AIM User Manual
Evolution and Diversity
Middle School Student Assessment

Overview

The AIM Evolution and Diversity Middle School Student Assessment is a 26-item multiple-choice assessment developed for middle grades science students. The assessment is based on the *Science Framework for the 2009 National Assessment of Educational Progress* (National Assessment Governing Board, 2008) and measures understandings of selected concepts in single content area: Evolution and Diversity - differences and adaptations of organisms, preferential survival and relatedness of organisms.

This user manual describes the background, development, measurement properties, and appropriate uses of the assessment. User manuals for other AIM assessments may be found at http://www.horizon-research.com/aim/instruments/.

Background

Horizon Research, Inc. (HRI) developed the AIM Evolution and Diversity Student Assessment as part of a larger study. The project—Assessing the Impact of the MSPs: K–8 Science (AIM) was funded by the National Science Foundation under Grant no. DUE-0928177.[1] One goal of AIM was to develop instruments that researchers could use to study the theory of action that underlies much professional development for science teachers. Briefly, the model asserts that changes in teacher knowledge lead to changes in classroom practice (mediated by instructional materials), and ultimately, changes in student learning (see Figure 1). Despite the prominent role this model plays in professional development design, it has not been studied systematically, in part because of a lack of instruments. Among other products, AIM developed pairs of assessments—one for teachers and one for students—focused on the same science content areas. These pairs of assessments enable the study of relationships between teacher knowledge and student learning in specific science contexts. AIM assessments exist for four content areas: (1) evolution and diversity of life; (2) force and motion (Newton's first and second laws); (3) populations and ecosystems; and (4) properties of and changes in matter. For each content area, separate pairs of assessments were developed for elementary school and middle school levels.

[1] Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

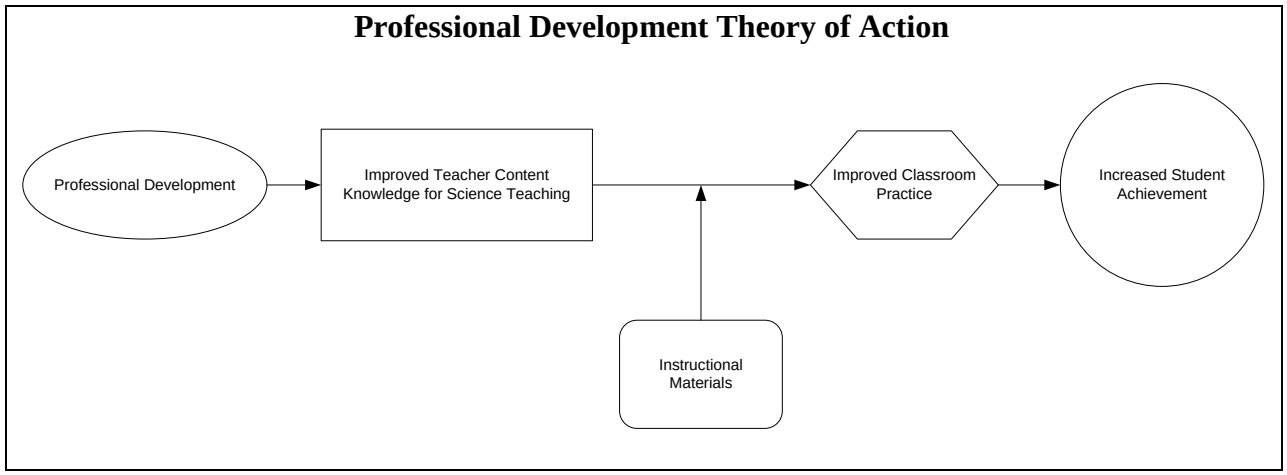


Figure 1

To enable large-scale research, HRI set out to create assessments that would be minimally burdensome, both for the test-taker and the researcher. Accordingly, HRI opted for a multiple-choice format, recognizing the limitations of such items. For instance, well-constructed, open-ended items may probe more depth of understanding than multiple-choice items, but they are more burdensome for both the researcher (in terms of scoring costs) and the test-taker (in terms of time required to complete the assessment). In addition, scoring open-ended items requires the training of raters to establish inter-rater reliability.

