



# Aligning science assessment standards: New Mexico and the National Assessment of Educational Progress (NAEP)



Institute of Education Sciences  
U.S. Department of Education





# Aligning science assessment standards: New Mexico and the National Assessment of Educational Progress (NAEP)

July 2007

**Prepared by**

**Michael Timms  
WestEd**

**Steven Schneider  
WestEd**

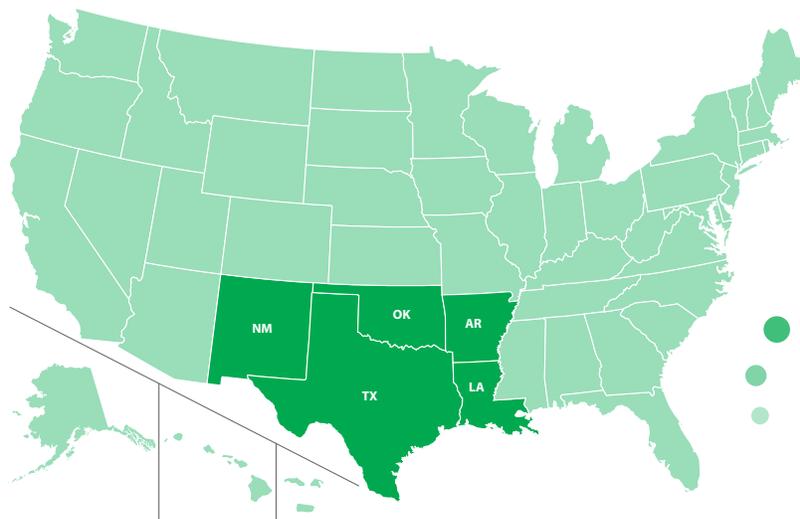
**Cindy Lee  
WestEd**

**Eric Rolfhus  
REL Southwest**



Institute of Education Sciences

U.S. Department of Education



**Issues & Answers** is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

July 2007

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0017 by Regional Educational Laboratory Southwest administered by Edvance Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Timms, M., Schneider, S., Lee, C., & Rolffhus, E. (2007). *Aligning science assessment standards: New Mexico and the National Assessment of Educational Progress (NAEP)* (Issues & Answers Report, REL 2007–No. 021). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

This report is available on the regional educational laboratory web site at <http://ies.ed.gov/ncee/edlabs>.

## Summary

# Aligning science assessment standards: New Mexico and the National Assessment of Educational Progress (NAEP)

**This policy research document is intended for New Mexico policymakers to use when examining possible changes to the state assessment's alignment with the National Assessment of Educational Progress (NAEP). The 2009 NAEP test is not yet in existence, so the purpose of this report is to give policymakers a headstart in determining where they might, if they so decide, begin to make changes in their assessment standards and specifications to develop an assessment system more closely aligned with that used for the NAEP.**

Overall, reviewers found New Mexico's science assessment framework to be fairly well aligned with the NAEP framework. For grade 4, all NAEP content items are to some degree addressed by New Mexico's science assessment framework, with no ratings of 1 and an overall alignment rating of 2.2 (a rating of 1 indicates no alignment and a rating of 3, full alignment). For grade 8 the majority of NAEP content statements are partially aligned with the content in the New Mexico science assessment framework, and the overall alignment rating is 2.1, mostly because the NAEP standards typically contain more detail and more specific content than the corresponding New Mexico standards. In the comparison with NAEP grade 12, New Mexico was given

an overall alignment rating of 2.3, indicating a fairly high degree of alignment; only two NAEP content statements are not addressed by corresponding New Mexico standards.

A rating of partial alignment between New Mexico and the NAEP was due primarily to reviewers finding that the state often implied content that was stated explicitly by the NAEP and that the NAEP often provided more specific content items or more detail. However, reviewers believed that New Mexico was, on the whole, fairly well aligned with the NAEP.

This report reveals current alignment issues between the state's tests and the future NAEP tests and may be especially important to policymakers who are considering revising science standards and assessments in line with No Child Left Behind requirements for state science tests in elementary, middle, and high schools. If state policymakers wish to increase the alignment between the state assessments and the NAEP, areas to consider are increasing Earth and space science coverage in grade 8 and including a wider variety of test item types, such as hands-on and interactive computer tasks. Revising assessments requires considerable time and resources, so policymakers must carefully consider their capacity to make changes and the degree to which such changes will benefit students.

---

### Grade 4 alignment

All NAEP content items are addressed to some degree by New Mexico's science assessment framework, and New Mexico is fairly well aligned with the NAEP. The New Mexico statements are most often partially aligned with NAEP statements, primarily because New Mexico covers the NAEP content at higher grade levels. However, raters believed that New Mexico's content was spiraled well through all grade levels. The New Mexico framework also contains many content items that are not listed in the NAEP's content statements, including those within New Mexico's scientific thinking and practice section, which reviewers praised for its promotion of rigorous scientific methodology. The overall alignment rating for grade 4 is 2.15.

---

### Grade 8 alignment

The majority of NAEP grade 8 content statements were partially aligned with the content in the New Mexico science assessment framework. Partial alignment was due mainly to the fact that NAEP standards typically contain more detail and more specific content than the corresponding New Mexico standards. The overall alignment rating is 2.1, indicating partial alignment.

---

### Grade 12 alignment

New Mexico's grade 11 performance standards are fairly well aligned with the NAEP's grade 12 content; only two NAEP statements were unaddressed by New Mexico, and the overall alignment rating is 2.3. Reviewers found most content to be partially aligned, and the

majority of aligned New Mexico content was found to imply content explicitly stated by the NAEP.

---

### Test specifications

The New Mexico test blueprints ensure that testing student knowledge and skills does not rely solely on multiple-choice items by including short and longer constructed-response items. That enables a wider range of knowledge types to be tested than with multiple-choice alone. New Mexico breaks down the content differently from the NAEP (by number of items and number of points), so it is hard to directly compare the relative amounts of testing time devoted to each topic. However, when focusing just on the three topics tested in the NAEP, the New Mexico blueprints are fairly similar to the NAEP at grades 4 and 8, although in grade 8 the NAEP devotes more time to Earth and space science than New Mexico does. Comparisons for high school could not be completed, because the grade 11 New Mexico Standards Based Assessment (NMSBA) will not be given until the 2007/08 school year. Overall, there is a reasonable match between the New Mexico test blueprints and the NAEP assessment and item specifications.

