Critique of an Assessment

Diana Theisinger, M. Ed.

College of William & Mary

Critique of an Assessment

In this critique, I have provided an overview of the intended learning outcomes (ILOs) associated with a particular unit assessment. Then, I have created a Table of Specifications (see Figure 1) aligning each ILO with its corresponding cognitive level and distributing test items throughout the Table of Specifications according to the ILO and cognitive level with which they align. Finally, I have offered a critique of the assessment in terms of sampling and construct validity and outlined an instructional leadership plan for working with the teacher who created this assessment to make improvements.

The Assessment: Science Unit Test

A middle-school science teacher with nine years of experience created this assessment (see Appendix). She teaches 6th-grade science and intended this as an end of unit assessment. The teacher used this assessment to check students' mastery and make course adjustments for the next unit. The assessment was created with the assistance of the Interactive Achievement testing software that this teacher's school division uses. There are 30 multiple-choice items (Waugh & Gronlund, 2013); 10 items require students to use a chart, table, or diagram to answer the question. The assessment is cleanly organized, with adequate white space and a clear flow from one question to the next. Each item is clearly numbered, but there are no directions on the assessment.

Intended Learning Outcomes & Curriculum Alignment

The teacher who provided this assessment also provided a test breakdown from Interactive Achievement that shows the Virginia Standard of Learning (SOL) associated with each item on the assessment. In this case, the SOL served as a proxy for the teacher's ILOs. I found several errors in the Interactive Achievement listing. Table 1 shows the ILOs

corresponding with the assessment according to Interactive Achievement. Table 2 shows the ILOs actually assessed according to my close reading of the assessment and related standards documents (VDOE, 2010a, 2010b).

Table 1

Assessed Content Strands According to Interactive Achievement's Test Breakdown

6.1b record precise and approximate measurements
6.1c use scale models to estimate distance, volume, and quantity
6.4a understand atoms consist of particles, including electrons, protons, and neutrons
6.5b understand properties of water in all three phases
6.8h understand the cause of tides
PS.1b measure length, mass, volume, density, temperature, weight, and force
PS.2b understand elements, compounds, mixtures, acids, bases, and salts
PS.2c understand solids, liquids, and gases
PS.3b understand the modern model of atomic structure
PS.4a understand symbols, atomic number, atomic mass, chemical families (groups), and periods
PS.4b understand classification of elements as metals, metalloids, and nonmetals
PS.5a understand physical changes

Table 2

Assessed Content Strands According to Standards of Learning and Item Analysis

6.1b record precise and approximate measurements
6.1c use scale models to estimate distance, volume, and quantity
6.4a understand atoms consist of particles, including electrons, protons, and neutrons
6.5b understand properties of water in all three phases
~~6.8h understand the cause of tides~~
PS.1b measure length, mass, volume, density, temperature, weight, and force
PS.2b understand elements, compounds, mixtures, acids, bases, and salts
PS.2c understand solids, liquids, and gases
PS.3b understand the modern model of atomic structure
PS.4a understand symbols, atomic number, atomic mass, chemical families (groups), and periods
PS.4b understand classification of elements as metals, metalloids, and nonmetals
PS.5a understand physical changes
Standards Not Included on Test Breakdown
PS.2.d understand physical properties
6.1e understand a method is devised to test the validity of predictions and inferences

In Table 2, I have added a strikethrough to the ILO, "understand the causes of tides," which appears to have been misapplied in the Interactive Achievement database. The item that is attributed to that ILO by the testing software is actually a question about density. The test breakdown also listed eight items as "uncategorized," meaning they could not be linked with any of the SOL in the database. However, I was easily able to find corresponding SOL for all eight items by scanning the standards documents (VDOE, 2010a, 2010b); further, the corresponding SOL in all cases was either already included in the standards being assessed elsewhere on the assessment or was closely related (i.e., from the same content strand) to ones being assessed.