Development of the Evolution and Diversity Middle School Student Assessment

As described above, this development effort was part of a much larger and well-funded project, which afforded a thorough development process (see Figure 2).

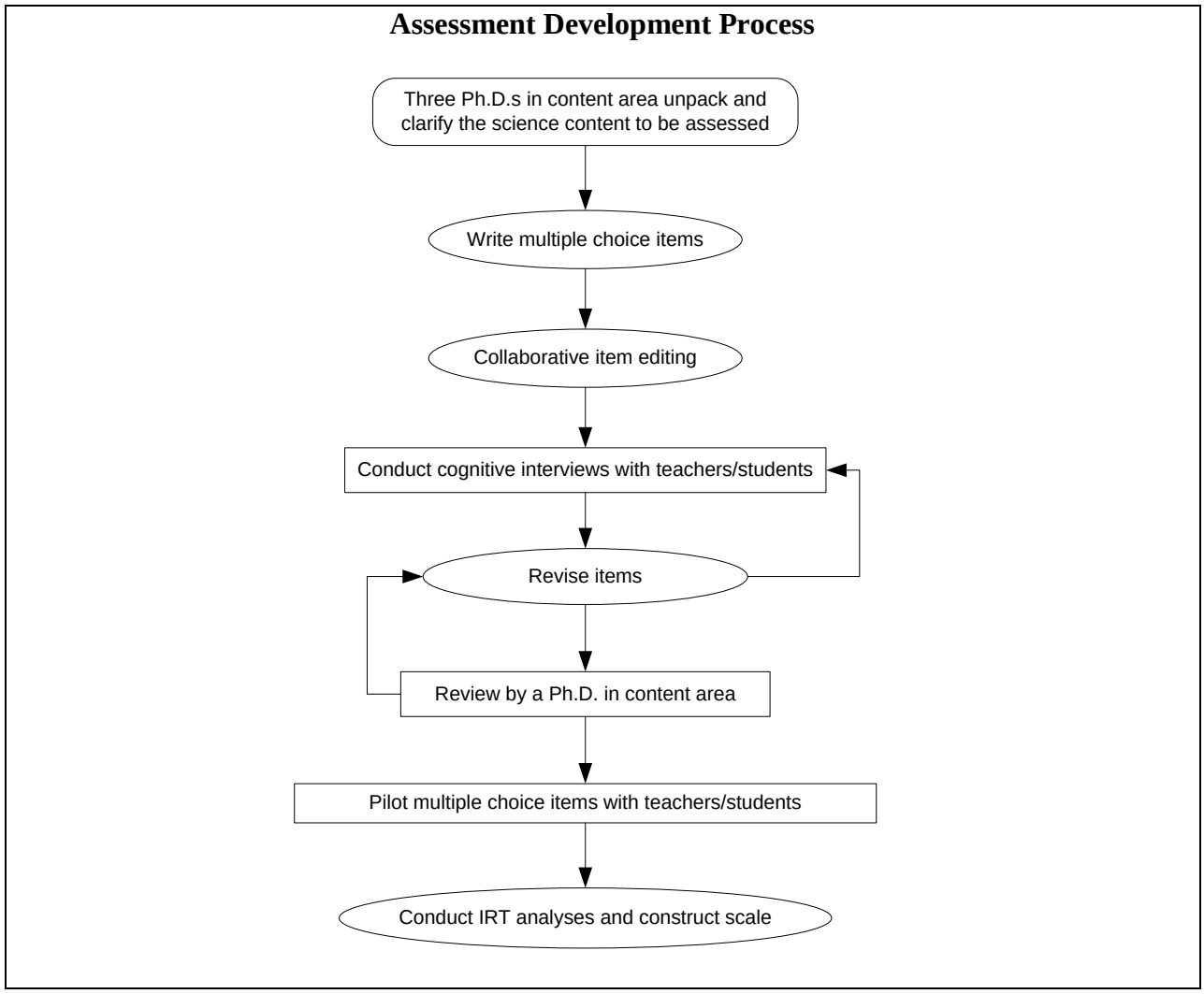


Figure 2

Clarifying the Content Domain

Development began with identifying the target content for the evolution and diversity assessments. We used the 2009 *NAEP Framework* for direction on the content of the AIM assessments. The *NAEP Framework* was based primarily on the *National Science Education Standards* (National Research Council, 1996) and the *Benchmarks for Science Literacy* (American Association for the Advancement of Science, 1993), but also reflected developments in science and policy that have taken place since those documents were published. HRI specified the assessment domain using a single strand in the *NAEP Framework*: differences and adaptations of organisms, preferential survival and relatedness of organisms. This process had three biologists/biology educators "unpack" the content into series of "sub-ideas" for middle school students. These are the ideas that were considered in developing the middle school student assessment. The final description of the content domain is shown in Table 1.

Table 1
Evolution and Diversity Content Domain

Evolution and Diversity. Differences and adaptations of organisms, preferential survival and relatedness of organisms

Sub-ideas for students:

- Different types of organisms (including plants and animals) have characteristics that enable them to survive and reproduce in different environments.
  - Different environments have different features that affect organisms' abilities to survive and reproduce. Some important features are climate, light level, soil nutrients, and the presence of other organisms.
  - Organisms that have characteristics that best meet the challenges of their environment are most likely to survive and reproduce. Therefore, the living organisms found in any particular environment are most likely to be those with characteristics that best meet the challenges in that environment.
  - A characteristic provides an advantage if it usually allows the number of individuals of that type of organism to increase over time; a characteristic provides a disadvantage if it usually causes the number of individuals of that type to decrease.
  - A characteristic or set of characteristics that provides an advantage in one environment is likely to be different than one that provides an advantage in other environments.
  - As environments change over time, the most successful adaptations will change, too.
  - Different types of organisms may meet the same environmental challenges in different ways. (For instance in a dry environment, animals may have leathery skin and plants may have a waxy covering, both of which decrease water loss.)
- Individuals of the same type differ in their characteristics, and sometimes the differences give individuals an advantage in surviving and reproducing.
  - There is variation in characteristics of organisms, even of the same type.
  - In a given environment, organisms with variations of a characteristic are likely to differ in their ability to survive and reproduce because some variations work better than others in that environment.
- Individual organisms with certain characteristics in particular environments are more likely than others to survive and have offspring.
  - Organisms of the same type in a particular environment have variation in many of their characteristics. No two individuals are exactly alike.
  - Some characteristics are more likely than others to lead to survival and reproduction in a particular environment.