Standards and test specifications represent the starting point for the development of tests and test items. In the ideal alignment study state science assessments would be compared with NAEP assessments directly at the item level. At some future date the NAEP 2009 assessment items may be available for such a study.

---

**TABLE OF CONTENTS**

<b>Summary</b>	<b>iii</b>
<b>Background to the study</b>	<b>1</b>
<b>Content alignment at grade 4</b>	<b>3</b>
Areas of full alignment	3
Areas of partial alignment	4
Areas of nonalignment	5
Areas where New Mexico benchmarks go beyond the NAEP content statements	5
Summary of grade 4 alignment	5
<b>Content alignment at grade 8</b>	<b>6</b>
Areas of full alignment	6
Areas of partial alignment	7
Areas of nonalignment	7
Areas where New Mexico benchmarks go beyond the NAEP content statements	7
Summary of grade 8 alignment	8
<b>Content alignment at grade 12</b>	<b>8</b>
Areas of full alignment	9
Areas of partial alignment	9
Areas of nonalignment	9
Areas where New Mexico benchmarks go beyond the NAEP content statements	9
Summary of NAEP grade 12 alignment	10
<b>Test specifications alignment</b>	<b>10</b>
<b>Appendix A The documents compared</b>	<b>15</b>
<b>Appendix B How the study was conducted</b>	<b>18</b>
<b>Appendix C Content alignment table for grade 4</b>	<b>24</b>
<b>Appendix D Content alignment table for grade 8</b>	<b>36</b>
<b>Appendix E Content alignment for grade 12</b>	<b>48</b>
<b>References</b>	<b>65</b>
<b>Box 1 Methodology</b>	<b>3</b>
<b>Figures</b>	
<b>1</b>	The majority of New Mexico grade 4 standards partially address National Assessment of Educational Progress content statements 4
<b>2</b>	The majority of New Mexico grade 8 standards partially address National Assessment of Education Progress content statements 6
<b>3</b>	The majority of New Mexico grade 11 standards partially address National Assessment of Educational Progress content statements 8
<b>B1</b>	Crosswalk instrument 21

**Tables**

<b>1</b>	Average ratings of alignment of the New Mexico grade 4 science assessment framework and the National Assessment of Educational Progress grade 4 science content statements	4
<b>2</b>	Average ratings of alignment of the New Mexico grade 8 science assessment framework and the National Assessment of Educational Progress grade 8 science content statements	6
<b>3</b>	Average ratings of alignment of the New Mexico grade 11 science assessment framework and the National Assessment of Educational Progress grade 12 science content statements	8
<b>4</b>	Percentages of multiple-choice and open-response items and points on the New Mexico Standards Based Assessment	12
<b>5</b>	Proportions of different item types on the New Mexico science assessment	13
<b>6</b>	Approximate testing time allocated to different science topics on the New Mexico science assessment	13
<b>7</b>	Comparison of the proportions of testing time allocated to the NAEP science topics (percent)	14
<b>A1</b>	NAEP distribution of items and standards by content area and grade	15
<b>A2</b>	NMBSA distribution of items and points by strand and standard	16
<b>A3</b>	Number of assessment framework content standards by strand	16
<b>C1</b>	Alignment of National Assessment of Educational Progress grade 4 science and New Mexico grade 4 standards	24
<b>C2</b>	New Mexico grade 4 standards not covered by NAEP grade 4 content	35
<b>D1</b>	Alignment of National Assessment of Educational Progress grade 8 science and New Mexico grade 8 standards	36
<b>D2</b>	New Mexico grade 8 standards not covered by NAEP grade 8 content	47
<b>E1</b>	Alignment of National Assessment of Educational Progress grade 12 science and New Mexico science assessment framework grade 11 standards	48
<b>E2</b>	New Mexico grade 11 standards not covered by NAEP grade 12 content	64

**This policy research document is intended for policymakers to use when examining possible changes to the state assessment's alignment with the National Assessment of Educational Progress (NAEP).**

## **BACKGROUND TO THE STUDY**

This report presents the findings of an alignment study comparing the new science framework for the 2009 NAEP and the accompanying science assessment and item specifications with the New Mexico state science assessment. More details about the documents compared are in appendix A. The study was conducted for the Regional Education Laboratory Southwest, funded by the Institute of Education Sciences to provide research and support to Arkansas, Louisiana, New Mexico, Oklahoma, and Texas. The study was undertaken in anticipation of a growing need in the region to be better informed about how state assessment

standards in science compare with those tested in the NAEP.

The 2009 NAEP test is not yet in existence, so the purpose of this report is to give policymakers a headstart in determining where they might, if they so decide, begin to make changes in their assessment standards and specifications to develop an assessment system more closely aligned with that used for the NAEP.

Five factors make this study timely. First, the importance of state science assessments has been increased by the No Child Left Behind Act of 2001. Beginning in the 2007/08 school year, states are required to administer science assessments to all students in each of the elementary, middle, and high school levels, holding states and local school districts accountable for student academic achievement in science (NCLB, 2001).

Second, the NAEP is increasingly being used as a benchmark against which student achievement across the nation can be compared (Linn, 2005; Linn, Baker, & Herman, 2005). The NAEP has been dubbed the “nation’s report card,” and when fresh NAEP results are released—as they were for science in 2006, following an administration of the test in 2005—the media report the results (Cavanagh, 2006a, 2006b). Although states are not sanctioned for failing to demonstrate NAEP student performance improvement, NAEP data do provide an external accountability benchmark and serve to verify student achievement on state assessments. In fact, the National Center for Education Statistics has a website (<http://nces.ed.gov/nationsreportcard/nde/statecomp/>) that allows anyone to create customized comparative reports based on the latest NAEP data. So anyone can create tables that compare states and jurisdictions based on the average scale scores for selected groups of public school students within a single assessment year, or compare the change in performance between two assessment years.

Third, NAEP data are being used more in education research to investigate how the No Child Left

Behind Act provisions have played out in different states. For example, Olson (2005) compared the percentages of students at or above the proficient level on the 2005 state grade 8 mathematics assessments in 33 states. The study showed that, on average, 33 percent more students scored at or above the proficient level according to the state assessments than did so according to the NAEP. As yet, no similar study has been done of science, but with the release of the 2005 NAEP results it is now possible to do so.

