

AIM User Manual

Force and Motion

Elementary School Teacher Assessment

Overview

The AIM Force and Motion Elementary School Teacher Assessment is a 30-item multiple-choice assessment developed for teachers of elementary grades science. 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 concepts in two related content areas.

1. Motion at the macroscopic level
 - descriptions of position and motion;
 - speed as a quantitative description of motion; and
 - graphical representations of speed
2. Forces affecting motion
 - the association of changes in motion with forces;
 - the association of objects falling toward Earth with gravitational force;
 - qualitative descriptions of magnitude and direction as characteristics of forces;
 - addition of forces;
 - contact forces;
 - forces that act at a distance; and
 - net force on an object and its relationship to the object's motion.

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 Force and Motion Teacher 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

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.

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.

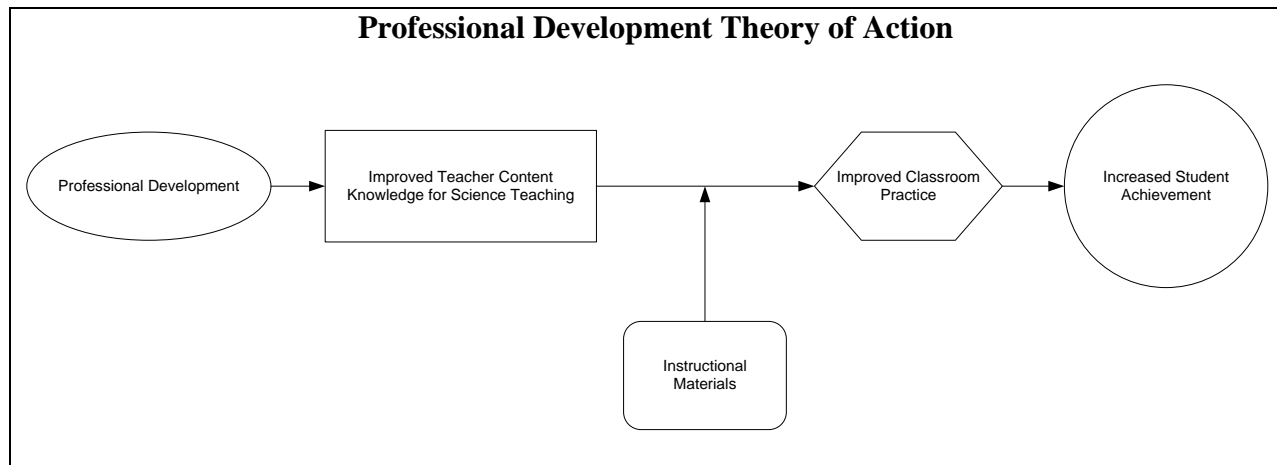


Figure 1

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 Force and Motion Elementary School Teacher 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).