  - Organisms inherit their characteristics from their parents. An individual with a characteristic that provides an advantage for survival and reproduction is more likely to have offspring that also have that advantageous characteristic. As these individuals survive and reproduce, their advantageous characteristics become more common in the population of organisms.
- When an environment changes, the advantage or disadvantage of characteristics can change.
  - Environments are not stable; they change from time to time. Environmental changes can be physical (e.g., temperature, rock slide), chemical (e.g., soil pH), biological (e.g., predators immigrating, competitors introduced) or a combination of these.
  - A characteristic that provides an advantage in one environment may provide less or more of an advantage, provide a disadvantage, or be neutral in a different environment.
- Extinction of a type of organism occurs when the environment changes and the characteristics of a type of organism are insufficient to allow its survival. Fossils indicate that many organisms that lived long ago are extinct. Extinction of organisms is common; most of the types of organisms that have lived on the Earth no longer exist.
  - Although organisms have characteristics that have variation, those variations are not always sufficient to meet the challenges of the environment. Individuals may die before reproducing or may not reproduce enough to replace all individuals who die.
  - If all members of a type of organism die, the type of organism becomes extinct.
  - Environmental changes that lead to the extinction of a type of organism often do so over time, such as through increased competition for limited resources, increased incidence of disease or parasites, or reduced ability to escape predators. Large-scale "catastrophic" changes in an environment may also produce extinctions.
  - There were many types of organisms in the past that were different than the ones that are alive today. We know this because we have a fossil record of some of the types of organisms that inhabited Earth in the past.
  - Fossils can provide scientists with records of organisms' structures that help identify the organisms and provide information about where and how they lived.
  - Types of organism can exist for long periods of time (many thousands to millions of years) before they become extinct. Becoming extinct does not mean a type of organism was unsuccessful; any type of organism may become extinct if its characteristics no longer provide an advantage in its environment.
- Similarities among organisms are found in anatomical characteristics, which can be used to infer the degree of relatedness among organisms.
  - Organisms of the same type have similar anatomical characteristics. Some of these characteristics can be seen with the naked eye, and some are microscopic.
  - The more closely related organisms are, the more similar their anatomy is likely to be; however, similar anatomy on its own does not prove relatedness.
- In classifying organisms, biologists consider details of internal and external structures to be more important than behavior or general appearance.
  - Grouping organisms according to their relatedness is called *classification*.
  - In judging similarity or differences of organisms, internal and external anatomical features should be examined.
  - Behavior and general appearance can be used to consider relatedness of organisms, but they are less reliable than specific anatomical features. Similar behaviors (e.g., flight, nest-building, earth-digging) are more likely than anatomical features to be shared by diverse organisms.
- Similar characteristics are not always an indication of relatedness because sometimes unrelated organisms may evolve similar adaptations to similar environmental constraints in similar ways. If the similar characteristics evolved from different structures that have different developmental origins, the process of evolving similar characteristics is called *convergent evolution.*