Fourth, political attention is beginning to focus on using the NAEP as a yardstick for measuring state standards (Olson, 2007). In January 2007 two bills were introduced in Congress, one seeking to encourage states to benchmark their own standards and tests to the NAEP and the other calling for states to adopt voluntary “American education content standards” in mathematics and science that would be developed by the National Assessment Governing Board, the body responsible for the NAEP. These issues will doubtless be topics of debate in the upcoming reauthorization of the No Child Left Behind Act.

**Several factors are working to raise the National Assessment of Educational Progress to a de facto national benchmark, and states want to know how well their state standards align with it**

knowledge. It defines the science knowledge and skills that science-literate students should possess at grades 4, 8, and 12. The assessment itself, while retaining some familiar paper-and-pencil assessment formats, will also include student performance assessments in both classroom settings and computer simulations. The 2009 NAEP framework will determine the shape of NAEP science assessments through 2017, setting the direction of science assessment across the nation.

Fifth, the standards and test specifications that form the blueprint for the content the NAEP science assessment covers and the types of items it uses were revised in 2006. The 2009 NAEP framework takes account of the latest knowledge on science learning and assessment, which suggests that measuring student understanding involves much more than assessing factual

These factors are working together to gradually raise the status of the NAEP to a de facto national benchmark, and states naturally want to know how well their state standards align with the NAEP so they can make informed decisions about possible changes to their own standards and assessment systems. This report describes the results of a systematic alignment study conducted for that purpose. Details of the study are in appendix B.

The intent of this report is to inform those in the New Mexico Public Education Department responsible for shaping the state assessment in science how the current assessment standards and test specifications compare with those of the NAEP 2009 assessment.

Similar reports have been completed for Arkansas, Louisiana, Oklahoma, and Texas, but there is no intent to compare New Mexico with these other states. This report shows where there is good content alignment with NAEP standards, identifies where there is partial alignment, pinpoints NAEP standards where there are no corresponding state standards, and highlights where the New Mexico standards go beyond the NAEP. It also examines the assessment specifications, showing the percentages of NAEP assessment at each grade level devoted to different science topics and comparing that to the coverage of the topics in the New Mexico assessment. And it compares the proportions of types of items used to test students’ science knowledge and skills. Through comprehensive comparative analysis, the report provides a way for the New Mexico Public Education Department to gauge how well its tests are covering the depth of science understanding expected on the NAEP test.

The results are presented in the summary tables and narratives in the sections that follow. Those sections provide an analysis that highlights the differences found between the NAEP assessment and the New Mexico state assessment. For more detail about the alignment of the state framework to the individual content statements of the

NAEP, turn to the tables in appendixes C–E. They show exactly which New Mexico standards align with a particular NAEP statement and, in cases of partial alignment, explain why. For a discussion of methodology, see box 1 and appendix B.

fully addressed by New Mexico standards in the science assessment framework, and 28 of these content statements (85 percent) are partially addressed by the state. No NAEP content statements are unaddressed by New Mexico.

The average alignment rating for grade 4 is 2.2 (table 1). The majority of content statements were given ratings of 2, which means that state standards partially address the NAEP content statements (figure 1 and appendix C).

## CONTENT ALIGNMENT AT GRADE 4

The NAEP grade 4 science standards were compared with the New Mexico science assessment framework.

For grade 4 the NAEP provides 33 distinct content statements (displayed in parentheses in table 1). Five of these content statements (15 percent) are

### Areas of full alignment

Five NAEP grade 4 content statements are fully addressed by New Mexico's grade 4 assessment standards. Three of 15 physical science NAEP

#### BOX 1

#### Methodology

The chief research questions driving this study were these: “To what extent do current state assessment standards cover the content on which NAEP 2009 assessments will be based?” and “To what extent do current state assessment specifications align with the NAEP 2009 assessment specifications?”

The methodology used to answer these questions followed the successful pattern of a similar study conducted by WestEd in New England, which examined the alignment of math and reading standards with the NAEP. The methodology developed by WestEd for the New England study was designed to include all the most prominent alignment methodologies, discussed in appendix B. Thus far, alignment studies and methods have focused on aligning standards and tests, whereas the objective of this study was to compare one set of

assessment standards and specifications with another. The methodology in this study, however, is based upon methodologies for aligning standards with tests, because similar principles are used in both types of alignments.

In this study reviewers followed the methodology of the portion of the previous study examining alignment between two sets of standards. Following the methodology of Achieve, test blueprints were examined to find correspondence between the two documents (see appendix B). Reviewers performed gap analyses to identify content included in one set of standards but not the other, identified issues of order so they could reveal differences in the grade levels at which standards appear, and examined the degree to which the standards and assessments cover content to the same depth and have similar cognitive demands (depth-of-knowledge consistency) and the degree to which assessments cover the same range of content as the corresponding

standards (range-of-knowledge correspondence) to determine whether there was a match between New Mexico and the NAEP in the level of detail, cognitive demands, and range of content covered. A coding scheme was used to indicate alignment issues and reviewer ratings, and a matrix-like format was created to facilitate alignment.

Reviewers attended several training sessions, conducted individual reviews, and then met in teams of two to reach consensus on ratings. This consensus method was designed to create one consensus rating per NAEP standard with the help of a moderator and was not intended to allow for disagreements. This methodology was determined to be best suited to the scope and timing of this study. The consensus methodology is designed to highlight areas for states to examine, not to gather large amounts of data, record multiple ratings, or measure inter-rater reliability.

TABLE 1

**Average ratings of alignment of the New Mexico grade 4 science assessment framework and the National Assessment of Educational Progress grade 4 science content statements**

NAEP content area (number of NAEP standards)	Average rating
Overall physical science (15)	2.2
Matter (6)	2.3
Energy (5)	2.2
Motion (4)	2.0
Overall life science (7)	2.0
Structures and functions of living systems (4)	2.0
Changes in living systems (3)	2.0
Overall Earth and space science (11)	2.2
Earth and space in time (3)	2.3
Earth structures (3)	2.0
Earth systems (5)	2.2
All content (33)	2.2

Note: Rating is based on a scale of 1 to 3, where 1 indicates that state standards do not address NAEP content statement, 2 that they partially address NAEP content statement, and 3 that they fully address or exceed NAEP content statement by targeted grade level.

statements have full alignment with New Mexico, as do 2 of 11 Earth and space science statements.

The three NAEP grade 4 content statements fully addressed by New Mexico are P4.4—objects composed of single or multiple substances, P4.5—magnets repelling and attracting objects, and P4.7—forms of energy.