By examining the accurate ILOs (those listed in Table 2), I was able to determine a close alignment between the curriculum and the assessment. In general the assessment items seem to represent a cogent set of ILOs. However, I think it would be difficult for any assessment to

measure students' knowledge adequately on 13 strands of content using only 30 items. One area for possibly improving this assessment would be narrowing its focus to include fewer content strands. Perhaps a separate, performance assessment (Waugh & Gronlund, 2013) could be given for those items concerned with accurate measuring, which would reduce the field from 13 content strands to only 10. This would allow three assessment items per content strand and have the added benefit of providing students an alternate means of engaging with the content and demonstrating their knowledge through a separate performance task.

Construct & Sampling Validity

Figure 1 shows the Table of Specifications I created for this assessment.

Content	Bloom's Taxonomy Level					
	Remember	Understand	Apply	Analyze	Evaluate	Create
6.1b record precise and approximate measurements	2		✔			
6.1c use scale models to estimate distance, volume, and quantity		✔	26			
6.4a understand atoms consist of particles, including electrons, protons, and neutrons	✔ 3, 4					
6.5b understand properties of water in all three phases	16	✔ (22)	(7)			
6.8h understand the cause of tides	~~11~~	✔				
PS.1b measure length, mass, volume, density, temperature, weight, and force	(11), 14		✔ (1), 23, 24			
PS.2b understand elements, compounds, mixtures, acids, bases, and salts	5	✔ 6, 15, 18				
PS.2c understand solids, liquids, and gases	19, 29	✔ 28				
PS.3b understand the modern model of atomic structure	✔ 25					
PS.4a understand symbols, atomic number, atomic mass, chemical families (groups), and periods		✔ (9)	(12), 17, 30			
PS.4b understand classification of elements as metals, metalloids, and nonmetals		✔	13			
PS.5a understand physical changes	21	✔ 20				
Standards Not Included on Test Breakdown						
PS.2.d understand physical properties	(8)	✔ (10)				
6.1e understand a method is devised to test the validity of predictions and inferences		✔ (27)				

Figure 1. Table of Specifications showing the relationship between content knowledge and skills used to develop test questions and the corresponding cognitive level of each strand according to Bloom's Taxonomy (Bloom et al., 1956). Item numbers from the assessment are placed in the appropriate box, with parenthesis indicating the items were originally misclassified in Interactive Achievement's test breakdown and have been relocated to the correct content strand. The standard highlighted in yellow was misapplied to item 11 and was not actually assessed by any items on the test.

I have included all of the ILOs listed in the test breakdown from Interactive Achievement, with modifications to indicate my analysis. I have highlighted the standard attributed to item 11, Science 6.8h, and added a strikethrough to that item number. Item 11 was incorrectly attributed to this standard, which refers to the movement of tides; it should, instead, be attributed to the Physical Science 1.b standard, which refers to density. Throughout the Table of Specifications, I have placed item numbers inside parenthesis to indicate that they were originally misclassified on the test breakdown, either attributed to the wrong content strand (e.g., item 11) or listed as "uncategorized" (e.g., items 1, 7, 8, 9, 10, 12, 22, and 27), when the item did, in fact, align with a related content standard. I have moved these items to the cell associated with the correct content standard.

Sampling validity. This assessment does not have strong sampling validity. Using a Table of Specifications adjusted to the actual content of the assessment, of the 13 content strands on the assessment, four are associated using only one item and three are assessed using only two items. This means students have only one or two chances to demonstrate mastery of those content strands.

Additionally, only half of the items on the test are written at the cognitive level suggested by the ILO. Nine are written below the suggested cognitive level; six are written above. This means that if a teacher relied only on this assessment to determine students' learning for this unit, his or her view of student learning would be skewed. If the teacher relied on the test breakdown report produced by Interactive Achievement, the outlook would be even grimmer. Applying Interactive Achievement's analysis of assessment items, only 21 items are related to any of the ILOs for this unit; of those 21 items, only 10 are worded to assess knowledge at the cognitive level suggested by the standard.

This assessment also does not give an accurate picture of students' knowledge at cognitive levels aligned with the ILOs. For example, item 2 is the only item dealing with 6.1b, which states that students will "record precise and approximate measurements," applying skills they have learned in the classroom. Yet, item 2 is a procedural question, dealing with students knowing where to stand for a level view when making measurements. No application of skills is required.

