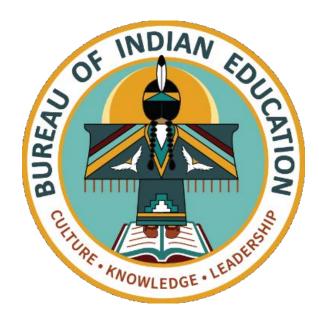


Test Specifications

Bureau of Indian Education (BIE) Science Assessment



Purpose of the BIE Science Assessment

The BIE Science Assessment is the Bureau of Indian Education's summative assessment for Science, administered at the end of grades 5, 8, and 11. The BIE Science Assessment measures students' proficiency with the NGSS Science Standards. As the BIE Science Assessment is a single measure at the end of a grade band, interpretations and uses of BIE Science Assessment scores should be supplemented with additional measures, including information from classroom summative and formative assessments in science.

Claims/Score Interpretation and Use Statements

The BIE Science Assessment is designed to measure whether students are on track to be ready for college or career by showing they have mastered the NGSS Science Standards. The standards require integration of Science and Engineering Practices, Disciplinary Core Ideas, and Crosscutting Concepts to explain phenomena and solve problems.

In addition to overall scale score, student performance on three science subdomains is reported:

- Practices and Crosscutting Concepts in Physical Sciences
- Practices and Crosscutting Concepts in Life Sciences
- Practices and Crosscutting Concepts in Earth and Space Sciences



Test Specifications – Test Design

Assessable Standards

The BIE Science Assessment assesses the NGSS Science Standards as follows:

- Grade 5 test: All standards in grades 3, 4, and 5.
- Grade 8 test: All standards in the middle school grade band (6-8).
- Grade 11 test: All standards in the high school grade band (9-12).

Test Design

The BIE Science Assessment is administered in three sessions. The test is administered online as a computer-based test (CBT).

Online accommodations are available for the CBT. Paper, large-print, and Braille test forms are also provided.

No calculator is provided for the BIE Science Assessment, as no items require calculator use. A periodic table will be provided as a reference for high school (Grade 11).

The types of items on the BIE Science Assessment are item clusters (CL), 2-point machine-scored standalone items (MS-2), and 4-point open-ended standalone items (OE). Additional item type descriptions and sample items can be found in the item specifications section of this document.

Both core operational items (which count for a student's score) and a small number of matrix field test items (which are try-out items that do not count for a student's score) are included on the BIE Science Assessment.

The total numbers of test items, points, and estimated testing time for the BIE Science Assessment are shown in the following tables.

Student Testing Experience								
Grade 5	Cluste	r/Passage	Items	Standalo	ne Items	Total	Total	
Grade 5	Stim/Psg	MS-1	MS-2	MS-2	OE	Items	Points	
Core Operational Items	6	12	12	8	3	35	64	
Matrix Field Test Items	2	4	4	4	1	13	24	
Total Student Experience 8 16 16 12						48	88	
	Estimated Testing Time (min)					150		

Student Testing Experience							
Grade 8	Cluste	r/Passage	Items	Standalo	ne Items	Total	Total
Graue o	Stim/Psg	MS-1	MS-2	MS-2	OE	Items	Points
Core Operational Items	6	12	12	8	3	35	64
Matrix Field Test Items	2	4	4	4	1	13	24
Total Student Experience 8 16 16 12 4 48						88	
Estimated Testing Time (min) 1					150		

Student Testing Experience							
Grade 11	Cluste	r/Passage	Items	Standalo	ne Items	Total	Total
Graue 11	Stim/Psg	Stim/Psg MS-1 MS-2 MS-2 OE		Items	Points		
Core Operational Items	6	12	12	10	3	37	68
Matrix Field Test Items	2	4	4	5	1	14	26
Total Student Experience 8 16 16 15 4 51					94		
Estimated Testing Time (min)					165		

Practice Test

A full-length practice test mirroring the operational test design is available for each grade. The practice tests and supporting materials can be accessed at https://bie.onlinehelp.cognia.org/.

Test Specifications – Reporting Categories

The reporting categories for BIE Science Assessment are based on the three content domains. Percentages for the distribution of operational (core) test points for each of the reporting categories reflect the distribution in the standards, so as not to over- or underrepresent content.

Based on this representativeness, the fourth content domain of Engineering, Technology, and Applications of Science is not reported as a subscore (as there are very few standards out of the total in each grade band). Items coded to these standards do count toward total test score.