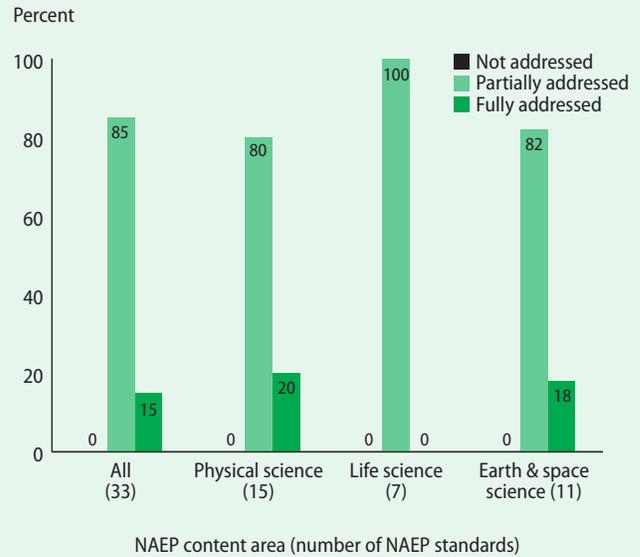
The two Earth and space science statements with full alignment are E4.3—changes in Earth’s surface and E4.11—humans’ dependence on and changes of their environments.

### Areas of partial alignment

Eighty-five percent of all NAEP grade 4 content statements have partial alignment, in large part because many New Mexico benchmarks address NAEP grade 4 content at higher grade levels. In addition, New Mexico’s content sometimes implies content explicitly stated by the NAEP.

FIGURE 1

**The majority of New Mexico grade 4 standards partially address National Assessment of Educational Progress content statements**



Raters found that many New Mexico benchmarks do not fully address the NAEP’s grade 4 content until grades 5 and 6. Nearly three-fourths of all NAEP statements were given the code “HG,” indicating that coverage of content was not found until a higher grade level. For example, P4.3 contains content regarding the three states of matter and their unique properties. Corresponding New Mexico statements (5 PS I.1 and 5 PS I.2) asking students to describe the way matter changes from one phase to another and the properties of the three states of matter are found in grade 5. In life science, L4.1 states, “Organisms need food, water, and air; a way to dispose of waste; and an environment in which they can live.” The New Mexico content statement that most closely corresponds to the NAEP content statement is 6 LS I.1, which mentions food, water, and air, but does not mention waste disposal. The content in 7 LS III.3 contains mention of excretion. However, both these standards are found in grade levels higher than grade 4.

One of the raters commented that, although many ratings of HG and 2 were given, New Mexico’s content was spiraled well through the curriculum

in grades 3–8 in all content areas. In other words, topics and concepts recurred across grades, and each time, new depth of knowledge was added. Additionally, the rater found that astronomy (stars and telescopes) is the focus of the Earth and space science section of the New Mexico framework, while the NAEP focuses on the sun and moon. Content involving the sun and moon is covered by New Mexico in grade 6.

Some of the content in NAEP was found to be implied in New Mexico content statements. For example, P4.15 states, “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.” New Mexico’s science framework contains content regarding gravity exerting more force on objects with greater mass and forces acting at a distance, but it does not mention objects falling to Earth, nor does it mention Earth as a reason for gravity. Additionally, the NAEP’s E4.8 states, “Weather changes from day to day and over the seasons.” A corresponding New Mexico statement states, “Know that local weather information describes patterns of change over a period of time,” but does not explicitly state that the change is over days or seasons.

---

### Areas of nonalignment

All NAEP content areas are at least partially covered by New Mexico’s content statements. No ratings of 1 were given to any content statements.

---

### Areas where New Mexico benchmarks go beyond the NAEP content statements

New Mexico has 41 content statements in the grade 4 science assessment framework, 23 (or 56 percent) of which are not addressed by the NAEP. The NAEP does not address, in its content statements, 8 of the 10 New Mexico statements in scientific thinking and practice, 4 of the 12 in physical science, 5 of the 10 in life science, 4 of the 6 in Earth and space science, or 2 of the 4 in science and society.

The NAEP does not address the majority of the scientific thinking and practice statements because it discusses scientific inquiry and practices in a section separate from the content statements, called “science practices,” intended to crosscut all NAEP content. One rater commented that New Mexico contains rigorous standards that promote strong scientific methodology.

In physical science the NAEP does not cover chemical and physical changes to matter (4 PS I.1), the particulate model of matter (4 PS I.2), the law of conservation of matter (4 PS I.3), or stored energy, such as potential energy (4 PS II.2).

In life science the NAEP does not cover structures and systems of organisms (4 LS I.1), senses and stimuli (4 LS I.2), cells (4 LS I.5), or parts and functions of the human body (4 LS III.1 and 4 LS III.2).

In Earth and space science the NAEP does not address telescopes and astronomy (4 ESS I.1, 4 ESS I.2, 4 ESS I.3) or U.S. weather patterns that move from west to east (4 ESS II.2).

In science and society the NAEP does not cover the various means of storage and retrieval of information (4 SS I.3) or that men and women of all races and social backgrounds choose science as a career (4 SS I.4).

---

### Summary of grade 4 alignment

All grade 4 NAEP content items are addressed to some degree by the New Mexico science assessment framework, and New Mexico is fairly well aligned with the NAEP, with an overall alignment rating of 2.15. The New Mexico statements are most often partially aligned with NAEP statements, primarily because New Mexico covers the

**All grade 4 NAEP content items are addressed to some degree by the New Mexico science assessment framework, and New Mexico is fairly well aligned with the NAEP**

TABLE 2

**Average ratings of alignment of the New Mexico grade 8 science assessment framework and the National Assessment of Educational Progress grade 8 science content statements**

NAEP content area (number of NAEP standards)	Average rating
Overall physical science (16)	2.1
Matter (7)	2.1
Energy (6)	2.0
Motion (3)	2.3
Overall life science (12)	1.7
Structures and functions of living systems (8)	1.9
Changes in living systems (4)	2.3
Overall Earth and space science (15)	2.1
Earth and space in time (4)	1.8
Earth structures (6)	2.0
Earth systems (5)	2.4
All content (43)	2.1

Note: Rating is based on a scale of 1 to 3, where 1 indicates that state standards do not address NAEP content statement, 2 that they partially address NAEP content statement, and 3 that they fully address or exceed NAEP content statement by targeted grade level.