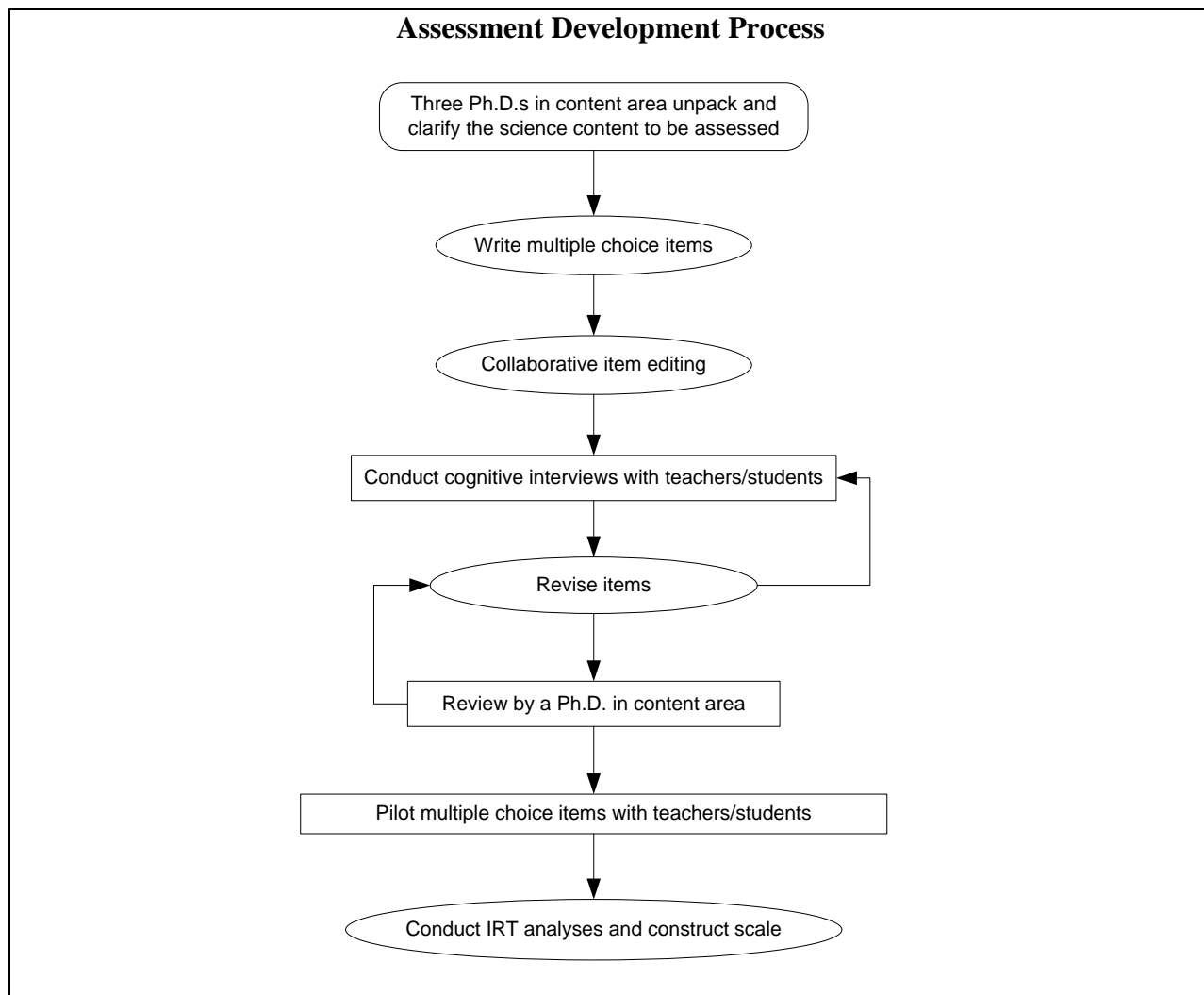


Figure 2

Clarifying the Content Domain

Development began with identifying the target content for the force and motion 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 two strands in the *NAEP Framework*: (1) motion at the macroscopic level and (2) forces affecting motion. This process had three physicists/physics educators “unpack” the content into series of “sub-ideas” for upper elementary school students. Additional sub-ideas that are important for teachers to understand in order to teach the upper elementary school student ideas were also specified by the content experts. Both sets of ideas were considered in developing the middle school teacher assessment. The final description of the content domain is shown in Table 1.

Table 1
Force and Motion Content Domain

<p>Motion at the Macroscopic Level. From descriptions of position and motion to speed as a quantitative description of motion and graphical representations of speed.</p> <p>Sub-ideas for students</p> <ul style="list-style-type: none"> • An object’s position can be described by locating the object relative to other objects or a background. • The description of an object’s motion from one observer’s view may be different from that reported from a different observer’s view. • An object is in motion when its position is changing. • The speed of an object is defined by how far it travels divided by the amount of time it took to travel that far. • A change in motion is a change in its speed, or its direction, or both <p>Sub-ideas for teachers</p> <ul style="list-style-type: none"> • • An object’s motion can be described completely by its speed and the direction in which it is moving. • An object’s position can be measured and graphed as a function of time. • An object’s speed can be measured and graphed as a function of time. • Dividing the distance traveled by the time taken gives the average speed of an object, as opposed to the speed at a particular instant.
<p>Forces Affecting Motion. From the association of changes in motion with forces and the association of objects falling toward Earth with gravitational force to qualitative descriptions of magnitude and direction as characteristics of forces, addition of forces, contact forces, forces that act at a distance, and net force on an object and its relationship to the object’s motion.</p> <p>Sub-ideas for students</p> <ul style="list-style-type: none"> • The motion of objects can be changed by pushing or pulling. • The size of the change is related to the size of the force (push or pull) and the weight (mass) of the object on which the force is exerted. • When an object does not move in response to a push or a pull, it is because another push or pull is being applied by the environment. • Earth pulls down on all objects with a force called gravity. • With a few exceptions (helium filled balloons), objects fall to the ground no matter where the object is on Earth. • A force is a push or pull exerted on one object by another object when they interact with one another. • Some forces between objects act when the objects are in direct contact or when they are not touching. • Forces have magnitude and direction. • Forces can be added. The net force on an object is the sum of all the forces acting on the object. • A non-zero net force on an object changes the object’s motion; that is, the object’s speed and/or direction of motion changes. • A net force of zero on an object does not change the object’s motion. • The force of friction acts to oppose the relative motion of two objects in contact. <p>Sub-ideas for teachers</p> <ul style="list-style-type: none"> • Air exerts a weak upward force on all objects. For some very light but large objects this force may be equal to or stronger than the gravitational force and so they will not fall to the ground when released and may even move upward. • An object’s mass is an inherent property, distinct from (but proportional to) its weight. The mass of an object is a measure of the amount of material comprising it. • The term weight refers to the strength of the gravitational force exerted by the Earth on an object. • Supporting objects, such as hands, tables, and shelves, exert upward forces on objects on top of them. These supporting forces exactly balance the downward pull of gravity and so these objects do not fall.

