Comparison of the PISA 2009 and NAEP 2009 Reading Assessments

Prepared by the International Activities Program,
National Center for Education Statistics

December 2010

Background

In the United States, nationally representative data on student achievement come primarily from two sources: the National Assessment of Educational Progress (NAEP)—also known as "The Nation's Report Card"—and U.S. participation in international assessments, including the Program for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS). While the international assessments may appear to have significant similarities with NAEP, each was designed to serve a specific purpose and each is based on a separate and unique framework and set of assessment items. Thus, each gives a somewhat different view of U.S. student performance.

In December 2010, the National Center for Education Statistics (NCES) is releasing results from the 2009 administration of PISA, an assessment of 15-year-olds' reading, mathematics, and science literacy (Fleischman et al. 2010). In PISA 2009, reading literacy was the major domain; therefore, detailed information about U.S. student performance in reading will be available. The PISA release follows the March 2010 release of the NAEP 2009 4th- and 8th-grade reading assessment results and the November 2010 release of the NAEP 2009 12th-grade reading assessment results (NCES 2009; NCES 2010). In anticipation of questions about the similarities and differences between the PISA and NAEP reading assessments, and what each can tell us about U.S. students' reading skills, NCES prepared this paper, which discusses aspects of each assessment. In particular, the paper discusses the purposes, target populations and reporting levels, and content assessed by each. For NAEP, the paper focuses on the reading assessments for students in grades 8 and 12, as these are the two target populations closest to PISA's target population of 15-year-olds. To examine and compare the content measured by PISA and NAEP, NCES convened a panel of reading experts[1] to compare how PISA and NAEP define reading, aspects of the texts used as the basis of the assessments, and the reading processes required in each assessment. This information is intended to help the press and others understand the similarities and differences between the assessments and to help identify what PISA and NAEP each contribute to the overall knowledge base on U.S. student reading performance.

Purposes of PISA and NAEP

The goals of the two assessments have subtle but important distinctions with regard to U.S. curricula.

International assessments, such as PISA, support comparisons of student performance among countries but provide less information for within-U.S. comparisons. NAEP supports comparisons of student performance among states, public and private schools, student demographic groups, and

[1] The experts included Patricia Alexander (University of Maryland), Gina Biancarosa (Stanford University and University of Oregon), Michael Kamil (Stanford University), Pamela Mason (Harvard University), and Junko Yokota (National Louis University). The meeting was facilitated by staff from the NCES International Activities Program in a joint effort with staff from the NCES Assessment Division.

a set of urban public school districts.[2] Both PISA and NAEP are conducted regularly to allow the monitoring of student outcomes over time. PISA is conducted every three years and NAEP is conducted, for some subjects (including reading), every two years; for some subjects, every four years; and for other subjects, less frequently.[3]

PISA provides internationally comparative information in the United States on the reading, mathematics, and science literacy of students at an age that, for most countries, is near the end of compulsory schooling. The objective of PISA is to measure the "yield" of education systems, or what skills and competencies students have acquired and can apply in these subjects to real-world contexts as they near the transition from compulsory schooling. PISA's *literacy* concept, which applies to the reading, mathematics, and science assessments, emphasizes the mastery of processes, understanding of concepts, and application of knowledge and functioning in various situations. By focusing on literacy, PISA assesses what students have learned in and outside of the school environments.

NAEP reports information on achievement in reading and other subjects at the 4th-, 8th-, and 12th-grade levels across the country. NAEP assessments are based on assessment frameworks and achievement levels (i.e., *Basic*, *Proficient*, and *Advanced*) established by the National Assessment Governing Board. NAEP assessment frameworks and achievement levels are based on the collaborative input of a wide range of experts and participants from the government, education, business, and public sectors in the United States. The frameworks are intended to be a blueprint for the assessment (specifying what should be assessed), and the achievement levels act as performance standards for each subject area and grade, showing what students should know and be able to do (NCES 2009, p. 4–5).

The focus of NAEP on subject matter expectations in the United States distinguishes it from PISA, the content of which is determined in collaboration with other countries. The focus in PISA on the yield of education systems and the application of competencies in real-world contexts distinguishes it from NAEP, which measures school-based performance and abilities to read and understand written texts and to interpret and use what students have read in ways that are appropriate to the type of text and situation.

Target Populations Assessed by PISA and NAEP

The students assessed represent different target populations.

PISA and NAEP are both sample-based assessments, meaning that each assessment administered is to a sample of students (rather than to all students) and the results are generalized to the larger population. However, each assessment defines the population to which it is generalizing, and thus from which the sample is drawn, differently. One distinction between main NAEP and PISA is that NAEP uses grade-based samples, whereas PISA uses an age-based sample. These choices relate to the purpose of each program—NAEP, to report on student achievement based on what students

[2] In "main NAEP," students in grades 4, 8, and 12 are assessed; in "long-term trend NAEP," students ages 9, 13, and 17 are assessed. This paper focuses on main NAEP (on which the 2009 NAEP results are based); thus, all statements about NAEP in this paper refer to main NAEP.

[3] PISA is on a 3-year cycle, with one domain being featured as the major domain every 9 years. In 2000 and 2009, reading literacy was the major domain in PISA. Main NAEP currently assesses 4th- and 8th-grade reading and mathematics every 2 years. Every 4 years, NAEP assesses 12th-grade reading and mathematics, as well as 4th-, 8th, and 12th grade science. Other subjects, such as writing, civics, economics, and U.S. history, are also assessed, but less frequently. In 2009, there were 18 public school districts that participated in the Trial Urban District Assessment (TUDA) reading, mathematics, and science assessments.

learn by a specific grade in school; and PISA, to describe the yield of education systems toward the end of compulsory schooling.