Item Development

HRI staff drafted items individually then met to edit them collaboratively. As the pool of items grew, we began recruiting middle school students for telephone cognitive interviews. We typically interviewed three students on each item in the pool using the interview protocol shown in Figure 3. After a round of interviews, HRI staff met to discuss students' responses and feedback. If substantive edits were made to an item, we interviewed additional students about the revised version. When interviews suggested no further edits were needed, we asked a content expert to review all of the items in the pool for content accuracy.

AIM Student Assessment Items
Cognitive Interview Protocol

Prologue Script:

We are developing test questions for middle school students who have been studying evolution and diversity, and we need your help to get the questions just right. I realize that you may not have studied some of this yet in school, and I don't expect you to get all of the answers right. If you get a few wrong, it will help me know whether we have written the answer choices well. You can ask me to explain any words or situations that may be unfamiliar or confusing, but I can't give you the answer to any of the questions until the end of the interview. Remember, the point is to help us write a good test, not to test what you do or don't know. You won't get a grade or anything like that on the test. Do you have any questions before we get started? If at any point in the interview you would like to stop, just say so.

Procedure:

- Ask student to read aloud and "think aloud" as they read the questions and answer choices, if they are comfortable doing so. Remind the student to go back and reread the question to himself/herself if he/she needs to. If reading the question aloud is too distracting or uncomfortable, allow the student to read the question to himself/herself.

- It is not necessary to time how long it takes for the student to arrive at an answer, but if it takes an especially long time on a question, please make a note of it in the comment area of the notes.

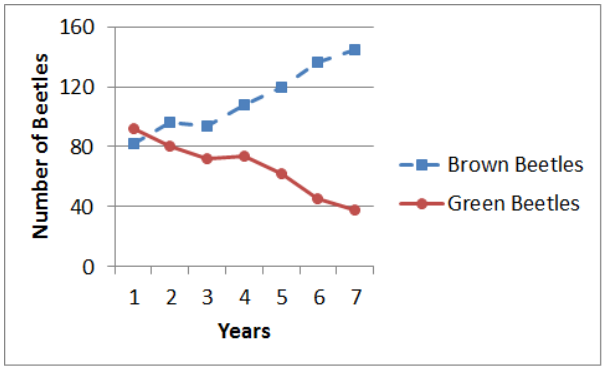
- For each item, ask:
  1. Why did you choose that answer? (probe for words or diagrams they keyed in on, as well as their thinking behind the response)
  2. What did you think of each of the other answer choices?
  3. Was there an answer choice you were expecting to see, but did not? What was it?
  4. Were there any words or diagrams you did not really understand, or situations that made the question confusing?
  5. Is there anything about the question that did not confuse you, but that you think might confuse other students?
  6. Do you have any other comments on the item?