NAEP content at higher grade levels. However, raters believed that New Mexico's content was spiraled well through all grade levels. The New Mexico framework also contains many content items that are not listed in the NAEP content statements, including those within New Mexico's scientific thinking and practice section, which reviewers praised for its promotion of rigorous scientific methodology.

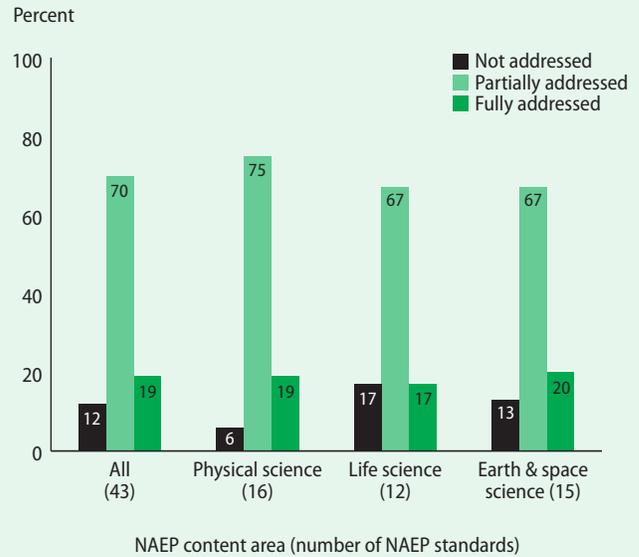
## CONTENT ALIGNMENT AT GRADE 8

The NAEP grade 8 science standards were compared with the New Mexico science assessment framework, primarily at grade 8.

For grade 8 the NAEP provides 43 distinct content statements (displayed in parentheses in table 2). Eight (19 percent) are fully addressed by New Mexico standards, 30 (70 percent) partially addressed, and 5 (12 percent) unaddressed.

FIGURE 2

**The majority of New Mexico grade 8 standards partially address National Assessment of Educational Progress content statements**



The majority of content statements were given ratings of 2, which means that most New Mexico performance standards are partially aligned with NAEP content statements (figure 2 and appendix D). The average alignment level for grade 8 is 2.1, indicating partial alignment.

### Areas of full alignment

Eight NAEP grade 8 content statements are fully addressed by New Mexico grade 8 assessment standards. Three of 16 physical science NAEP statements have full alignment with New Mexico, as do 2 of 12 life science statements and 3 of 15 Earth and space science statements.

The eight NAEP grade 8 content statements fully addressed by New Mexico's science assessment framework are P8.1—properties of solids, liquids, and gases, and the particulate model of matter, P8.7—chemical changes and conservation of mass, P8.16—forces and change in an object's motion, L8.9—reproduction, L8.11—traits, environmental change and extinction, E8.3—fossils as evidence of change, E8.12—seasons and their cause, and E8.14—the water cycle.

---

### Areas of partial alignment

Seventy percent of all NAEP grade 8 content statements have partial alignment with New Mexico's assessment standards.

Raters repeatedly found that New Mexico's assessment standards do not have as much detail as NAEP's statements; 87 percent of the 30 partially aligned NAEP statements were given a code for "more detail." For example, NAEP's P8.8 and P8.9 were matched to New Mexico's 8-PS-II.2, which mentions kinetic and potential energy. However, the NAEP statements include examples of kinetic and potential energy while New Mexico's statements do not. In life science the NAEP's L8.4 and L8.5 describe and give examples of producers, consumers, and decomposers, while New Mexico's matching standards (8-LS-III.2 and 8-LS-I.2) are more general and do not include examples. An example of this mismatch in detail can also be found in Earth and space science, where, for example, the NAEP's E8.10 provides content regarding Earth's magnetic field, which is detectable at the surface with a compass and similar to the field of a magnet, allowing many people to use compasses for navigation. The matching New Mexico statement says, "Know that Earth has a magnetic field" (8-PS-III.6), but gives no further details.

Several instances of implied content were also found for New Mexico content statements, as one rater found that statements in the state standards often emphasize different aspects of the same concept but do not explicitly match the NAEP statement. For example, P8.11 in the NAEP gives content regarding light energy from the sun reaching Earth, providing energy that heats Earth's surfaces and results in wind, ocean currents, and storms. New Mexico's 8-ESS-I.1 states, "Understand how energy from the sun and other stars, in the form of light, travels long distances to reach Earth," and 6-PS-II.4 provides content on energy traveling as waves and the sun as a source of energy for many Earth processes. New Mexico's standards appear to imply the parts of the NAEP standard that discuss the heating of Earth surfaces, winds, ocean currents, and storms.

---

### Areas of nonalignment

Five NAEP statements are found to be unaddressed by the New Mexico science assessment framework's content statements, one in physical science, two in life science, and two in Earth and space science.

In NAEP physical science the unaddressed content statement is P8.2—the arrangement of atoms and molecules that explain chemical properties. In life science the unaddressed statements are L8.2—cell division and differentiation, and L8.12—anatomical features and classifications of organisms. In Earth and space science the unaddressed statements are E8.1—a model of the solar system, and E8.4—earth processes and the measurement of geologic time.

**The majority of New Mexico grade 8 performance standards are partially aligned with NAEP content statements**

---

### Areas where New Mexico benchmarks go beyond the NAEP content statements

New Mexico has 51 performance standards listed in the science assessment framework for grade 8. The NAEP does not address about half of these standards, including the 8 New Mexico standards in scientific thinking and practice, 9 of the 24 in physical science, 4 of the 9 in life science, 2 of the 6 in Earth and space science, and all 4 in science and society.

The NAEP does not address scientific thinking and practice standards or science and society standards because it discusses inquiry and technology in a section separate from the content statements, called "science practices," intended to crosscut all NAEP content.

In physical science the NAEP does not address distinguishing between metals and nonmetals (8-PS-I.2), identifying and locating protons, neutrons, and electrons (8-PS-I.4), describing natural physical and chemical changes (8-PS-I.8), chemical reactions (8-PS-I.9, 8-PS-I.10), distinguishing

between renewable and nonrenewable sources of energy (8-PS-II.3), electrical energy (8-PS-II.4), or electrical fields and magnetic fields (8-PS-III.4, 8-PS-III.5).

In life science the NAEP does not address the composition of living organisms (8-LS-II.1), DNA and heredity (8-LS-II.2), the role of carbon in living systems (8-LS-II.3), or the influence of chemical substances on cellular activity (8-LS-III.3).