The PISA target population is all 15-year-old students. In 2009, this included all students who were 15 years and 3 months to 16 years and 2 months at the beginning of the testing period (fall 2009) and who were enrolled in school, regardless of grade level or full- or part-time status. The majority of respondents in the U.S. PISA 2009 sample were in 10th grade (68.5 percent), but some were in the 11th (20.3 percent), 9th (10.9 percent), or another grade (0.3 percent). The NAEP target populations are all students in the 4th, 8th, and 12th grades, and NAEP reflects the performance of U.S. students enrolled in these grades. Thus, the PISA results are for students who are mostly in grades between those being tested for NAEP (the 8th and 12th grades), and closer in grade proximity to those taking the NAEP 8th-grade assessment because of the timing of the respective assessments (with PISA given earlier in the school year than NAEP).

Reporting Levels and Sample Sizes for PISA and NAEP

PISA and NAEP are designed to provide results at different levels of aggregation and, as a result, have different levels of precision.

PISA and NAEP are both designed to provide information about U.S. students' performance aggregated to the national level and for subgroups of the population (e.g., subgroups defined by gender and race/ethnicity). NAEP, however, is also designed to provide reading results for individual states and some large urban districts.[4] Under the federal Elementary and Secondary Education Act of 2002, states receiving Title I funds are required to participate in NAEP reading, as well as mathematics, at the 4th- and 8th-grade levels. No requirements are placed on states to participate in PISA or other international assessments; and while states or districts could opt to participate in PISA and receive state- or district-level results, to date none have.

The reporting requirements for NAEP and PISA have implications for sample sizes and, in turn, for which subgroups' results can be reported and for the precision of the estimates. The NAEP national sample comprises the state and district samples and thus is extremely large. For example, the sample size for the grade 8 NAEP 2009 reading assessment was more than 160,000 students. In contrast, in 2009, PISA assessed 5,233 students. Because of its large sample size, NAEP is able to reliably measure achievement for more subgroups than PISA can. For example, while both NAEP and PISA can report achievement for students based on their racial and ethnic classifications, in 2009 the PISA sample did not include a sufficiently large sample size to report achievement for students who are identified as American Indian/Alaska Native or Native Hawaiian/Other Pacific Islander. Moreover, because of NAEP's large sample size, it can detect smaller differences between subgroups or over time than can PISA. For example, while the standard error associated with the NAEP 2009 reading national mean score for grade 8 was 0.3 (NCES 2009), the standard error for the PISA 2009 reading mean score was 3.7 for the United States (Fleischman et al. 2010).

Related to sampling is the degree of inclusion of students with special or language needs. Both NAEP and PISA strive to be inclusive and ultimately achieve similar inclusion rates, although their specific policies differ. PISA is designed to be as inclusive as possible and requires that no more than 5 percent of the target population be excluded from testing. Exclusions are allowed at both the school level (e.g., a geographically remote school) and within schools at the student level, including students with a functional disability, intellectual disability, or insufficient language experience

[4] State- and district-level results are also provided for some other subjects.

(defined as non-native, limited proficient speakers with less than 1 year of instruction in the testing language). Currently, there are no special accommodations provided for students taking PISA. NAEP's policy endeavors to assess all students selected as part of its sampling process and allows a range of accommodations, as necessary, for students with disabilities (SD) or English language learners (ELL). Accommodations include modifications in presentation format, response formats, test-taking setting, timing, or other aspects, as well as direct and indirect linguistic support. The weighted exclusion rate was 5 percent of students in PISA in 2009. The exclusion rate in the NAEP 8th-grade assessment in 2009 was 4 percent of students.

What Is Measured by the PISA and NAEP Reading Assessments?

PISA and NAEP measure some similar aspects of reading, but there are differences in how reading is defined in the frameworks, in the types of passages presented to students, and in the types of cognitive skills required of students.

To examine the similarities and differences in the content assessed by PISA and NAEP, NCES commissioned a panel of experts to examine and compare the PISA and NAEP reading frameworks, passages, and items. The panel looked at how each assessment defined reading; how the domain was organized in the frameworks; the nature, length, and difficulty of the reading passages; the format of the items used; and the cognitive processes in which students were asked to engage.

Definitions of Reading

There is overlap between the NAEP 2009 and PISA 2009 definitions of reading (see exhibit 1). The definitions for both assessments identify reading as a constructive process that involves interaction between the reader and the text, and both focus on understanding and using written text. There are subtle differences, however. PISA's definition emphasizes the use of reading for personally defined goals and growth and for participation in society, while the NAEP definition reflects the notion that readers draw on the ideas and information they have acquired from text to meet a particular purpose or situational need. The NAEP reading framework calls for the 12th grade assessment to address the preparedness of 12th-graders for postsecondary education and training, although NAEP does not currently report results on a preparedness scale.

Exhibit 1. Definitions of reading

Reading Literacy in PISA 2009
Reading literacy is understanding, using, reflecting on, and engaging with written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society (OECD 2009, p. 23).