Figure 3

An example student assessment item resulting from this process is shown in Figure 4. (correct answer is A)

Evolution and Diversity Item

Individuals of a type of beetle that live in the rainforest can be either green or brown. Birds that live there find and eat the green beetles more easily than the brown ones. The graph below shows that the number of green beetles goes down and the number of brown beetles goes up over time.



What might explain why there are more brown than green beetles after 7 years?

A. The trait of brown coloration is more likely to lead to survival and reproduction in the rainforest.
B. The trait of brown coloration is more likely to lead to survival and reproduction in any environment.
C. The number of brown beetles goes up due to chance; eventually the number of brown beetles will go down and the number of green beetles will go up.
D. The number of green beetles goes down because organisms of the same type in a particular environment have many different traits.

Figure 4

This item illustrates some features common to all AIM student assessment items. It is not included in the AIM assessment, but is shown here to illustrate item features. This example item may be flawed and is not intended to be used in any assessments. As mentioned previously, all items are multiple choice. All include only four choices and preclude as choices "none of the above," "all of the above," or multiple correct answers such as, "A and B but not C."

Pilot

We selected 35 items to pilot with approximately 799 students of teachers recruited from mailing lists of middle grades teachers across the country. The pilot was administered as a paper form by recruited teachers.

Table 2
Characteristics of the Pilot Test Sample

	Percent of Students
Grade Level[†]	
6th grade	20
7th grade	46
8th grade	35
English is primary language	
Yes	94
No	6
Gender	
Female	53
Male	47
Race/Ethnicity[†]	
American Indian or Alaskan Native	4
Asian	5
Black or African American	21
Hispanic or Latino	14
Native Hawaiian or Other Pacific Islander	1
White	69

[†] Percentages may add up to more than 100 as students could select multiple categories.

Measurement Properties of the Assessment

Following is a description of the content coverage of the assessment, information about the validity and reliability of the assessment, and the results of the item response theory (IRT) analysis.

Content Coverage

Using results from the pilot, 26 items were selected for the final form. The distribution of items by sub-idea is shown in Table 3. The number of items totals to more than 26 because one item may address more than one sub-idea. There are fewer sub-ideas in Table 3 than in the content unpacking (see Table 1), as limiting the assessment to a total of 26 items required restricting the coverage of sub-ideas. In some cases a sub-idea may not be represented in the final assessment because it was deemed to be less central than others. In other cases, items associated with the sub-idea did not perform as well as others in the pilot study.

Table 3
Number of Items Addressing Each Sub-Idea

Sub-Ideas:	Number of Items
A. Individual organisms with certain characteristics in particular environments are more likely than others to survive and have offspring.	1
B. Organisms of the same type in a particular environment have variation in many of their characteristics. No two individuals are exactly alike.	1
C. Organisms inherit their characteristics from their parents. An individual with a characteristic that provides an advantage for survival and reproduction is more likely to have offspring that also have that advantageous characteristic. As these individuals survive and reproduce, their advantageous characteristics become more common in the population of organisms.	2
D. When an environment changes, the advantage or disadvantage of characteristics can change.	2
E. Environments are not stable; they change from time to time. Environmental changes can be physical (e.g., temperature, rock slide), chemical (e.g., soil pH), biological (e.g., predators immigrating, competitors introduced) or a combination of these.	1
F. A characteristic that provides an advantage in one environment may provide less or more of an advantage, provide a disadvantage, or be neutral in a different environment.	2
G. Extinction of a type of organism occurs when the environment changes and the characteristics of a type of organism are insufficient to allow its survival. Fossils indicate that many organisms that lived long ago are extinct. Extinction of organisms is common; most of the types of organisms that have lived on the Earth no longer exist.	3
H. Although organisms have characteristics that have variation, those variations are not always sufficient to meet the challenges of the environment. Individuals may die before reproducing or may not reproduce enough to replace all individuals who die.	1
I. If all members of a type of organism die, the type of organism becomes extinct.	1
J. Environmental changes that lead to the extinction of a type of organism often do so over time, such as through increased competition for limited resources, increased incidence of disease or parasites, or reduced ability to escape predators. Large-scale "catastrophic" changes in the environment may also produce extinctions.	1
K. There were many types of organisms in the past that were different than the ones that are alive today. We know this because we have a fossil record of some of the types of organisms that inhabited Earth in the past.	3
L. Fossils can provide scientists with records of organisms' structures that help identify the organisms and provide information about where and how they lived.	2
M. Types of organism can exist for long periods of time (many thousands to millions of years) before they become extinct. Becoming extinct does not mean a type of organism was unsuccessful; any type of organism may become extinct if its characteristics no longer provide an advantage in its environment.	2
N. Similarities among organisms are found in anatomical characteristics, which can be used to infer the degree of relatedness among organisms.	1
O. Organisms of the same type have similar anatomical characteristics. Some of these characteristics can be seen with the naked eye, and some are microscopic.	2
P. The more closely related organisms are, the more similar their anatomy is likely to be; however, similar anatomy on its own does not prove relatedness.	2
Q. In judging similarity or differences of organisms, internal and external anatomical features should be examined.	1
R. Behavior and general appearance can be used to consider relatedness of organisms, but they are less reliable than specific anatomical features. Similar behaviors (e.g., flight, nest-building, earth-digging) are more likely than anatomical features to be shared by diverse organisms.	1
S. Similar characteristics are not always an indication of relatedness because sometimes unrelated organisms may evolve similar adaptations to similar environmental constraints in similar ways. If the similar characteristics evolved from different structures that have different developmental origins, the process of evolving similar characteristics is called *convergent evolution*.	1