Types of Teacher Assessment Items

The sections that follow discuss the three types of teacher multiple-choice items included in the assessment:

1. knowledge of science content (Level 1 items);
2. using content knowledge to analyze/diagnose student thinking (Level 2 items); and
3. using content knowledge to make instructional decisions (Level 3 items).

The example items below are not included in the AIM assessment, but are shown here to illustrate item features. These example items may be flawed and are not intended to be used in any assessments.

Knowledge of science content

All of the AIM items for teachers assess knowledge of science content, but the most basic type of question attempts to isolate disciplinary content knowledge from a teachers’ ability to apply that knowledge in making instructional decisions. An example of a “Level 1” item is shown in Figure 3 (correct answer is B).

Level 1 Item

A teacher gave students the following question.

Student Question

A hockey puck is hit and moves across the ice. Imagine there is no friction and no air resistance. What must be done to make the puck move slower and slower?

Which of the following is a correct response?

- A. Apply a force on the puck in the direction of its motion.
- B. Apply a force on the puck in the direction opposite its motion.**
- C. Remove the force on the puck that is keeping it moving.
- D. Nothing, the puck will move slower and slower on its own

Figure 3

This item illustrates some features common to all AIM teacher assessment items. As mentioned previously, all of the items are multiple-choice. In addition, all items include only four choices, and answer choices are never worded as “all of the above” or “none of the above.” Multiple correct answers, such as “A and B, but not C,” are also not used. Perhaps most importantly, all of the items are set in an instructional context. The intent in using these contexts was two-fold. First, we wanted teachers to feel like they were taking a test that was written for them, as opposed to, for example, a test constructed for undergraduates. The second goal was for teachers to recognize in the items the kind of work they do every day, making it more likely that they would intellectually engage with the items.

Using science content knowledge to analyze/diagnose student thinking

“Level 2” items require teachers to apply their content knowledge in analyzing or diagnosing a sample of student thinking. Figure 4 shows an illustrative item (correct answer is C).

Level 2 Item

A teacher pushes on a very large box. The students observe that the box does not move. The teacher asks the students to explain why the box does not move.

One student says "The box does not move because you have not transferred enough of your force to it yet."

What does this response indicate about this student's thinking?

- A. The student does not understand that a force is a push or a pull.
- B. The student does not understand that the box pushes back with an equal but opposite force.
- C. **The student does not understand that there must be another force acting on the box.**
- D. The student seems to understand why the box does not move.

Figure 4

Certainly a teacher must understand the science content in order to select the correct answer choice for Level 2 items. However, additional analysis of the question is required because more than one of the choices includes a correct physics statement, unlike the Level 1 item in Figure 3. In Figure 4, the statements in choices A, B, and C are correct in terms of the physics, but only C applies to what the students said. This feature is present in all Level 2 items and makes the cognitive load of these items higher than that of Level 1; teachers must evaluate the students' thinking in relation to the physical scenario in order to determine which answer choice is correct.

Using content knowledge to make instructional decisions

"Level 3" items ask teachers to apply their content knowledge in choosing among instructional moves. A sample Level 3 item is shown in Figure 5 (correct answer is C).