Reading in NAEP 2009
Reading is an active and complex process that involves understanding written text; developing and interpreting meaning; and using meaning as appropriate to type of text, purpose, and situation (National Assessment Governing Board 2008, p. 2).

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, 2009; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress(NAEP), *Reading Framework for the 2009 National Assessment of Educational* Progress, 2008.

Framework Organization and Assessment Features

There are similarities in how PISA and NAEP organize the reading domain, although both PISA and NAEP have unique features and requirements not included in the other assessment.

As shown in exhibit 2, the 2009 NAEP reading framework is based on a two-dimensional matrix with text types as one dimension and cognitive processes ("cognitive targets") as the other. These can be thought of as "what students read" and "what students do with what they read." PISA's framework also organizes the assessment around texts and cognitive processes ("aspects"). PISA's text dimension, however, includes four text taxonomies—type, format, medium, and environment. Type is largely analogous to NAEP's "text type." Format distinguishes between "continuous" texts—that is, text that is formed by sentences organized into paragraphs—and noncontinuous texts—that is, texts that are composed of less than sentences (e.g., lists, tables, graphs, diagrams, advertisements, schedules, catalogues, indexes and forms) (OECD 2009). Medium and environment are used because the PISA 2009 reading assessment included an assessment of student reading literacy with electronic texts; medium distinguishes between print and electronic text[5], and environment distinguishes between authored and message-based text.[6] NAEP does not currently include electronic texts in its reading framework or assessment.

PISA also uses a unique third dimension, the reading situation, which distinguishes the range of contexts or purpose for which reading takes place. NAEP, on the other hand, has its own unique feature: an assessment of "meaning vocabulary," which refers to students' ability to apply meaning to words vital for comprehending the overall passage. The framework specifies that each NAEP passage will have approximately two items that focus on meaning vocabulary.

Reading Passages

The reading passages selected for inclusion in PISA and NAEP represent the individual framework and design of each assessment. Both assessments strive to cover a wide range of text types, difficulty, and topics.

[5] The PISA 2009 electronic reading assessment was an international option; the United States did not participate.
[6] The distinction between *authored* and *message-based* text is whether the reader has the potential to influence the content of the site. An *authored* environment does not allow the reader to modify the content (e.g., home pages, government information sites, news sites, etc.) whereas a *message-based* environment allows the reader the opportunity to add to or change the content (e.g., e-mail, blogs, chat rooms, etc.) (OECD 2009, p.29).

Both assessments distinguish a range of text types, which are somewhat, but not perfectly, aligned. For example, NAEP's literary category includes fiction, poetry, and literary nonfiction and is more expansive than PISA's corresponding category, narration. Also, there is no clear counterpart to PISA's "description" category, which includes documents that typically provide an answer to "what?" questions, such as a depiction of a place or a schedule. NAEP texts in the informational category include exposition, argument and persuasion, and procedural texts and documents (e.g., news articles, research reports, historical documents, persuasive essays, and position papers). Across grade levels, NAEP incorporates increasingly complex text structures and features, genre/type of text, and author's craft. Although NAEP includes some noncontinuous material, it is only used as augmentation, embedded in continuous material, at the 8th-grade level, and there are only a few stand-alone examples in the 12th-grade assessment. PISA, on the other hand, makes heavier use of noncontinuous material, including, in particular, texts that fall into PISA's exposition, argumentation, and transaction categories. Even the continuous texts in those categories are often drawn from activities that a 15-year old student might engage in during a daily routine and may come from a wide range of sources that are not strictly academically grounded (e.g., from a mainstream newspaper versus from a student's educational magazine).

PISA is designed to cover a wide breadth of what students read and the purpose for reading, which is not always in school but outside of school as well. PISA includes both continuous and noncontinuous text, as well as the range of types described in table 1. The most common text type in PISA is exposition, which includes almost one-third of the passages. The least common text type in PISA is narration, which is represented in about 10 percent of the passages.

The framework for NAEP also addresses the different kinds of reading materials students will encounter both in and outside of the classroom and describes NAEP as an "assessment of varied reading skills." The broad categories of literary and informational texts are identified in two ways: first, by the different purposes for which literary and informational texts are read, and, second, by structural differences between literary and informational texts that mark the text and help readers understand what they are reading. Passages in the 8th-grade assessment are evenly distributed between these two main categories. At the 12th-grade level, about one-quarter of the passages are literary and about three-quarters are informational. This is intended to mirror the distribution of the kinds of texts students encounter as they progress through the education system.

The PISA reading assessment includes 29 passages, and the NAEP 8th- and 12th-grade assessments include 16 and 17 passages, respectively. In PISA and NAEP, each student receives only a subset of the passages in each assessment. In NAEP, both the 8th- and 12th-grade assessments include some passages that are used at another grade as well—e.g., there is a subset of 8th-grade passages that is also used in the 4th-grade assessment and a subset that is used at 12th grade. Also, NAEP pairs some passages. Students are presented two related passages and are asked inter-textual questions, as well as questions specific to each passage. PISA does not have paired passages in the same sense, although some PISA passages include multiple parts and may even include different text types or formats, and inter-textual questions may be included. In PISA, students respond to 2 to 5 items per passage, whereas in NAEP students respond to between 9 and 11 items per passage or pair of passages.