Reporting Categories, Grade 5 BIE Science Assessment							
Reporting Category	Typical Number of Clusters	Typical Number of Standalone MS-2	Typical Number of Standalone OE	Number of Core Points	Percent of Core Points (+/- 4%)		
Practices and Crosscutting Concepts in Physical Sciences	2	4-6	1	24-28	40%		
Practices and Crosscutting Concepts in Life Sciences	2	1-3	1	18-22	30%		
Practices and Crosscutting Concepts in Earth and Space Sciences	2	1-3	1	18-22	30%		

Reporting Categories, Grade 8 BIE Science Assessment							
Reporting Category	Typical Number of Clusters	Typical Number of Standalone MS-2	Number	Number of Core Points	Percent of Core Points (+/- 4%)		
Practices and Crosscutting Concepts in Physical Sciences	2	2-4	1	20-24	35%		
Practices and Crosscutting Concepts in Life Sciences	2	2-4	1	20-24	35%		
Practices and Crosscutting Concepts in Earth and Space Sciences	2	1-3	1	18-22	30%		

Reporting Categories, Grade 11 BIE Science Assessment							
Reporting Category	Typical Number of Clusters	Typical Number of Standalone MS-2	Number	Number of Core Points	Percent of Core Points (+/- 4%)		
Practices and Crosscutting Concepts in Physical Sciences	2	3-5	1	22-26	35%		
Practices and Crosscutting Concepts in Life Sciences	2	3-5	1	22-26	35%		
Practices and Crosscutting Concepts in Earth and Space Sciences	2	1-3	1	18-22	30%		

Test Specifications – Cognitive Complexity

Because the items on the BIE Science Assessment are NGSS-aligned, the cognitive complexity of items is evaluated with a different framework than Depth of Knowledge.

For the items on the BIE Science Assessment, four indicators are used to classify the cognitive complexity of each item: stimulus, science and engineering practice, disciplinary core idea, and crosscutting concept. For each indicator, the classification in terms of high, medium, or low complexity is based on how the students are using the indicator to respond to the item – specifically, to what degree does students' engagement with the indicator contribute to the level of sense-making required by the item.

On the BIE Science Assessment, after summing the operational (core) test points at each cognitive complexity level across all four indicators, at least 10% of the points should be high cognitive complexity and no more than 35% of the points should be low cognitive complexity.

The descriptors for each indicator at the three complexity levels (high, medium, low) are presented in the following tables.

	STIMULUS					
	- Phenomenon is novel, complex, and/or unfamiliar to students					
	- Students must synthesize multiple pieces of information and do a significant					
	amount of "figuring out" to make sense of the phenomenon					
	- Phenomenon is somewhat novel, but may be analogous to what many					
Madium	students are familiar with					
Medium	students are familiar with - Students must use multiple pieces of information and do an intermediate					
	amount of "figuring out" to make sense of the phenomenon					
	- Phenomenon is familiar and/or more straightforward for students					
Low	- Students only need to use simple/straightforward information, and/or a					
LOW	single piece of information, and do a minimal amount of "figuring out" to					
	answer the question or contribute to making sense of the phenomenon					

	SEP (SCIENCE AND ENGINEERING PRACTICE)
	- Students must apply the SEP, or multiple SEPs, in a sophisticated way to
	make sense of the phenomenon (e.g., synthesis to perform more connections,
High	steps, combination of SEP elements, such as having to combine data, produce
	a new graph or model as evidence, etc.)
	- Often little to no scaffolding that helps students apply the SEP
Modium	Students must apply the SEP to make sense of the phenomenonTypically some scaffolding that helps students apply the SEP
Mediuiii	- Typically some scaffolding that helps students apply the SEP
	- Students only need to use the SEP in a simple, mechanical way to answer the
Low	question or contribute to making sense of the phenomenon
	Often a large amount of scaffolding that helps students apply the SEP

© 2021 Cognia, Inc.