Construct validity. Because the assessment lacks strong sampling validity, it cannot have strong construct validity. In fact, only two strands, PS.1b and PS.2b contain an adequate number of assessment items that are written at the correct cognitive level. If the teacher were to use this assessment fairly, he or she would only be able to reliably draw conclusions about students' mastery of those two content strands. It is concerning that the teacher appears to have used the SOL as a proxy for ILOs, especially in light of the problems with the Interactive Achievement database. It is unclear whether the teacher intended, for example, to include "understand the cause of the tides" as an ILO; what seems more likely is that the teacher found item 11 in the database and intended it as a measure of students' knowledge about density, not realizing that it was inaccurately coded in the database.

Conclusion. This assessment has weak sampling and construct validity. Simply put, this assessment does not adequately measure the essential learning objectives for these content standards.

Repeated Trials Reliability

For seven of the content strands assessed by this instrument, students have only one or two opportunities to demonstrate mastery. This means that students' performance on this assessment has the potential to vary widely based on factors unrelated to their mastery of the

content. In other words, the repeated trials reliability of this assessment is low; a few lucky or unlucky guesses could dramatically impact students' performance.

Reliability & Sources of Measurement Error

There are only a few sources of error associated with individual items on the assessment. The charts, graphs, and diagrams used are clear and easy for students to read and understand. Overall, questions are concise and clearly written and the assessment items flow smoothly, with adequate white space.

One area that could be improved is the consistency of the periodic table diagram that is used. For three items, a miniature and abbreviated version of the periodic table is used, but for the fourth item that relies on this diagram, the full, detailed periodic table is used. This could present a source of confusion for some students. Additionally, the abbreviated version of the diagram increases the chance that students may be able to guess the right answer to questions such as item 9: "Which of the following elements has 16 protons?" In the abbreviated diagram, each element on the periodic table is shown with only one number in the cell; in this case, simple matching might lead students to the correct answer, Sulfur, since the number 16 only appears in the box with the symbol for Sulfur. Had the detailed periodic table been used, students would have had to filter through more data—like the atomic weight of each element—to arrive at the correct answer. This would provide a more accurate snapshot of student learning. The assessment would be clearer if the detailed periodic table was presented first and followed by the four questions related to interpreting the periodic table, omitting the smaller, abbreviated diagrams altogether.

Another item that could be improved is item 18: "Which of the following molecules does not contain three items?" This is the only negatively stated (Waugh & Gronlund, 2013) item on

the assessment, but no emphasis is provided to draw students' attention to the word "not." It would be easy for a student to overlook that modifier and choose one of three answer choices that does contain three atoms rather than the one that does not. Fortunately for students, item 18 is associated with one of the two content strands on the test that has an adequate sample of assessment items. Students have other opportunities to demonstrate their knowledge of this content on the assessment.

Scoring System

Each item on the assessment is weighted equally, at 3.3 points each. This is significant because if students miss only three items on the assessment, their letter grade will drop from an A-level to a B-level. And, since the items on the assessment are not distributed evenly across the content strands, a student who was absent on the day measurement was introduced, and therefore struggles with PS.1b, could miss as many as five items. This would automatically lower the student's score to a low B-level. Similarly, if a student struggled with reading the periodic table (PS.4a) or understanding elements and compounds (PS.2b), he or she could miss up to four items in each of those categories. Conversely, students who are struggling may evade detection with this assessment: many of the standards are assessed with only one or two items, so a lucky guess or an item that is written at a lower-than intended cognitive level could mask the need for re-teaching.

Instructional Leadership

This assessment is well laid-out and most of the questions are well written. I would commend the teacher for selecting test items from Interactive Achievement that cover the broad scope of the ILOs for this unit. The test would benefit from reorganizing similar items so they

appear near one another and from adding directions to differentiate between items that require students to use a chart or diagram to answer the question and those that do not.