Exhibit 2. PISA and NAEP reading framework dimensions and features

Dimensions	PISA	NAEP
Texts	Type Narration Exposition Argumentation Instruction Transaction Description Format Continuous, noncontinuous Medium Print, electronic Environment Authored, message-based	Text types Literary: fiction literary nonfiction poetry Informational: exposition argumentation/persuasive text procedural texts and documents
Cognitive processes	Aspects Access and retrieve Integrate and interpret Reflect and evaluate	Cognitive targets Locate and recall Integrate and interpret Critique and evaluate
Situation	Contexts Personal, Public, Occupational, Educational	Category does not exist
Other features		Meaning vocabulary

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, 2009; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress(NAEP), *Reading Framework for the 2009 National Assessment of Educational Progress*, 2008.

Related to the issue of text types, there is another important but subtle distinction between NAEP and PISA. The NAEP framework explicitly emphasizes the authenticity of text and notes a commitment to selecting high-quality, authentic stimulus materials that students are likely to encounter both in school and out of school. To this end, NAEP sets minimum passage lengths for inclusion and makes very few edits to the original texts. Although PISA is intended to measure authentic tasks, the PISA framework does not explicitly emphasize the use of existing, intact text. It is constrained in some ways by its international nature, as passages must be applicable across a wide range of cultures and languages. Therefore, while passages are selected to represent a range of texts and applicability in real-world settings, more manipulation and editing of passages is used than is in NAEP. Also, the 2009 NAEP reading framework explicitly required that the selection of passages be informed by readability analyses, such as the ones described in the next section; although readability

analyses had played a role in passage selection prior to 2009, the 2009 framework explicitly called for their use.

Length

PISA passages are notably shorter than NAEP 8th- and 12th-grade passages, averaging 354 words to NAEP's 924 and 1,174 words per passage or pair of passages,[7] respectively (see table 1). Passages or pairs of passages in NAEP range from 219 words to 1,429 words in the 8th grade and 771 to 1,429 words in the 12th grade, compared with a range of 53 to 758 words in PISA. Thus, even the longest passage in PISA is shorter than the average passage length for each of the two NAEP grades. The NAEP framework specifies passage length by grade level to represent what students encounter in their in-school and out-of-school reading, to ensure usage of strategic reading skills, to ensure that approximately 10 distinct items can be generated from the passage, and to ensure that the structural patterns of the passages are supportive of the range of text types and that the items cover the range of cognitive processes (National Assessment Governing Board 2008, p. 28). PISA does not have a similar requirement for passage length as part of its framework. While PISA's noncontinuous texts tend to be shorter than its continuous texts (see appendix table A), the presence of noncontinuous texts alone does not account for these differences. Rather, it is likely that these differences are driven by the differing framework requirements, constraints, and purposes of the assessments described earlier.

Table 1. NAEP 2009 and PISA 2009 reading passage word and item analyses

Length of passage and items per passage	NAEP grade 8[1] (*N* = 13)	PISA (*N* = 29)	NAEP grade 12[2] (*N* = 13)
Average number of words	923.6	354.4	1173.5
Range of words in passages	219 - 1,429	53 - 758	771 - 1,429
Average number of items	10.0	3.6	10.1
Range of number of items	9 - 11	2 - 5	9 - 11

[1]This column presents averages and ranges for all reading passages categorized as grades 4/8, 8, and 8/12.

[2]This column presents averages and ranges for all reading passages categorized as grades 8/12 and 12.

NOTE: In NAEP a student might be presented with a single passage or a pair of passages; length analyses were calculated based on the passage or pair of passages a student received.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress(NAEP) 2009 Reading Assessment; and Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2009 Reading Literacy Assessment.

Passage Difficulty

Several readability measures (i.e., the Dale-Chall Formula, Flesch Reading Ease, Flesch-Kincaid Grade Level, and FORCAST Formula) were used to compare passage difficulty between PISA and NAEP (at the 8th and 12th grades) (see table 2). These analyses excluded passages of less than 250 words (which included 3 poetry passages in NAEP and 8 passages in PISA) because of the lower

[7] The reading passage length calculations reported here for NAEP include combined scores of paired passages and individual reading passage scores.

reliability of applying the formulas to passages of such length. [8] In general, PISA passages were somewhat closer in difficulty to 12th-grade NAEP than to 8th-grade NAEP. The PISA passages tended to cover a broader range of readability or grade levels than did the NAEP passages.

Using the formulas best suited to *continuous* text (Dale-Chall and Flesch-Kincaid)—although applied to both continuous and noncontinuous passages—passages from the NAEP 8th-grade assessment corresponded, on average, to a 7th-grade level according to Dale-Chall and 8th-grade level according to Flesch-Kinkaid, with a range that extended from 5th grade to 12th grade. The average grade level for NAEP 12th-grade passages was the 7th grade according to Dale-Chall and the 9th grade according to Flesch-Kincaid. The range of grade levels extended from 5th grade to 13th grade. The average grade level for PISA passages was the 8th grade for Dale-Chall and the 9th grade for Flesch-Kincaid. The range of grades extended from 3rd grade to 15th grade. Using the formula best suited to *noncontinuous* texts (FORCAST)—although again applied to both passage types—the NAEP 8th-grade and 12th-grade and PISA passages all averaged at the 10th-grade level.

When the passages for each assessment are separated by format (continuous and noncontinuous), other differences emerge (see appendix table B). Within assessments, PISA's continuous passages tend to be more difficult than its noncontinuous passages. On the other hand, NAEP's relatively few strictly noncontinuous passages at the 12th-grade level are more difficult than its continuous texts at that grade level. In fact, NAEP 12th-grade noncontinuous passages are, on average, more difficult than PISA noncontinuous passages, which are more numerous, and the NAEP 12th-grade noncontinuous passages are among the most challenging in either assessment.