Table 4 shows the answer key and content association for each item on the assessment. The letter "P" denotes a primary association with the sub-idea being targeted by the item. An "S"

denotes a secondary association with a sub-idea that is also necessary in order to answer the item correctly, but is not the primary idea being assessed.

Table 4
Answer Key and Sub-Idea Associations

Item #	Key	Sub-Idea																		
		A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**
1	A											P								
2	B						P													
3	C		P																	
4	A			P			S													
5	A													P						
6	B																	P	S	
7	A				P															
8	C															P	S			
9	D												P							
10	D							P												
11	C	P																		
12	D					P														
13	D																P			
14	B											P								
15	C								P											
16	C				P															
17	B							P												
18	A														P					
19	A							P												
20	B															P				
21	C											S	P							
22	D										P									
23	C																			P
24	D													P						
25	D			P																
26	B									P										
Primary:		1	1	2	2	1	1	3	1	1	1	2	2	2	1	2	1	1	0	1
Secondary:		0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0
Total:		1	1	2	2	1	2	3	1	1	1	3	2	2	1	2	2	1	1	1

Validity
Three lines of evidence support the argument that the assessment is a valid measure of students' knowledge of these evolution and diversity ideas. First, cognitive interviews with students established that students interpret the items as intended and that they must use their knowledge of content to answer the items correctly. Second, a content expert (individual with a Ph.D. in biology) reviewed the assessment items to ensure content accuracy. Third, factor analysis indicates that all items on the assessment measure a single dominant trait. HRI termed this trait "content knowledge about evolution and diversity."

Reliability
Both classical test and item response theory (IRT) analyses were conducted on the pilot data and those results were used to select items for the final assessment. The assessment has an IRT reliability of 0.84; reliabilities above 0.60 are generally considered acceptable for making judgments about groups (higher reliabilities are required for making high-stakes decisions about individuals.

Speededness
In the pilot, teachers were instructed to give their students 50 minutes or the length of the class period (whichever was shorter) to complete the test. There was no evidence of speededness.

Using the Assessment

The AIM Evolution and Diversity Student Assessment is available at no cost through an online process to those who agree to the terms of use (see the Appendix). To complete the terms of use agreement, visit http://www.horizon-research.com/aim/instruments/.

Appropriate Use
The AIM Evolution and Diversity Student Assessment yields a score for each individual. However, the assessment is not valid for making judgments about individuals based on those scores. For instance, assigning student grades based on scores is not a valid use of the assessment. The assessment was not validated for such purposes.

HRI developed the assessment for use in research contexts involving groups of teachers. Appropriate uses with sufficiently large groups of teachers (20 or more) include:

- Measuring the change in group mean from pre-workshop to post-workshop;
- Comparing the gains of treatment and control groups; and
- Researching the relationship between teacher knowledge and other variables (e.g., student learning).