6



	DCI (DISCIPLINARY CORE IDEA)
	- Students must apply and connect DCIs in a sophisticated way to make sense
	of the phenomenon, i.e.,
	 application of science ideas (often multiple, grade-band appropriate
High	ideas) in unique ways or new combinations
	 knowledge transfer to construct new understanding, make sense of
	novel phenomena
	- Often little to no scaffolding that helps students apply the DCI
	- Students must apply or reason with the DCI(s) to make sense of the
Medium	phenomenon
	- Typically some scaffolding that helps students apply the DCI
	- Students use the DCI in a simple, straightforward way (i.e., little to no
	application or reasoning) to answer the question or contribute to making sense
	of the phenomenon
	- Often a large amount of scaffolding that helps students apply the DCI

	CCC (CROSSCUTTING CONCEPT)
High	- Students must apply the CCC in an in-depth way to expand thinking and
	make non-typical connections to make sense of the phenomenon
Modium	- Students must use the CCC as specified by the CCC sub-bullet detail to make sense of the phenomenon
Medium	sense of the phenomenon
T	- Students only use the CCC in a general way to answer the question or
Low	contribute to making sense of the phenomenon



Test Specifications – Fairness

Fairness is defined as the extent to which the test scores are valid for different groups of test takers. Consideration of universal design, bias, and sensitivity guidelines support the construction of fair, valid assessments.

Universal Design for Assessments

The concept of Universal Design for Assessments focuses on developing content and assessments that reach the widest population of students possible. Stimuli and items on the BIE Science Assessment are designed to simply and clearly present tasks and to provide maximum readability, comprehensibility, and legibility. The seven elements of Universal Design for Assessments are based on the original UDL guiding principles:

Universal Design for Assessments

Principle	Explanation
Inclusive Assessment Population	Tests designed for state, district, or school accountability must include every student except those in the alternate assessment, and this is reflected in assessment design and field-testing procedures.
Precisely Defined Constructs	The specific constructs tested must be clearly defined so that all construct-irrelevant cognitive, sensory, emotional, and physical barriers are removed.
Accessible, Non-Biased Items	Accessibility is built into items from the beginning, and bias review procedures ensure that quality is retained in all items.
Amenable to Accommodations	Test design facilitates the use of needed accommodations (e.g., all items can be brailled).
Simple, Clear, and Intuitive Instructions and Procedures	All instructions and procedures are simple, clear, and
Maximum Readability and Comprehensibility	A variety of readability and plain language guidelines are followed (e.g., sentence length and number of difficult words kept to a minimum) for readable and comprehensible text.
Maximum Legibility	Characteristics that ensure easy decipherability are applied to text, tables, figures, and illustrations, and to response formats.

Bias

The concept of Bias is defined as the presence of some characteristic of an item that results in differential performance for two individuals of the same ability but from different ethnic, sex, cultural, or religious groups.

Bias can occur whenever content offends or disadvantages a student or group of students due to gender, race, regional background, socioeconomic status, or any other such classification.



Test developers take care to craft content in a way that does not misrepresent specific groups or rest on assumptions made about specific groups, that in turn could negatively impact how students interpret content.

- Stimulus and item content on the BIE Science Assessment should not present stereotypes or unfair representations of gender, race, ethnicity, disability, culture, or religion.
- Stimulus and item content on the BIE Science Assessment should not depend on overlyexperiential information such as knowledge of technology, consumer goods, pop culture, geographic locations, or sports and extracurricular activities. While these topics are not completely excluded from use, care must be taken to ensure that the items are presented in a way that does not require a level of knowledge that would not be held by all students.

Sensitivity

Sensitivity refers to the presence of content that is contrary to the acceptable norms of the students, educators, parents, or other members of the community that may interact with the assessment. Sensitive subject matter can impact student performance or attitudes toward testing, and hence, their test scores.

Consideration of bias and sensitivity issues is very important when developing content for an assessment. Test developers must ensure that stimuli and items are free of content that will negatively affect a student's performance not because of what the student knows and can do but because the content evokes an emotional response from that student (or is in some other way distracting to the student).

Subjects/contexts that are likely to prompt emotional distress on the part of students cannot be used on the BIE Science Assessment (e.g., war, violence, human death or debilitating disease, animal-based medical research). Careful judgment should be applied to PEs that cover topics that may be considered controversial by some groups (e.g., evolution examples, population dynamics including death/extinction, environmental impact). Those PEs represent content knowledge to be assessed, but the assessment must be done in a sensitive, unbiased way.

Stimulus Specifications

All items for the BIE Science Assessment have a stimulus. For clusters, all items in the set are associated with a common stimulus that presents a science phenomenon or engineering design problem. For standalone items (MS-2, OE), the item includes a lead stimulus that provides a specific science phenomenon or engineering design problem, or context thereof. By phenomenon, we mean something observable that happens in the real world, whether natural or man-made. By engineering design problem, we mean a personal or societal need or want.