[8] Eighth-grade NAEP reading passages included all individual reading passages categorized as grades 4/8, 8, and 8/12. Twelfth-grade NAEP reading passages included all individual reading passages categorized as grades 8/12 and 12. All readability analyses were calculated based on individual reading passages and do not take into account cases in which passages were paired.

Table 2. NAEP 2009 and PISA 2009 reading passage readability analyses

Readability measure	NAEP grade 8[1] (*N* = 14)	PISA (*N* = 21)	NAEP grade 12[2] (*N* = 16)	Description of measure
Dale-Chall Formula				The Dale-Chall Formula uses a familiar words list common to students and rates the text against it as well as the sample's total number of words and sentences. The more familiar words there are in a text, the easier the text is scored.
Average Dale-Chall grade	6.9	7.6	7.4	
Range of Dale-Chall grades	5.2 - 8.3	5.8 - 9.9	5.4 - 9.4	
Flesch Reading Ease				The Flesch Reading Ease measure is based on the number of words, syllables, and sentences in adult reading materials. Reading Ease scores fall between 0 and 100, with a lower number indicating more difficult material. The readability scores correspond to the following readability levels: very difficult (0-29), difficult (30-49), fairly difficult (50-59), standard (60-69), fairly easy (70-79), easy (80-89), and very easy (90-100).
Average readability score	69.4	60.7	62.4	
Range of readability scores	51 – 85	32 – 88	41 – 83	
Flesch readability identifier	Standard	Standard	Standard	
Flesch-Kincaid Grade Level				The Flesch-Kincaid Grade Level formula is most reliable when used to assess upper elementary and secondary materials. Similar to the Flesch Reading Ease measure, it is based on the number of words, syllables, and sentences in a text, but with slightly different weighting.
Average grade level	7.5	8.7	8.8	
Range of grade levels	4.8 - 11.5	3.2 - 15.1	5.3 - 13.4	
FORCAST Formula				The FORCAST Formula focuses on the number of single-syllable words present in a text. It is usually used in evaluating questionnaires, forms, tests, and job materials not in narrative form and often absent of any end punctuation, such as periods or question marks.
Average FORCAST grade	9.5	10.2	9.8	
Range of FORCAST grades	7.9 - 10.6	8.2 - 11.7	8.3 - 11.7	

[1]This column presents averages and ranges of all individual reading passages categorized as grades 4/8, 8, and 8/12.
[2]This column presents averages and ranges of all individual reading passages categorized as grades 8/12 and 12.
NOTE: Only passages with word counts over 250 were included in the readability analyses. Excluded passages included three from NAEP (which include the poetry reading passages) and eight from PISA.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2009 Reading Assessment; and Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2009 Reading Literacy Assessment.

How Interchangeable Are the PISA and NAEP Passages?

The expert panel was asked to determine how well NAEP and PISA passages would fit in the other's framework or, in other words, the likelihood that the passages of one assessment could appear in the other. The expert panel reviewed a sample of about 70 percent of the passages from each assessment selected to represent the full range of the frameworks in terms of text type and length.[9] They found that PISA passages tended to fit better to the NAEP framework than NAEP passages did to the PISA framework, although a substantial number of passages from both assessments were deemed not interchangeable (see table 3). About half of the NAEP 8th-grade and two-thirds of the NAEP 12th-grade passages that were reviewed did not fit within the PISA framework. Just over two-fifths of the PISA passages that were reviewed did not fit within the NAEP framework at either the 8th or 12th grade. The most typical reason for lack of fit of NAEP passages was the prominence of "author's craft" in NAEP. ("Author's craft" refers to the specific techniques used by an author to relay an intended message.) NAEP texts such as poetry or rhetorical narratives would be difficult to translate into the various languages required by PISA, as would maintaining the tone and quality of a text as the author intended. The experts concluded that some PISA passages would not appear in the NAEP assessment because there was too much disconnected text (or presentation of multiple stimuli not strictly related), the texts were not authentic enough, or the passages were simply too short.

Table 3. Percentage distribution of NAEP 2009 and PISA 2009 passages that fit the other assessment's framework

Fits/Does not fit	PISA to NAEP (Grade 8 or 12)	NAEP to PISA	
		Grade 8	Grade 12
Fits other framework	57	50	33
Does not fit other framework	43	50	67

NOTE: The judgment of the expert panel, rather than any specific formula, was used assess "fit" and to calculate the percentage. Analyses are based on a representative sample of roughly 70 percent of the passages from each of the three assessments.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Database of expert panel ratings, June 2010.

Item Format

Both NAEP and PISA include multiple-choice items from which students choose one correct answer. All NAEP multiple-choice items include four response options, whereas PISA multiple-choice items include four or five response options (see table 4). In addition to traditional multiple-choice items, PISA also includes what it calls "complex multiple choice" items, which require students to answer a series of multiple-choice or true/false questions based on the same information.

Both NAEP and PISA also include constructed-response items, for which students must supply the response. NAEP has "short answer" and "extended response" items. As described in the 2009 NAEP reading framework:

[9] This included 6 individual and 1 set of paired passages (8 total individual) from the NAEP 8th grade assessment; 3 individual and 3 sets of paired passages (9 total individual) from the NAEP 12th grade assessment; and 21 passages from the PISA assessment.

