulations, or other item types open up different ways to gather information. Students are being asked to respond in different ways. Our communication with them needs to be thought through. We are very linguistically oriented, but there is also a vital role for other communication devices.

**Issue 6:** This is not a new issue……

a) What does a constructed response item do that is similar or different from a multiple choice item? CR and MC being heavy in their language requirement can make it difficult for students to show some kinds of conceptual knowledge. MC may be useful in assessing discrimination skills, but there are many other kinds of skills that can and should be assessed.

b) Performance assessments vs. other kinds of items
Performance assessments can be authentic ways to demonstrate knowledge, but can be expensive and time consuming to set up.

c) Portfolios vs. other assessment means
Portfolios can be a good way to get lots of information from students but comparability can be challenging.

d) Oral exams vs. other assessment means
Some need to sit, think, write and rewrite in order to get their ideas across.

e) Computer simulations which model real data vs. other assessment means
Computer simulations can be a good way to have students adjust or manipulate variables that it would be impractical for them to do in a hands-on lab setting. In addition they very closely emulate what real scientists do in the field and in their analyses as they model real-life data (flood systems, planetary exploration, tectonic plate movement etc.).