In Earth and space science the NAEP does not address the use of the properties of light for learning about the universe (8-ESS-I.2) or understanding geologic conditions that have resulted in energy resources in New Mexico (8-ESS-II.3).

**Summary of grade 8 alignment**

The majority of New Mexico grade 8 performance standards are partially aligned with NAEP content statements. Partial alignment was mainly due to the fact that NAEP standards most often contained more detail and more specific content than the corresponding standards in the New Mexico science assessment framework. The overall alignment rating is 2.1, indicating partial alignment.

**CONTENT ALIGNMENT AT GRADE 12**

The NAEP grade 12 science standards were compared with the New Mexico science assessment framework performance standards for grade 11, provided for this study by the New Mexico state science specialist.

For grade 12 the NAEP provides 49 distinct content statements (displayed in parentheses in table 3). Fifteen (31 percent) are fully addressed by New Mexico performance standards within the science assessment framework, 32 (65 percent) are partially addressed, and 2 (4 percent) are unaddressed.

The average alignment rating for all New Mexico grade 11 statements with NAEP content statements

TABLE 3

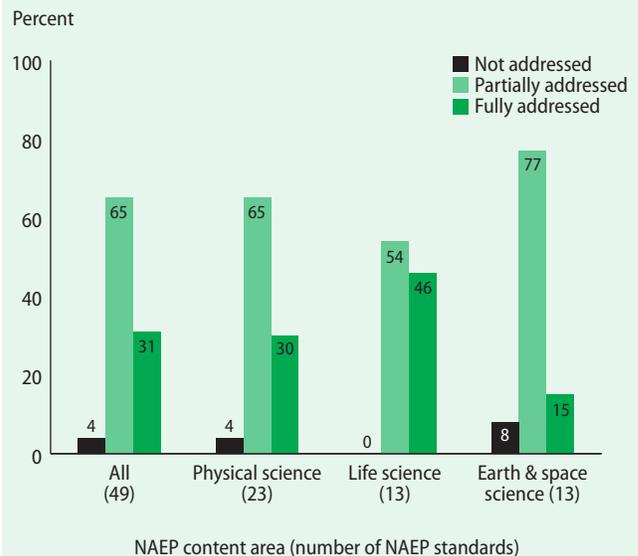
**Average ratings of alignment of the New Mexico grade 11 science assessment framework and the National Assessment of Educational Progress grade 12 science content statements**

NAEP content area (number of NAEP standards)	Average rating
Overall physical science (23)	2.3
Matter (7)	2.4
Energy (9)	2.0
Motion (7)	2.4
Overall life science (13)	2.5
Structures and functions of living systems (7)	2.3
Changes in living systems (6)	2.7
Overall Earth and space science (13)	2.1
Earth and space in time (7)	2.0
Earth structures (1)	2.0
Earth systems (5)	2.2
All content (49)	2.3

Note: Rating is based on a scale of 1 to 3, where 1 indicates that state standards do not address NAEP content statement, 2 that they partially address NAEP content statement, and 3 that they fully address or exceed NAEP content statement by targeted grade level.

FIGURE 3

**The majority of New Mexico grade 11 standards partially address National Assessment of Educational Progress content statements**



in grade 12 is 2.3. The majority of content statements were given ratings of 2, which means that state standards partially address the NAEP content statements (figure 3 and appendix E).

### Areas of full alignment

Fifteen NAEP grade 12 content statements are fully addressed by New Mexico's grade 11 assessment standards. Seven of 23 physical science NAEP statements have full alignment with New Mexico, as do 6 out of 13 life science statements and 2 of 13 Earth and space science statements.

In physical science the seven fully addressed standards are P12.3—organization of the periodic table, P12.6—an atom's electron configuration, P12.7—chemical reactions involving transferring electrons, transferring hydrogen ions, or sharing electrons, P12.17—motion of an object described by position, velocity, time, average speed and average acceleration, P12.19—net force changing the motion of an object, P12.20—acceleration, net force and mass, and P12.22—gravitational force.

In life science the six fully addressed standards are L12.3—regulation of cellular processes and the development of multicellular organisms, L12.7—stability and change in ecosystems and the impact of humans on other species, L12.8—genes, traits, chromosomes and heredity, L12.10—sorting and recombination of genes, L12.11—evolution's explanation of the history of life on Earth, and L12.13—evolution as the consequence of various factors.

In Earth and space science the two fully addressed standards are E12.4—relative and absolute dating and E12.9—Earth systems' internal and external sources of energy.

### Areas of partial alignment

Sixty-five percent of the NAEP grade 12 content statements have partial alignment.

Reviewers found that many New Mexico content statements address the NAEP content in an

implicit fashion; 67 percent of all NAEP content statements were given the implied content code "IC." For example, in physical science, although New Mexico has two performance standards that matched NAEP's P12.1 (11 PS I.1 and 11 PS I.10), the standards do not include (but seemed to imply) the NAEP content regarding forces of attraction between atoms, ions and molecules. In addition, in life science four New Mexico statements address NAEP L12.12, which includes content regarding the molecular evidence for the anatomical evidence for evolution, but the New Mexico statements do not specify "molecular evidence." In Earth and space science New Mexico's 11 ESS I.4 asks students to describe the key observations that led to the acceptance of the Big Bang theory, which seems to imply the content of NAEP E12.1, which contains details about the Big Bang theory itself.

### Areas of nonalignment

Only 2 of the 49 NAEP grade 12 content statements (4 percent) are unaddressed by New Mexico. One is P12.18, which states, "Objects undergo different kinds of motion—translational, rotational, and vibrational." The other unaddressed statement is E12.6, which states, "Early Earth was very different from today's planet. Evidence for one-celled forms of life—the bacteria—extends back more than 3.5 billion years. The evolution of life caused dramatic changes in the composition of Earth's atmosphere, which did not originally contain molecular oxygen." All other NAEP statements were partially or fully aligned with the New Mexico performance standards.

### Areas where New Mexico benchmarks go beyond the NAEP content statements

New Mexico has 121 performance standards in the grade 11 science assessment framework. NAEP does not address 55 of these, including the 16 New Mexico standards in scientific thinking and practice, 13 of the 38 in physical science, 3 of the

**New Mexico's grade 11 performance standards are fairly well aligned with NAEP's grade 12 content**

29 in life science, 4 of the 19 in Earth and space science, and the 19 in science and society.