> Short constructed-response items can be answered by one or two phrases or by one or two sentences; they should take students approximately 2 to 3 minutes to complete. Extended constructed-response items should elicit longer, more elaborated answers of a paragraph or two. They should take students approximately 5 minutes to complete. Scoring rubrics for short and extended constructed-response items will focus on the content included in answers, not on spelling or grammatical considerations. However, students must answer constructed-response questions by using information from the text to receive credit (National Assessment Governing Board 2008, p. 40).

PISA classifies its constructed-response items as open constructed response, short constructed response, and closed constructed response. Open constructed-response items may require a description or an explanation to support a response and may be scored for partial credit, though the acceptable length of response is much less than in NAEP. Short constructed-response items typically require students to supply a word or phrase or may require students to provide a specific response from the text. Closed constructed-response items are described as those that "require the student to generate a response, but that require minimal judgment on the part of a coder" (OECD 2009, p. 46).

Table 4. Percentage distribution of items by item format in NAEP 2009 and PISA 2009 reading assessments

Item format	NAEP grade 8 (N = 130)	NAEP grade 12 (N = 131)	PISA (N = 104)
Multiple choice	59	58	38
Complex multiple choice	†	†	9
Constructed response			
Extended response	10	10	†
Open response	†	†	35
Short answer	31	32	8
Closed	†	†	11

† Not applicable.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2009 Reading Assessment; and Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2009 Reading Literacy Assessment.

Cognitive Processes

While the texts used form the content of the assessment, the cognitive processes ("aspects" in PISA and "cognitive targets" in NAEP) define the skills and abilities that students must draw on in response to the texts. Each item is written to primarily address one process. Although PISA and NAEP have three similarly named and defined cognitive process categories (shown in exhibit 3), there are differences that influence the kinds of items presented to students in each assessment.

Exhibit 3. Cognitive process categories

PISA 2009		NAEP 2009	
Aspect categories		Cognitive targets	
Access and retrieve	Students navigate the information space provided to locate and retrieve one or more distinct pieces of information.	Locate and recall	Students locate or recall information from what they read; identify clearly stated main ideas or supporting details; and find essential elements of a story, such as characters, time, or setting.
Integrate and interpret	Students develop an understanding of the coherence of the text and make meaning from something that is not stated.	Integrate and interpret	Students integrate new information into their initial sense of what a passage says, often interpreting what they read in the process; make comparisons and contrasts of information or character actions; examine relations across aspects of text; and consider alternatives to what is presented in text.
Reflect and evaluate	Students are required to draw upon knowledge, ideas, or attitudes beyond the text in order to relate the information provided within the text to their own conceptual and experiential frames of reference.	Critique and evaluate	Students are required to stand back from what they read and view the text objectively. The focus remains on the text itself, but the reader's purpose is to consider the text critically by assessing it from numerous perspectives and synthesizing what is read with other texts and other experiences.

SOURCE: Organization for Economic Cooperation and Development (OECD), Program for International Student Assessment (PISA), *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*, 2009; and U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), *Reading Framework for the 2009 National Assessment of Educational Progress*, 2008.

Experts' Comparison of PISA and NAEP Reading Items

The experts were also asked to review the items associated with the reviewed passages. They assessed the extent to which PISA and NAEP items fit into the other framework's cognitive categories and whether or not the items fit the other framework in terms of the nature and format of the item. Additionally, items were reviewed for factors that contributed to their ease or difficulty. Panelists also looked at the PISA items to consider whether, in terms of level of challenge, they fit more closely to the NAEP 8th- or 12th-grade assessment.

Fit to Other Framework: Cognitive Processes

In general, the NAEP and PISA items reviewed tended to fit within the other framework's cognitive categories—that is, the items required a similar range and type of reading and thinking skills. Only

about 4 to 6 percent of items in each assessment (including 8th- and 12th-grade NAEP and PISA) were rated as completely outside the other assessment's framework cognitive categories. However, there were a number of items, especially in PISA, that were considered more borderline for cognitive fit or that were thought to fit within the other assessment's framework, but in a different cognitive category . For example, there were some PISA items classified as "reflect and evaluate" that the panel thought would be considered "integrate and interpret" in the NAEP framework, as well as some PISA "integrate and interpret" items that might be classified as "locate and recall" in NAEP. Adjusting for these borderline items, the experts thought that, overall, about 90 percent of both NAEP 8th- and 12th-grade items fit PISA's cognitive categories tightly and well and that about 80 percent of PISA items fit the NAEP cognitive categories tightly and well. The expert panel remarked that it is more difficult for a student to read and answer questions from a passage that is significantly longer in length than it is from a shorter passage. The panel considered this difference when deciding the "fit" of PISA passages into the NAEP reading assessment cognitive categories.

PISA items that were different from NAEP items cognitively included those that asked students to provide a personal stance or required a written response that was not dependent on text-based evidence and those that drew on multiple cognitive skills—scenarios that would not occur in NAEP. NAEP items that seemed different, or somewhat different, from PISA items cognitively were mainly NAEP vocabulary items, which required students to identify word meaning within the passage context and that have no corollaries in PISA.