Level 3 Item

A teacher gives her students the following scenario.

Scenario

Three books are sitting on a table. Each has a different mass. If I push each book just as hard for the same amount of time, which book's motion will change the most?

Most students agree that all of the books will have the same change in motion because the same force is applied to all of the books.

Which of the following would be the best next step to move these students forward in their understanding about the effect of forces on motion?

- A. Drop all three books from the same height at the same time and see which book hits the ground first.
- B. Push the books across different surfaces that have varying amounts of friction.
- C. Show a video that illustrates how the strength of an applied force and the mass of an object affect an object's motion.**
- D. Have a class discussion about the difference between mass and weight.

Figure 5

Level 3 items have the highest cognitive load; teachers must evaluate the physical scenario, the student's thinking in relation to the physical scenario, and then evaluate each instructional choice. As with Level 2 items, more than one answer choice is consistent with a correct interpretation of the physics content, but only one has a correct physics statement *and* is relevant to the instructional context. Although the cognitive load of Level 3 items is demanding, it is a small fraction of the demand placed on a teacher managing the learning of a classroom of students.

Item Development

HRI staff drafted items individually then met to edit them collaboratively. As the pool of items grew, we began recruiting elementary school teachers for telephone cognitive interviews. We interviewed at least three teachers on each item in the pool using the interview protocol shown in Figure 7. After a round of interviews, HRI staff met to discuss teachers' feedback. If substantive edits were made to an item, we interviewed additional teachers 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 Teacher Assessment Items Cognitive Interview Protocol

Prologue Script:

Thank you for agreeing to let us interview you. As we explained in the email, we are developing a test for elementary school science teachers, and we need your help to refine the test questions. I don't expect you to get all of the answers right. The point is to help us write a good test, not to test what you do or don't know. Do you have any questions before we get started? Remember that all of your answers are confidential. If you decide you would like to stop at any point, just say so.

Procedure:

- Ask teacher to read aloud and “think aloud” as they read the questions and answer choices, if they are comfortable doing so.

- 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 teachers?
 6. Do you have any other comments on the item?

Figure 6

The cognitive interviews revealed distinct patterns of errors in teacher responses to the Level 2 items (using content knowledge to analyze/diagnose student thinking). Some teachers chose an answer that included student thinking they were familiar with, whether or not it represented the thinking of the student in the item. Others chose a statement that was correct in terms of the physics, but not in relation to the student's thinking.

Interviews also suggested some common errors teachers make with Level 3 items (using content knowledge to make instructional decisions). First, they often saw more than one of the instructional choices (including the correct one) as equally good, particularly when the item requires teachers to evaluate which question should be asked next. When the choices are about actual activities, as in the example in Figure 5, teachers sometimes get bogged down in the details of the choices. For instance, they may rule out a choice that includes low-friction carts because they do not have access to such equipment, regardless of whether the activity would help move the student's thinking forward. Finally, teacher beliefs about effective instruction may get in the way, even when they seem to understand the content targeted by the item. For example, teachers often choose a hands-on activity, even if it does not address the student's thinking.

Pilot

We selected 40 items to pilot with 330 teachers recruited from mailing lists of elementary grades teachers across the country. The pilot was administered via the Internet. Approximately 20 percent of the sample was comprised of teachers who indicated they had taught the content at the

middle school level to ensure that some respondents would be at the upper end of the knowledge spectrum.

Table 2
Characteristics of the Pilot Test Sample (N = 330)

	Percent of Teachers
Grade Level Taught in 2009–10	
Kindergarten	2
1 st grade	2
2 nd grade	3
3 rd grade	26
4 th grade	29
5 th grade	29
6 th grade	12
7 th grade	12
8 th grade	12
9 th grade	1
10 th grade	0
11 th grade	1
12 th grade	0
Taken a college-level introductory physics course	
Yes	80
No	20
Gender	
Female	88
Male	13
Race/Ethnicity	
American Indian or Alaskan Native	1
Asian	1
Black or African American	3
Hispanic or Latino	2
Native Hawaiian or Other Pacific Islander	0
White	93