Specifications for Cluster Stimuli

- 1. The stimulus must present a single, rich science phenomenon or engineering design problem aligned to the PEs.
- 2. The stimulus may present any variety of elements to provide the necessary information to support sense-making (via the items) around the phenomenon or problem: text paragraphs, passages, graphs, data tables, models, drawings, etc.
- 3. The stimulus must be rich enough to support the development of enough items for the cluster, in the context of a storyline (sequence of sense-making) around the phenomenon or problem using the DCIs, SEPs, and CCCs of the targeted PEs.
- 4. All information in the stimulus should be necessary, but not conceptually sufficient, for students to respond (i.e., students must also use their own knowledge of the constructs in the PE(s) to answer the items, rather than simply identify given information).
- 5. The stimulus phenomenon or problem must be grade-appropriate, engaging, and relevant for students at that grade level.
- 6. The stimulus should adhere to the specifications in the following table regarding length, wording, and complexity.

Stimulus characteristic	Elementary School (Grades 3-5)	Middle School (Grades 6-8)	High School (Grades 9-12)				
Text word count**	100-300 words	100-400 words	100-400 words				
	be lower; in a stimulus	with few and/or very simp	and/or complex graphics, ole graphics, the word				
Vocabulary level (excluding science Grade 3 Grade 5 maximum Grade 8 maximum content vocabulary)							
Readability/Lexile maximum	820L (Gr 3)	1010L (Gr 5)	1185L (Gr 8)				

Slight mix of simple and Mix of simple and more more complex phrasing complex phrasing and

sentence structure,

complex graphics,

demands.

average vocabulary

average to moderately

10

and sentence structure,

average to moderately

complex graphics,

demands.

average vocabulary

© 2021 Cognia, Inc.

Simple sentence

graphics, lower

clear/uncomplicated

vocabulary demands,

use of only essential

science vocabulary.

structures,



Qualitative text

characteristics

Specifications for Standalone Item Stimuli

- 1. MS-2 items: The stimulus must present a hook or driving reason for the question being asked, and it must set a phenomenon- or problem-based context, aligned to the PE, for the item. The stimulus will typically not be as extensive as a stimulus for an item cluster.
- 2. OE items: The stimulus must present a hook or driving reason for the question being asked, and it must include a phenomena or problem, aligned to the PE, to drive the item. The stimulus for open-ended items will typically be more concise than for item clusters but more detailed than for MS-2 standalone items.

Item Specifications

Alignment

The items on the BIE Science Assessment are aligned to the NGSS Science Standards.

Each item is aligned to a performance expectation (PE) as well as dimensions of the performance expectation. All items must have either 2-dimensional or 3-dimensional alignment.

Item Types

The types of items on the BIE Science Assessment are item clusters, 2-point machine-scored standalone items (MS-2), and 4-point open-ended standalone items (OE):

- An item cluster is a set of items all associated with a common stimulus. Clusters contain
 four items. These items may be multiple choice, multiple select, or technologyenhanced, with two of the items being worth 1 point and two of the items being worth 2
 points. The clusters typically align to two PEs, and all clusters measure all three
 dimensions of the PEs being assessed.
- Standalone MS-2 items are worth 2 points. These items have two parts (Part a and Part b) for students to answer, and 0, 1, or 2 points total can be earned across Part a and Part b. These items may be multiple choice, multiple select, or technology-enhanced (e.g., drag-and-drop, hot spot, drop-down selections).
- Open-ended items are worth 4 points. These items require students to write an extended response to a prompt. The prompt may be a single prompt, or more typically, the items are written with multiple, scaffolded parts for students to respond to. These items are hand-scored, with scorers using a rubric and scoring notes to evaluate responses on a scale from 0–4.

Samples of each of these item types are included on the following pages.



Clusters: Clusters are a set of 4 items all associated with an introductory passage, or "stimulus."

- The stimulus typically contains both text and graphics such as diagrams, tables, or graphs. An example stimulus from the *grade 5 practice test* is on the next page. The items associated with the cluster assess two Physical Sciences PEs:
 - 5-PS1-3: Make observations and measurements to identify materials based on their properties.

SEP: Planning and Carrying Out Investigations

DCI: PS1.A: Structure and Properties of Matter

CCC: Scale, Proportion, and Quantity

5-PS1-4: Conduct an investigation to determine whether the mixing of two or more substances results in new substances.