Amount of Time Required to Complete the Assessment
Although there is no evidence of speededness, it is recommended that at least 45 minutes be allowed for completing the assessment.

Computing Scores

Scores may be computed either as number correct or percent correct. Results of an item-response theory (IRT) analysis are shown in Table 5. This table can be used to convert a raw score in terms of number correct to the corresponding scale score.

Table 5
Assessment Score Conversions

Raw Score	Scale Score
0	0
1	13
2	22
3	27
4	30
5	33
6	36
7	38
8	40
9	42
10	44
11	46
12	48
13†	50†
14	52
15	53
16	55
17	57
18	59
19	61
20	64
21	66
22	69
23	73
24	78
25	86
26	100

† Mean value

References

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

National Assessment Governing Board, U.S. Department of Education. (2008) *Science framework for the 2009 national assessment of educational progress*. Washington, DC: U.S. Government Printing Office.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

Appendix
Terms of Use Agreement
Evolution and Diversity Middle School Student Assessment

By using the AIM Evolution and Diversity Student Assessment developed by Horizon Research, Inc. (HRI), you agree to abide by the stipulations below concerning use, test security, test administration, and citations.

Use of the Assessment

The Evolution and Diversity Student assessment may be used to gauge growth in knowledge about a specific content area as a result of an intervention such as professional development, curriculum use, or mentoring. It may also be used to learn about the contribution of teacher knowledge to student knowledge and classroom instruction.

We ask that you abide generally by the standards put forward in the *Standards for Educational and Psychological Testing* (AERA/APA, 1999).

You may not use the assessment to evaluate individuals. Assessment results may not be associated with any high-stakes consequence such as tenure, pay, hiring, or grades. The assessments were not developed for making decisions/judgments about individuals.
You should also refrain from using these measures to publicly demonstrate teachers' ability or lack of ability in science, which may adversely affect willingness to participate in future studies.

IRB and/or District/School Study Approval

It is your responsibility to obtain proper IRB and/or the appropriate district/school approval for your study and to follow the necessary requirements for obtaining principal, teacher, parent, and/or student permission/approval to administer to the assessment(s).

Responsibilities to Teachers and Students

Your responsibilities to study participants will largely depend on the details of the IRB and/or district/school approval of your study. In most cases, completion of the assessment will be strictly voluntary. As such, participants should be informed of the voluntary nature of the study. Teachers should be assured that if their students' data are not anonymous, individual identities will be kept strictly confidential; i.e., an individual's score or responses will never be reported in association with his or her name or any other identifying information. To encourage a high response rate among teachers, it may be helpful to:

- Clearly explain what the data will be used for and why the data are important for your study;
- Explain that there are no high-stakes consequences associated with completing the assessment; and
- Offer teachers compensation for time spent outside of the regular school day completing the assessment.

Test Security

The AIM Evolution and Diversity Student Assessment may NOT be shared without prior authorization from HRI. Anyone who administers the assessment must agree to:

- Refrain from using any non-released item in any presentation, paper, article, or other public forum. Items are expensive to develop and pilot, and we are attempting to keep our item pool secure.
- Refrain from distributing copies of any non-released item to individuals other than participants in your research project.
- Refrain from using the assessment, in original or in copied form, to provide test-taking practice or to enhance test-taking skills.
- Refrain from using test items, actual or similar, for discussion or review.

(HRI acknowledges that, in some cases, school administrators and IRBs may require that the test materials be reviewed prior to granting permission for study participants to take the test. Such a review is not considered a violation of this Test Security Policy as long as the other provisions of this policy are not violated.)

Citing AIM Assessments

In any writing in which data from HRI's AIM assessments are included, the following citation must be used:

The assessment was developed by the Assessing the Impact of the MSPs: K–8 Science (AIM) project at Horizon Research, Inc., funded by the National Science Foundation under grant number DUE-0928177. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation or Horizon Research, Inc.

By signing below, I acknowledge that I have read the user manual, and I agree to abide by terms of use described above.

_____

Printed Name Signature Date

Address: _____

Street City State Zip code

Phone number (including area code): _____

Your email address: _____