The NAEP does not address the scientific thinking and practice (strand I) statements because it discusses inquiry in a section separate from the content statements, called “science practices,” intended to cross-cut all NAEP content.

In physical science the NAEP does not address 11 PS I.2—identifying, measuring and using a variety of physical and chemical properties, 11 PS I.3—separating mixtures into pure substances, 11 PS I.9—the arrangement of atoms and bonds determining macroscopic properties, 11 PS II.4—heat transfer, 11 PS II.6—the decrease of energy’s ability to do work after conversion of forms, 11 PS II.10—wavelengths of electromagnetic radiation, 11 PS II.11—the concept of equilibrium, 11 PS III.4—relationship between force and pressure, and between the pressure of gas and the temperature and amount of gas, 11 PS III.6—representing magnitude and direction of forces by vectors, 11 PS III.9—relative motion and frames of reference, 11 PS III.10—wave propagation using amplitude, wavelength, frequency, and speed, 11 PS III.11—interference, reflection and refraction of waves, or 11 PS III.12—waves used for practical purposes.

**To improve the assessment of the range of science knowledge and skills, the last two NAEP science frameworks have expanded the range of item types on the test**

In life science the NAEP does not address 11 LS I.8—understanding and explaining the hierarchical classification scheme, 11 LS II.7—chromosomes, including sex chromosomes, within cells in the human body, or 11 LS III.4—cell membrane permeability and transport.

In Earth and space science the NAEP does not address 11 ESS I.2—changes in positions and appearances of objects in the sky, 11 ESS I.3—knowledge of the universe stemming from evidence collected from advanced technology, 11 ESS I.5—objects’ emission of electromagnetic radiation, or 11 ESS I.7—the role of New Mexico research facilities in space exploration.

The NAEP does not address any of New Mexico’s science and society performance standards. This is because the NAEP addresses the societal applications of science and technology in its “science practices” section, which is separate from its content statements but intended to be applied to all content.

---

### Summary of NAEP grade 12 alignment

New Mexico’s grade 11 performance standards are fairly well aligned with NAEP’s grade 12 content. Only two NAEP statements were found to be unaddressed by New Mexico. Reviewers found most content to be partially aligned, and the majority of aligned New Mexico content implied content explicitly stated in the NAEP. The overall alignment rating is 2.3.

Reviewers noted that many New Mexico performance standards were not covered by the NAEP. They also noted that the New Mexico science framework is well organized; one reviewer commented, “The topics fit well together.” One reviewer found similarities in wording between New Mexico and the NAEP, suggesting that New Mexico’s grade 11 science assessment framework is well-aligned with the content of the NAEP.

---

### TEST SPECIFICATIONS ALIGNMENT

The assessment specifications alignment involved two parts: examining the types of items found in the NAEP and in the New Mexico Standards Based Assessment (NMSBA), and comparing the NAEP’s and the NMSBA’s distribution of items between the different science strands. The test specifications alignment for the NMSBA is completed only for grades 3–9, because the NMSBA grade 11 test will not be given until the 2007/08 school year and the high school NMSBA test specifications were not readily available.

Science is a discipline with a strong tradition of investigation, experimentation, and application of knowledge and skills. Before the 2005 assessment,

NAEP Science Assessments consisted primarily of short-answer, paper-and-pencil questions that were mostly multiple-choice, which can only go so far in assessing skills. To improve the assessment of the range of science knowledge and skills, the last two NAEP science frameworks have expanded the range of item types on the test. In particular, the 2009 NAEP framework takes advantage of advances in educational measurement and the development of computer-based assessments. Due to the varying ways that differing item types assess and reveal what students know and can do, the NAEP 2009 assessment specifications require future NAEP tests to incorporate a range of item types, allowing students to reveal their understanding in ways beyond traditional selected-response methods. Multiple-choice items, short constructed-response items, extended constructed-response items, hands-on performance tasks, and interactive computer tasks will all be used to more accurately assess student knowledge, thinking, and skills.

Each type of assessment item demands a unique response from students (selecting a response from a set of alternatives, writing an explanation or justification, performing a virtual lab experiment). Individual items may draw on different types of stimuli (verbal, graphic, manipulative) to access the knowledge and skills required and may be scored in a variety of ways (right/wrong, partial credit, human scorers, computer software). By using several types of items the 2009 NAEP science assessment will require students to draw on multiple types of knowledge and a variety of skills for using and expressing that knowledge, thereby giving a more accurate picture of the breadth and depth of their learning. In this study, the following item types from NAEP were compared with the types in use from the states.

In multiple-choice items, students reflect on the material and then select an answer from a limited number of alternatives. Well constructed multiple-choice items can probe important facts, broad concepts, and themes of science, as well as deductive reasoning skills.

Constructed-response items, in which students answer without reference to a provided list of alternatives, include short constructed-response items and extended constructed-response items. Constructed-response items can provide insights into students' levels of conceptual understanding and assess their abilities to communicate about science. They can also be used to probe student abilities to generate information related to science content statements and their interconnections (how two or more cyclic events are related). Constructed-response items may be particularly useful for probing the practices of using scientific inquiry or using technological design (interpret given data or provide a solution to a real-world problem).

In hands-on performance tasks, students manipulate selected physical objects and try to solve a scientific problem involving the objects. These exercises, if carefully designed, can probe student abilities to combine science knowledge with the investigative skills reflective of the nature of science and inquiry.

Interactive computer tasks, in the 2009 NAEP science assessment, may involve information search and analysis, empirical investigation, simulation, or concept mapping. The broad purpose of interactive computer tasks in this context is to tap performance expectations that are more advantageously assessed in a virtual format, such as scientific modeling of microscopic or temporal phenomena, repeated experiments, or simulations of hazardous or messy lab situations. Interactive computer tasks are intended as a complement to the hands-on performance tasks, not as a replacement.

The NAEP specifications also include two other types of items, item clusters and predict-observe-explain item sets. Item clusters are groups of

**The 2009 NAEP science assessment will require students to draw on multiple types of knowledge and a variety of skills for using that knowledge, thereby giving a more accurate picture of the breadth and depth of their learning**

related items that provide more in-depth analysis of student performance than would a collection of discrete, unrelated items. They can be particularly useful in exploring student conceptions, predictions, or explanations of the natural world. The predict-observe-explain item sets (White & Gunstone, 1992) describe a situation and ask the student to predict, observe, and/or explain the outcome, sometimes with additional supporting detail. Predict-observe-explain items may involve using science principles or the cognitive demand of “knowing why (schematic knowledge).” Because these are really ways of clustering items and are not usually included in state test specifications, they were not used for comparison in this study.