The primary weakness of this assessment is in the alignment of items' cognitive level with the cognitive level of the ILOs. Likewise, the distribution of items across content strands (i.e., the sampling validity) does not give students adequate opportunities to demonstrate their mastery. Unfortunately, the source of this misalignment is most likely an overreliance on the Interactive Achievement software. In many cases, the use of this expensive software is endorsed, if not mandated, by the school division. This means my conversation with the teacher would focus on cautioning him or her against blind trust of the software. Instead, the teacher needs to look critically at each item in the software's database to insure it is accurately categorized and aligns with the cognitive level implied by the ILO.

Additionally, it will be important that this teacher not rely on the data collected from this assessment in planning changes to future instruction. There simply is not an adequate sample of students' mastery of the ILOs on the assessment for the teacher to make valid inferences about changing instruction.

References

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain.* New York, NY: David McKay Company.

Virginia Department of Education [VDOE]. (2010a). Science Standards of Learning curriculum framework 2010: Grade six. Retrieved from http://www.doe.virginia.gov/testing/sol/standards_docs/science/2010/k-6/stds_science6.pdf

Virginia Department of Education [VDOE]. (2010b). Science Standards of Learning curriculum framework 2010: Physical science. Retrieved from http://www.doe.virginia.gov/testing/sol/standards_docs/science/2010/courses/stds_physical_science.pdf

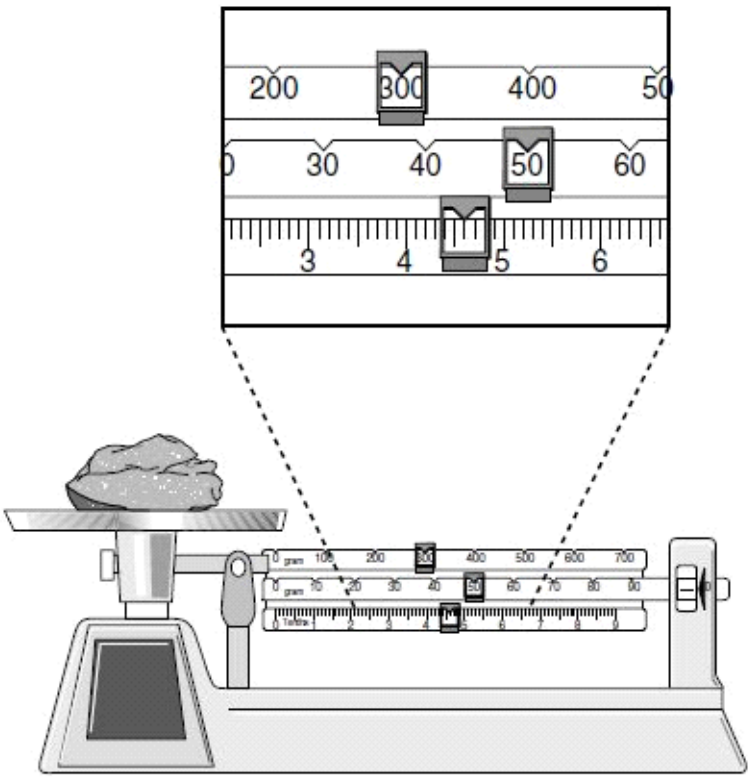
Waugh, C. K., & Gronlund, N. E. (2013). *Assessment of student achievement* (10th edition). Boston, MA: Pearson.

SCI-PS

Unit 1 Test

[Exam ID:46V9V2] Scan Number:33

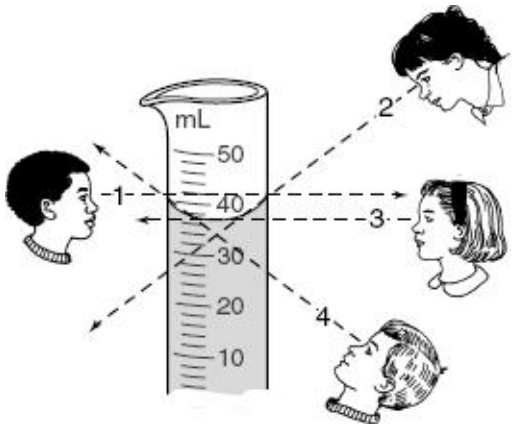
1



What is the mass of the rock?