Within individual cognitive categories, the most challenging items to fit to the other framework's cognitive categories appeared to be the PISA "reflect and evaluate" items. About 14 percent of these items did not match the cognitive categories of the NAEP framework at all, and about 20 percent matched a different cognitive category in NAEP. PISA's emphasis on inclusion of the student's own experiences in "reflect and evaluate" items sometimes fit better NAEP's "integrate and interpret" category than its "critique and evaluate" category.

Fit to Framework: Item Format

The experts found more differences between the assessments in item format than in cognitive skills measured. Although PISA and NAEP items tended to measure similar cognitive skills, they often were presented or formatted in ways that were dissimilar between the assessments. Over one-third of NAEP 8th-grade items and nearly two-fifths of NAEP 12th-grade items were judged incompatible with the PISA framework in terms of their nature or format; over half of the PISA items were judged incompatible with the NAEP framework. In general, the NAEP items that did not fit PISA were either vocabulary items or items that required a response that used information from the text to support it, for which PISA does not have corollary formats. PISA items did not fit NAEP for more varied reasons. Some PISA items did not fit NAEP because of the relatively short length of response acceptable for a correct answer, others because of the use of scaffolding or introduction to test items. Many items did not fit because of format differences. PISA used the closed constructed-response formats for low-level cognitive items in cases in which NAEP would only use multiple choice, and PISA used formats such as "complex multiple-choice" or included visuals in the item or responses, which also would not occur in the NAEP reading assessment.

Factors Contributing to Cognitive Challenge

The experts reviewed the items to identify the factors that contributed to their cognitive challenge, or what drives item difficulty. Each item could be assigned multiple factors from a list of nine , the

first 8 of which are based on factors described in the PISA framework (OECD 2009) and the 9th which was added by the experts:

1. number of pieces of information needed to locate/consider;
2. amount of inference required;
3. amount and prominence of competing information;
4. length and complexity of text;
5. type of interpretation required;
6. familiarity with structure and genre;
7. nature of knowledge needed to bring to item (narrow v. broad);
8. depth of understanding required; and
9. type of information.

For both 8th- and 12th-grade NAEP, the most prevalent factors contributing to item challenge were the type of interpretation required by the test-taker, the number of pieces of information to be located or considered, and the depth of understanding required to answer the item correctly. Each of these factors was present in at least 35 percent of the items. In most cases, these were viewed as factors that contributed to increased item challenge—that is, the interpretations or understanding required were relatively complex or deep and the amount of information to be sorted through was relatively great.

For PISA, the factors contributing to item challenge were typically the type of information that the student is required to handle, the number of pieces of information to be located or considered, and the amount of inference required. As in NAEP, these factors were present in at least 35 percent of items. Identifying the type of information that the student is required to handle was most often viewed as the factor that would increase the challenge level, because of PISA's use of visual or graph-based information, which is not routinely found in NAEP and is considered fairly challenging. The other two factors were most often viewed as contributing to a relatively lower challenge level, as the amount of text to be sorted through was not as great as in NAEP and the amount of inference required was not as great as it might have been in the context of longer passages.

Appropriate Level of Challenge for NAEP (PISA Items)

In a final analysis, the experts examined whether or not the PISA items would be appropriate in terms of level of challenge for the NAEP 8th- or 12th-grade assessment. The experts considered what the PISA items required of students and how well that aligned with the items in the NAEP 8th- and 12th-grade assessments. The experts found that about 55 percent of PISA items would be suitable for the NAEP 8th-grade assessment and about 15 percent would be suitable for the NAEP 12th-grade assessment. However, about 30 percent of the items were thought to be inappropriate for NAEP in terms of level of challenge, and some of the items deemed suitable for the 8th grade were considered more borderline, or on the lower end of what would be acceptable, for this grade.

Summary

NAEP measures in detail the reading knowledge of U.S. students as a whole, but can also provide trend information for individual states and some districts, different geographic regions, and demographic population groups. PISA provides a method for comparing the performance of U.S.

students in reading with that of students in other nations. The two assessments differ in some key design elements. Differences include the following:

- The content assessed by PISA and NAEP differ in subtle, but important ways. NAEP is tailored specifically to practices and standards used in the United States; in PISA, the content is determined internationally, in collaboration with other countries and reflecting consensus views of key content. Also, PISA's specific focus on the "yield" of education systems and the application of competencies in real-world contexts distinguishes it from NAEP, which focuses more closely on measuring school-based performance.

- Different target populations of students are assessed. Main NAEP uses grade-based samples targeting 4th-, 8th-, and 12th-grade students. PISA uses an age-based sample, which targets 15-year-olds, who are most likely between the ages of the NAEP target populations of 8th- and 12th-graders.

- Measurement precision is greater in NAEP than in PISA. NAEP and PISA are both designed to provide valid and reliable measures of U.S. students' performance in the aggregate as well as for major subpopulations, and each study draws a sample sufficient for this purpose. NAEP, however, is also designed to provide estimates for individual states, which requires an increased sample size, and thus measures performance at a higher level of precision than PISA. This difference may have an impact on the assessments' sensitivities in detecting changes in student performance.

- There is some overlap in how reading is defined in the two assessment programs and some similarities in how the frameworks are organized, with both NAEP and PISA specifying a cognitive dimension and a range of text types. However, there are subtle differences in how the cognitive categories are defined and more notable differences in the text types targeted for inclusion, as well as features (e.g., an assessment of vocabulary embedded within NAEP) that are unique to each assessment.