SEP: Planning and Carrying Out Investigations

DCI: PS1.B: Chemical Reactions

CCC: Cause and Effect

Read the information. Then answer the questions that follow.

Investigating Gas Production

In class, a teacher demonstrates a chemical reaction by mixing vinegar and baking soda to produce bubbles of gas. Eliana wonders whether mixing other substances could also produce a gas. She decides to test different combinations of sugar, water, vinegar, and baking soda.

Some properties of these substances are shown in the table.

Properties of Substances

Substance	Color	Solid or Liquid	Attracted to a Magnet	Conducts Electricity
Sugar	White	Solid	No	No
Water	Clear	Liquid	No	Yes
Vinegar	Clear	Liquid	No	Yes
Baking soda	White	Solid	No	No

Investigation 1

Eliana mixes a small amount of each liquid and solid in a bowl and observes whether bubbles of gas are produced. Her observations are shown in the table.

Investigation 1 Observations

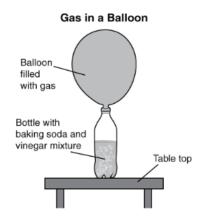
Liquid Used	Solid Used	Gas Produced	
Water	Sugar	No	
Water	Baking soda	No	
Vinegar	Sugar	No	
Vinegar	Baking soda	Yes	

Investigation 2

Next, Eliana wonders whether changing the amount of baking soda would change the amount of gas produced. To investigate, she follows these steps:

- 1. Record the mass of a balloon.
- 2. Pour 50 milliliters of vinegar into a bottle.
- Put 5 milliliters of baking soda inside the balloon. Hold the balloon so that the baking soda stays inside the balloon and attach the open end of the balloon to the top of the bottle.
- Lift the balloon so that the baking soda falls into the bottle with vinegar.
- Wait one minute.
- Carefully remove the balloon from the bottle without allowing any gas to escape.
- 7. Measure the mass of the balloon filled with gas.
- Calculate the mass of gas produced by subtracting the mass of the balloon from the mass of the balloon filled with gas.
- 9. Repeat steps 1-8 until three trials have been completed.
- Repeat steps 1–9 with 10 milliliters and 15 milliliters of baking soda

The results of one trial are shown in the diagram.



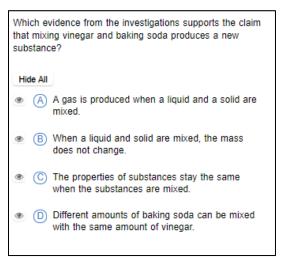
Eliana's data are shown in the table

Investigation 2 Data

Amount of Baking Soda	M	Average Mass of Gas Produced		
(milliliters)	Trial 1	Trial 2	Trial 3	(grams)
5	1.0	0.8	1.2	1.0
10	1.5	1.9	1.4	1.6
15	2.4	1.9	2.6	2.3

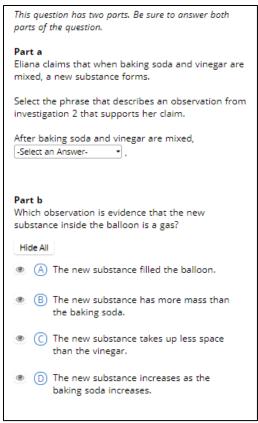


• Two of the items in the cluster are machine-scored items worth 1 point each. These items may be multiple-choice, multi-select, or technology-enhanced items (e.g., drag-and-drop, hot spot, drop-down selections).



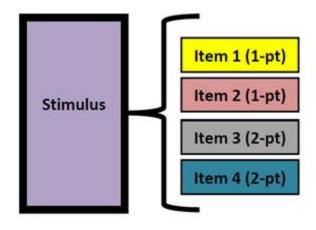
MS-1 cluster item, grade 5 practice test, aligned to PE 5-PS1-4: Conduct an investigation to determine whether the mixing of two or more substances results in new substances. The dimensions for the PE are SEP: Planning and Carrying Out Investigations; DCI: PS1.B: Chemical Reactions; CCC: Cause and Effect. This particular MS-1 item in the cluster assesses the DCI and CCC dimensions.

• The other two items in the cluster are machine-scored items worth 2 points each. These items have two parts, with Part a worth 1 point and Part b also worth 1 point. Each part of the item may be presented as multiple-choice, multi-select, or technology-enhanced (e.g., drag-and-drop, hot spot, drop-down selections).