A 350.6 g
B 335.6 g
C 354.6 g
D 356.0 g

2



The diagram shows a graduated cylinder containing water. From which position will the most accurate measure of the volume of the water be made?

F Position 2
G Position 1
H Position 3
J Position 4

3 **Which of these belongs in the outermost shell (energy level) of an atom?**

A Electrons
B Photons
C Protons
D Neutrons

4 **A proton has which of the following charges?**

F Negative
G Magnetic
H Positive
J Neutral

5 **A substance made up of two or more elements that have been chemically combined is called—**

A an atom
B an element
C a compound
D a mixture

6 **Which of these is best classified as a mixture?**

F Water
G Carbon dioxide
H Iron
J Soil

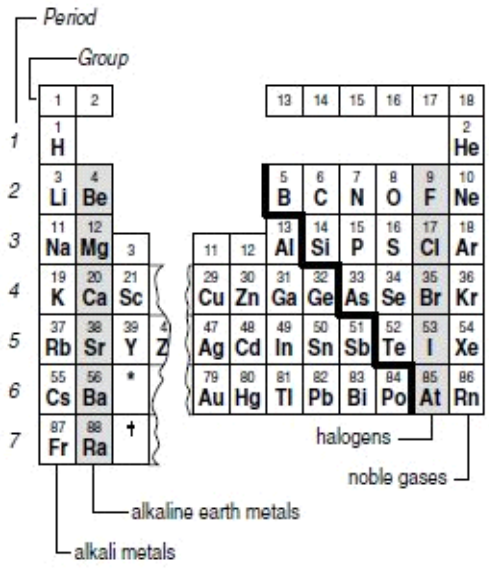
7 **Which of these will cause water to change to ice?**

A Stirring the water
B Putting the water in the sunlight
C Adding salt to the water
D Removing heat from the water

8 **A student determined the mass and volume of a gold ring. What physical property of the gold can be determined using these two measurements?**

F Its expansion rate
G Its density
H Its reactivity
J Its melting point

9



Which of the following elements has 16 protons?

A Oxygen
B Chlorine
C Sulfur
D Germanium

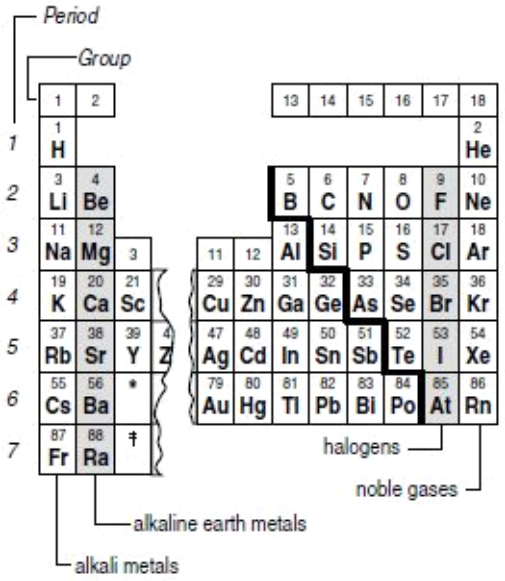
10 **Which of the following is a physical property of copper?**

F Ductile, can be drawn into a wire
G Liquid at room temperature
H Readily reacts with oxygen
J Readily reacts with water to form an acid

11 **The density of an object is dependent upon the object's mass and—**

A height
B weight
C volume
D area

12



Which of the following elements belongs to the family indicated?

F Helium belongs to the noble gas family.
G Nitrogen belongs to the halogen family.
H Calcium belongs to the alkali family.
J Sodium belongs to the alkaline earth family.

13

Characteristics of Samples of Matter

Sample	Luster	Malleability/ Ductility	Electrical Conductivity
1	None	No	Insulator
2	High	No	Semi-conductor
3	High	Yes	Conductor
4	Low	No	Insulator

According to the data in the table, which sample of matter is *most likely* a metal?