- The passages selected for NAEP and PISA would likely fit in each other's frameworks to only a limited degree. For example, NAEP passages, on average, are longer than PISA passages. Another, related difference is PISA's more frequent use of graphic and other visual displays of text rather than continuous text passages. In terms of readability and grade level, PISA passages were generally more comparable to 12th-grade NAEP than to 8th-grade NAEP.

- NAEP and PISA items generally tend to measure similar cognitive skills; however, they often are presented or formatted in ways that would not be interchangeable between the assessments. Key differences include PISA's less extensive use of multiple-choice and more extensive use of short-constructed response formats than NAEP, while NAEP requires much longer, text-based responses for its extended constructed response formats.

- Finally, there are differences in the source of challenge for NAEP and PISA items; these differences appear to be driven by the inclusion of longer passages in NAEP and the inclusion of more visual and other noncontinuous text formats in PISA. PISA items were found more frequently to be appropriate for the NAEP 8th-grade assessment than the 12th-grade assessment.

Contact Information

Dan McGrath
Director, International Activities Program
National Center for Education Statistics
U.S. Department of Education
1990 K Street NW
Washington, DC 20006
Tel.: (202) 502-7426
E-mail: Daniel.McGrath@ed.gov

Eunice Greer
National Assessment of Educational Progress
National Center for Education Statistics
U.S. Department of Education
1990 K Street NW
Washington, DC 20006
Tel.: (202) 502-7488
E-mail: Eunice.Greer@ed.gov

Useful Websites

NAEP: http://nces.ed.gov/nationsreportcard
PISA: http://www.nces.ed.gov/surveys/pisa (national)
http://www.pisa.oecd.org (international)

References

Fleischman, H., Hopstock, P., Pelczar, M., and Shelley, B. (2010). *Highlights From PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context* (NCES 2010-027). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S.

National Assessment Governing Board (2008). *Reading Framework for the 2009 National Assessment of Educational Progress*. U.S. Department of Education, Washington, DC.

National Center for Education Statistics (2009). *The Nation's Report Card: Reading 2009* (NCES 2010-458). Institute of Education Sciences, U.S. Department of Education, Washington, DC.

National Center for Education Statistics (2010). *The Nation's Report Card: Grade 12 Reading and Mathematics 2009* (NCES 2011–455). Institute of Education Sciences, U.S. Department of Education, Washington, DC.

Organization for Economic Cooperation and Development (2009). *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*. OECD, Paris.

Appendix Table A. PISA and NAEP reading passage information by text format

	Continuous passages			Noncontinuous passages		
	NAEP 8 (*N* = 13)	**PISA (*N* = 14)**	**NAEP 12 (*N* = 11)**	**NAEP 8 (*N* = 0)**	**PISA (*N* = 15)**	**NAEP 12 (*N* = 2)**
Length of passage and items per passage						
Average number of words	923.6	385.9	1127.7	†	325.0	1425.5
Range of words in passages	219 - 1,429	115 - 758	771 - 1,429	†	53 - 577	1,262 - 1,589
Average number of items	10.0	3.6	10.0	†	3.6	10.5
Range of items	9 - 11	2 - 5	9 - 11	†	2 - 5	10 - 11

†Not applicable.
NOTE: NAEP 8 columns present averages and ranges of all reading passages categorized as grades 4/8, 8, and 8/12. NAEP 12 columns present averages and ranges of all reading passages categorized as grade 8/12 and 12. NAEP presents students with both individual passages and paired passages. Length analyses were calculated based on the passage or set of passages a student received with each set of items in order to accurately reflect reading load.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2009 Reading Assessment; and Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2009 Reading Literacy Assessment.

Appendix Table B. PISA and NAEP readability analyses by text format

	Continuous passages			Noncontinuous passages		
	NAEP 8 (*N* = 14)	**PISA (*N* = 10)**	**NAEP 12 (*N* = 13)**	**NAEP 8 (*N* = 0)**	**PISA (*N* = 11)**	**NAEP 12 (*N* = 3)**
Dale-Chall						
Average grade	6.9	7.8	7.0	†	7.4	8.7
Range of grades	5.2 - 8.3	6.1 - 9.9	5.4 - 9.4	†	5.8 - 8.3	8.5 - 9.1
Flesch Reading Ease						
Average readability score	69.4	58.1	65.5	†	63.1	48.7
Range of readability scores	51 - 85	32 - 88	41 - 86	†	51 - 77	42 - 56
Flesch Readability identifier	Standard	Fairly difficult	Standard	†	Standard	Difficult
Flesch-Kincaid Grade Level						
Average grade	7.5	9.4	8.7	†	8.1	9.3
Range of grades	4.8 - 11.5	3.2 - 15.1	5.3 - 13.4	†	4.8 - 10.6	7.6 - 10.6
FORCAST						
Average grade	9.5	10.2	9.4	†	10.1	11.4
Range of grades	7.9 - 10.6	8.2 - 11.7	8.3 - 10.8	†	9.3 - 11.4	10.9 - 11.7

†Not applicable.
NOTE: Only passages with word counts over 250 were included in the readability analyses. Excluded passages included three from NAEP and eight from PISA. NAEP 8 columns present averages and ranges of all individual reading passages categorized as grades 4/8, 8, and 8/12. NAEP 12 columns present averages and ranges of all individual reading passages categorized as grade 8/12 and 12.
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2009 Reading Assessment; and Organization for Economic Cooperation and Development, Program for International Student Assessment (PISA) 2009 Reading Literacy Assessment.