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, 30 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 30 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 30 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. An object's position can be described by locating the object relative to other objects or a background.	2
B. The description of an object's motion from one observer's view may be different from that reported from a different observer's view.	2
C. An object is in motion when its position is changing.	1
D. The speed of an object is defined by how far it travels divided by the amount of time it took to travel that far.	2
E. The motion of objects can be changed by pushing or pulling.	1
F. The size of the change is related to the size of the force (push or pull) and the weight (mass) of the object on which the force is exerted.	3
G. When an object does not move in response to a push or a pull, it is because another push or pull is being applied by the environment.	1
H. Earth pulls down on all objects with a force called gravity.	3
I. With a few exceptions (helium filled balloons), objects fall to the ground no matter where the object is on Earth.	2
J. A change in motion is a change in its speed, or its direction, or both.	1
K. A force is a push or pull exerted on one object by another object when they interact with one another.	1
L. Dividing the distance traveled by the time taken gives the average speed of an object, as opposed to the speed at a particular instant.	1
M. Air exerts a weak upward force on all objects. For some very light but large objects this force may be equal to or stronger than the gravitational force and so they will not fall to the ground when released and may even move upward.	1
N. An object's mass is an inherent property, distinct from (but proportional to) its weight. The mass of an object is a measure of the amount of material comprising it.	2
O. The term weight refers to the strength of the gravitational force exerted by the Earth on an object.	1
P. Supporting objects, such as hands, tables, and shelves, exert upward forces on objects on top of them. These supporting forces exactly balance the downward pull of gravity and so these objects do not fall.	1
Q. An object's motion can be described completely by its speed and the direction in which it is moving.	1
R. An object's position can be measured and graphed as a function of time.	1
S. An object's speed can be measured and graphed as a function of time.	2
T. Some forces between objects act when the objects are in direct contact or when they are not touching.	1
U. Forces have magnitude and direction.	1
V. Forces can be added. The net force on an object is the sum of all the forces acting on the object.	2
W. A non-zero net force on an object changes the object's motion; that is, the object's speed and/or direction of motion changes.	1
X. The force of friction acts to oppose the relative motion of two objects in contact.	2

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	T	U	V	W	X		
1	A	P																									
2	B	P																									
3	C		P																								
4	A																			P							
5	A																			P							
6	D		P																								
7	A			P																							
8	B				P							S															
9	A				P																						
10	C																			P							
11	B																		P								
12	B									P																	
13	C					P																					
14	C								P						P												
15	A								P					P													
16	D								P																		
17	C									P																	
18	B																		P								
19	A									P																	
20	C															P	P										
21	D						P																				
22	A							P																			P
23	D											P											P				
24	D																					P					
25	D																							P			
26	C																							P			
27	B																										P
28	B																										P
29	D							P																			
30	C							P																			
Primary		2	2	1	2	1	3	1	3	2	1	1	0	1	2	1	1	1	1	2	1	1	2	1	1	2	2
Secondary		0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total		2	2	1	2	1	3	1	3	2	1	1	1	1	2	1	1	1	1	2	1	1	2	1	1	2	2

Validity

Three lines of evidence support the argument that the assessment is a valid measure of teachers' knowledge of these force and motion ideas. First, cognitive interviews with teachers established that teachers interpret the items as intended and that teachers must use their knowledge of content to answer the items correctly. Second, a content expert (individual with a Ph.D. in physics) 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 for teaching about force and motion."

Reliability

Both classical test and item response theory (IRT) analyses were conducted on data and those results were used to select items for the final assessment. The analyses were repeated with data collected using the final assessment. The assessment has an IRT reliability of 0.80; 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

The pilot was not a timed administration. As such, there is no information about speededness.

Using the Assessment

The AIM Force and Motion Teacher 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 Force and Motion Teacher Assessment yields a score for each individual. However, the assessment is not valid for making judgments about individuals based on those scores. For instance, evaluating teacher performance 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
- Investigating the relationship between teacher knowledge and other variables (e.g., student learning).

Amount of Time Required to Complete the Assessment

As described above, the pilot test was administered on-line and was not timed. Although there is no evidence of speededness, it is recommended that at least 30 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	12
2	20
3	25
4	28
5	32
6	34
7	37
8	39
9	41
10	43
11	44
12	46
13	48
14	50
15	51
16	53
17 [†]	54 [†]
18	56
19	58
20	60
21	61
22	63
23	65
24	68
25	70
26	73
27	77
28	81
29	88
30	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. (2009) *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

Force and Motion Elementary School Teacher Assessment

By using the AIM Force and Motion Teacher 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 Force and Motion Teacher 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 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.