A 3
B 4
C 2
D 1

14 **Which of the following is used to measure the mass of an object?**

F Metric ruler
G Graduated cylinder
H Barometer
J Balance

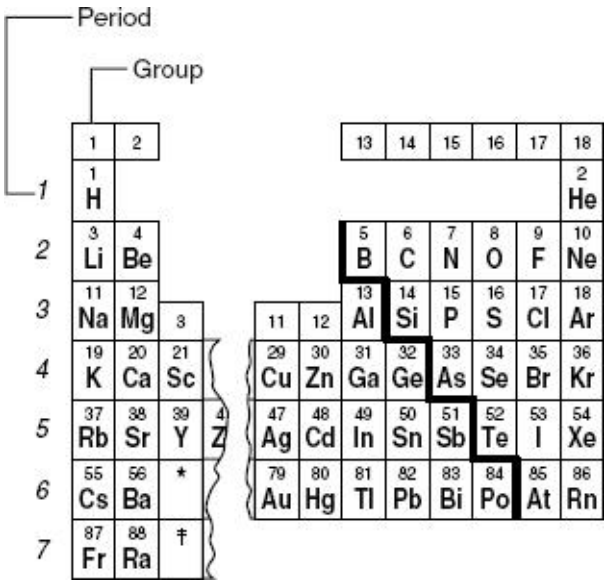
15 **Compounds are made up of more than one element. Which of the following represents a compound?**

A NH_3
B N_2
C H_2
D O_3

16 **Which process causes dew to form in the morning?**

F Precipitation
G Evaporation
H Transpiration
J Condensation

17



Which of the following lists the elements in order, from those having the *least* protons to those having the *most* protons in the atoms?

A Rb, K, Na, Li
B Na, S, Al, Cl
C O, N, B, Li
D O, S, Se, Te

18 **Which of the following molecules does not contain three atoms?**

F H_2O
G CO_2
H O_3
J N_2

19 **Of the states of matter, gas is the *only* state that —**

A is fluid
B has a definite shape
C has mass
D is highly compressible

20 **If you break a piece of glass, the shape of the glass changes, but the properties in the fragments remain the same. Which of the following has occurred?**

F A chemical change
G A temperature change
H A phase change
J A physical change

21 **Which process is a physical change?**

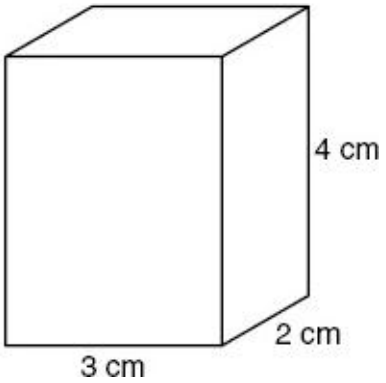
A Burning coal
B Rusting iron
C Tarnishing silver
D Melting ice

22 **During the process of evaporation, liquid water becomes —**

F separate atoms of water
G atoms of hydrogen and oxygen
H molecules of water vapor
J tiny drops of water

23 $\text{Density} = \frac{\text{mass}}{\text{volume}}$

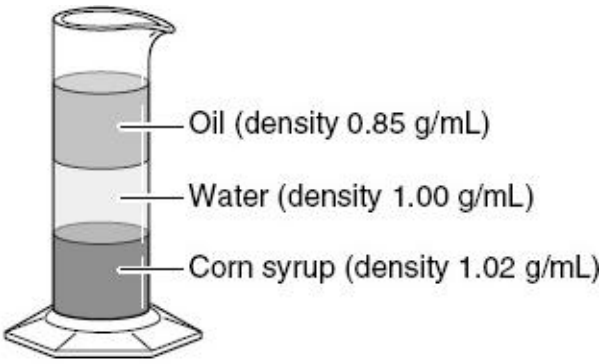
Mass = 12 g



The density of the box is-

A 1.0 g/cm^3
B 2.0 g/cm^3
C 1.5 g/cm^3
D 0.5 g/cm^3

24



Density Column

Densities of Some Unknowns	
Liquids	**Density (g/mL)**
Sample A	1.02
Sample B	0.96
Sample C	1.15
Sample D	0.82