The NAEP stipulates that 50 percent of student response time should be spent on multiple-choice items and the other 50 percent on constructed-response items (including short constructed-response, extended constructed-response, and concept-mapping tasks). Within these two categories are item clusters, predict-observe-explain item sets, hands-on performance tasks, and interactive computer tasks. There will be at least one item cluster, one predict-observe-explain item set, one hands-on performance task, and one interactive computer task at each grade level, and the total number of interactive computer tasks plus hands-on performance tasks will be at least four at each grade level.

The current New Mexico tests contain a combination of multiple-choice and open-response items.

Open-response items include both short answer items and longer, open-ended items. Table 4 shows the percentages of multiple-choice items and open-response items on the NMSBA, as well as the percentage of multiple-choice and open response points on the NMSBA. The percentages of different item types used to test students in science in New Mexico stay the same in grades 3 through 9, while the weighting of the points gradually increases for open-response items as the grade level increases.

Table 5 shows the percentages of various item types found in the NAEP and in New Mexico. The 2009 NAEP will have 50 percent of student response time allocated to multiple-choice items and 50 percent of student response time allocated to constructed-response items (short and extended). The NMSBA also contains a combination of multiple-choice and “open response” items, which are comparable to the NAEP’s “constructed-response” items. New Mexico provides item type distribution data by item and by points. Table 4 indicates the tests’ percentages of multiple-choice and open-response, or constructed-response, items, and the percentages of multiple-choice points and open-response points. New Mexico’s grade 8 percentages of item types by number of points are distributed similarly to the NAEP’s.

To consider how the state test coverage of the NAEP science topics matched, table 6 shows the proportions of testing time devoted to each of the three content areas for the NAEP and the

TABLE 4

**Percentages of multiple-choice and open-response items and points on the New Mexico Standards Based Assessment (percent)**

Grade	Share of multiple-choice items	Share of open-response items	Share of multiple-choice points	Share of open-response points
3	79	21	61	39
4	79	21	61	39
5	79	21	59	41
6	79	21	59	41
7	79	21	59	41
8	79	21	58	42
9	79	21	58	42

New Mexico test. The first column of the table lists all the science topic areas that are included on the New Mexico test. The first three topic areas (physical, life, and Earth and space science) are those that are covered in the NAEP, and the two topics below those (science as inquiry and science and society) are not separately assessed as content strands on the NAEP test.

Under the column heading for grade 4, three subcolumns are shown. The first shows the proportion of testing time devoted to each topic for the three NAEP topic areas. The second shows the proportion of points devoted to each of the five New Mexico topics at grade 4. The third shows the comparison of percentages devoted to the three NAEP topics, a positive number if the New Mexico test devotes more and a negative number if the NAEP devotes more. This pattern of columns is repeated for grade 8. It is noted that percentage of test time (for NAEP) and percentage of points

(for New Mexico) may not be directly comparable; New Mexico's test blueprints did not include the proportions of student testing time devoted to each science strand, so the percentage of points per strand is used in this comparison.

At grades 4 and 8 the proportion of New Mexico points devoted to physical, life and Earth and space sciences is lower than the proportion of time given to those science strands in the NAEP. This is because New Mexico allocates 29 percent of its points at grade 4 and 35 percent of its points at grade 8 to science as inquiry and science and society topics. The NAEP does not separately address these strands in its organization of science content statements.

Table 7 ignores the testing time devoted to the two topics that are not separately tested in NAEP (science as inquiry and science and society) and shows how the proportion of testing time in NAEP

TABLE 5

**Proportions of different item types on the New Mexico science assessment (percent)**

NAEP item types	NAEP		New Mexico		
	All grades	Grade 4		Grade 8	
	Share of time	Share of total items	Share of total points	Share of total items	Share of total points
Multiple-choice items	50	79	61	79	58
Constructed-response items (short and extended)	50	21	39	21	42
Hands-on performance tasks <sup>a</sup>	(≥1)				
Interactive computer tasks <sup>a</sup>	(≥1)				

a. Hands-on performance tasks and interactive computer tasks are combination items and can be categorized as multiple-choice or constructed-response.

TABLE 6

**Approximate testing time allocated to different science topics on the New Mexico science assessment (percent)**

Content area	Grade 4			Grade 8		
	NAEP (time)	New Mexico (points)	Difference	NAEP (time)	New Mexico (points)	Difference
Physical science	33.3	29.0	-4.3	30.0	21.0	-9.0
Life science	33.3	26.0	-7.3	30.0	27.0	-3.0
Earth and space science	33.3	16.0	-17.3	40.0	17.0	-23.0
Science as inquiry	0.0	23.0		0.0	27.0	
Science and society	0.0	6.0		0.0	8.0	

TABLE 7

**Comparison of the proportions of testing time allocated to the NAEP science topics (percent)**

	Grade 4			Grade 8		
	NAEP (time)	New Mexico (points)	Difference	NAEP (time)	New Mexico (points)	Difference
Physical science	33.3	40.9	+7.6	30.0	32.6	+2.6
Life science	33.3	36.4	+3.1	30.0	41.9	+11.9
Earth and space science	33.3	22.7	-10.6	40.0	25.6	-14.4

compares with the proportion of points in the New Mexico state test. At grade 4, in Earth and space science, there is a difference of approximately 10.6 percentage points between New Mexico and the NAEP. At grade 8, in Earth and space science, there is a difference of approximately 14.4 percentage points between New Mexico and the NAEP.

In grades 4 and 8 there is a greater proportion of points devoted to physical and life sciences in New Mexico than the proportion of time given to physical and life sciences in the NAEP, and there is a smaller percentage of points given to Earth and space science by New Mexico than the percentage

of time given to Earth and space science by the NAEP. In the NAEP the percentage of time devoted to physical science decreases from grade 4 to grade 8, while in New Mexico the percentage of points devoted to physical science decreases from grade 4 to grade 8. The NAEP's percentage of time devoted to life science decreases from grade 4 to grade 8, while New Mexico's percentage of points devoted to life science greatly increases from grade 4 to grade 8. For Earth and space science the percentages of time and points in the NAEP and New Mexico increase from grade 4 to grade 8, although the NAEP's increase is substantially larger than New Mexico's.