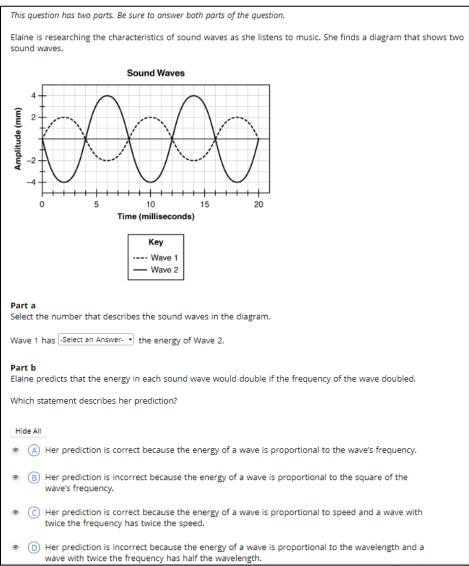
MS-2 cluster item, grade 5 practice test, aligned to PE 5-PS1-4: Conduct an investigation to determine whether the mixing of two or more substances results in new substances. The dimensions for the PE are SEP: Planning and Carrying Out Investigations; DCI: PS1.B: Chemical Reactions; CCC: Cause and Effect. This particular MS-2 item in the cluster assesses the DCI and CCC dimensions.

• The entire cluster is worth a total of 6 points. The diagram below summarizes the structure of a cluster.



MS-2 Items: MS-2 items are standalone, or individual, machine-scored items.

• As in the cluster, the standalone MS-2 items are worth 2 points and have two parts, with Part a worth 1 point and Part b also worth 1 point. Each part of the item may be presented as multiple-choice, multiselect, or technology-enhanced (e.g., drag-and-drop, hot spot, drop-down selections).



MS-2 item, grade 8 practice test, aligned to PE MS-PS4-1: Use mathematical representations to describe a simple model for waves that includes how the amplitude of a wave is related to the energy in a wave. The dimensions for the PE are SEP: Using Mathematics and Computational Thinking; DCI: PS4.A: Wave Properties; CCC: Patterns. This particular MS-2 standalone item assesses the SEP, DCI, and CCC dimensions.

OE Items: OE, or open-ended, items are standalone items that require students to provide a written response to a prompt or question.

- The prompt or question may be a single prompt, or more typically, the item will be written with multiple, scaffolded parts for students to answer.
- The items are worth 4 points each and are hand-scored for 4, 3, 2, 1, or 0 points by trained scorers using a rubric and scoring notes.

This question has two parts. Be sure to answer both parts of the question.

Some students work at a local aquarium. One of their tasks is to care for mollusks and corals in ocean water in a tank at the aquarium. The students need to make sure that the ocean water has the right balance of calcium ions (Ca²⁺) and carbonate ions (CO₃²⁻) that the mollusks and corals need to build their shells and skeletons.

To do this, the students need to ensure that calcium and carbonate ions are continuously added to the ocean water in the tank. The students know that ocean water contains calcium carbonate, which naturally breaks down into calcium and carbonate ions. The equilibrium relationship between the components in the water is shown in the equation.

Equilibrium Equation

$$CO_2 + H_2O + CaCO_3 \rightleftharpoons Ca^{2+} + 2 H^+ + 2 CO_3^{-2}$$

The students decide to test the equilibrium relationships in the equation. With ocean water as an input, the students remove calcium ions (Ca^{2+}) as the ions form in the water in the tank. The students observe that as they remove calcium ions, more calcium ions form in the tank. They realize that this is an example of Le Chatelier's principle that describes the equilibrium relationships in the water.

The people who work at the aquarium tell the students that ocean water contains carbon dioxide (CO₂) and that increasing amounts of CO₂ in ocean water can cause some of the calcium carbonate (CaCO₃) in the shells and skeletons of ocean organisms to dissolve

The students want to solve this problem by decreasing the amount of carbon dioxide in ocean water.

- a. Describe one way students could decrease the amount of CO2 in ocean water by applying Le Chatelier's principle.
- b. Describe one constraint on implementing the change described in Part (a).

OE item, grade 11 practice test, aligned to PE HS-PS1-6: Refine the design of a chemical system by specifying a change in conditions that would produce increased amounts of products at equilibrium. The dimensions for the PE are SEP: Constructing Explanations and Designing Solutions; DCI: PS1.B: Chemical Reactions and ETS1.C: Optimizing the Design Solution; CCC: Stability and Change. This particular OE item assesses the SEP, DCI, and CCC dimensions.