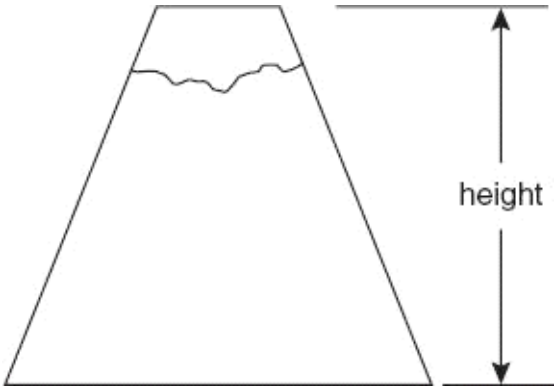
Students measured and recorded the density of 4 samples. Using the density column and the data table shown above, which of the samples will probably float on top of the oil?

F Sample D
G Sample A
H Sample B
J Sample C

25 **The atom is composed of electrons, protons, and neutrons. What is the electric charge on the neutron?**

A +2
B +1
C -1
D 0

26



A student designs and builds a 1:1000 scale model of the Mt. Fuji volcano in Japan. Since Mt. Fuji is about 3,800 m high, what is the height of his scale model?

F 3,800 m
G 3.8 m
H 0.38 m
J 380 m

27 **A student wanted to test the hypothesis that adding antifreeze to water lowers the freezing point of water. The student found that pure water froze at 0°C, while the water/antifreeze mixture froze at -20°C. In order to confirm the results, the student should repeat the experiment using —**

A a container of a different size
B a different amount of water
C exactly the same set-up as the first experiment
D a different amount of antifreeze

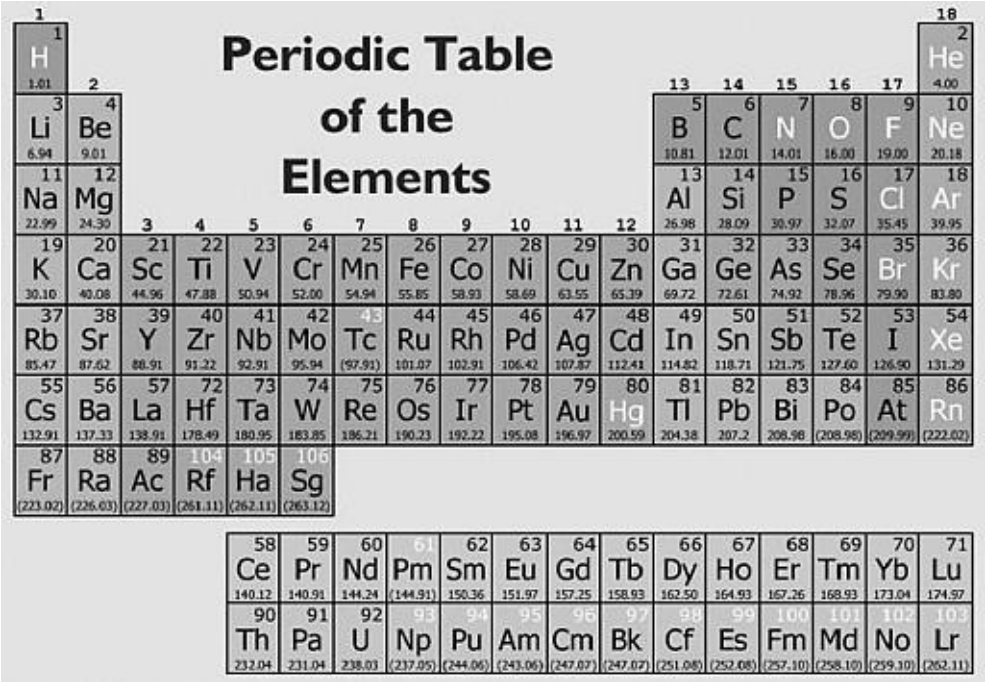
28 **When the temperature of a solid increases, which state of matter could the solid transition into?**

F Acid
G Liquid
H Plasma
J Solid

29 **A substance with a definite volume and a definite shape is classified as a —**

A plasma
B solid
C liquid
D gas

30



On the periodic table, the element sulfur is most like the element —

F phosphorous (P)
G calcium (Ca)
H chlorine (Cl)
J oxygen (O)