Study of the Alignment of the 2015 NAEP Mathematics Items at Grades 4 and 8 to the Common Core State Standards (CCSS) for Mathematics

Phil Daro
Strategic Education Research Partnership (SERP) Institute

Gerunda B. Hughes
Howard University

Fran Stancavage
American Institutes for Research

October 2015
Commissioned by the NAEP Validity Studies (NVS) Panel

The NAEP Validity Studies Panel was formed by American Institutes for Research under contract with the National Center for Education Statistics. Points of view or opinions expressed in this paper do not necessarily represent the official positions of the U.S. Department of Education or American Institutes for Research.

Since 1995, AIR has maintained the **NAEP Validity Studies Panel (NVS)**, an independent panel of experts that meets three times a year to commission and discuss research addressing important validity considerations for the NAEP program. Completed research papers are published under the NVS report series and available through the AIR website at http://www.air.org/project/naep-validity-studies-nvs-panel

Panel Members:

Project Director:
Frances B. Stancavage
American Institutes for Research

Project Officer:
Grady Wilburn
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS)
American Institutes for Research
2800 Campus Drive, Suite 200
San Mateo, CA 94403
Phone: 650/843-8163

Executive Summary

For more than four decades, the National Assessment of Educational Progress (NAEP) has provided the best available information about the academic achievement and educational progress of the nation's students. NAEP results also are used to make comparisons among participating states and to track the progress of states over time. To support this mission, the National Assessment Governing Board periodically adjusts the NAEP content frameworks to reflect curricula commonly taught in the United States while also, where possible, avoiding abrupt changes that would require a break in the trend line.[1]

The Common Core State Standards (CCSS) for mathematics and English language arts arrived on the educational scene in 2010 and were adopted shortly thereafter by nearly all of the states, four territories, and the District of Columbia. Although actual classroom implementation of the CCSS has so far varied substantially, the current and anticipated influence of the CCSS on instruction, as well as the administration of CCSS-aligned assessments by many states, suggest the need to examine the alignment between the content covered by NAEP assessments and the content covered by the CCSS and its associated assessments.

In spring 2011, the NAEP Validity Studies (NVS) Panel began an investigation of the validity and utility of NAEP in the context of the CCSS. The study reported here is the second in a planned series of comparisons between the content of the NAEP mathematics assessment and the CCSS for mathematics. The first study compared the NAEP mathematics framework and the CCSS. This second study compares the items in the 2015 NAEP item pool and the CCSS. An anticipated third study would compare the NAEP items with the items in assessments built specifically to align with the CCSS.

Research Questions and Methodology

The study is built around two complementary research questions (RQs) that allow for bidirectional comparisons between NAEP and the CCSS:

RQ 1: To what extent does the NAEP item pool include content that is targeted by the CCSS for instruction at or below the grade level tested by NAEP? What is the alignment profile of NAEP to the CCSS across NAEP content areas?

RQ 2: To what extent do the CCSS target for instruction, at or below the grade level tested by NAEP, content that is assessed by NAEP? What is the alignment profile of the CCSS to NAEP across CCSS domains?

To address these questions, the study authors assembled a panel of 18 mathematicians, classroom teachers (or recent classroom teachers), mathematics supervisors, and mathematics educators to classify all of the items (150 items at each grade level) that were

[1] In 2005, NAEP did break the trend line for grade 12 mathematics as a result of revisions to the framework for that grade level.

used in the 2015 NAEP mathematics assessment for grades 4 and 8 by matching them to appropriate CCSS standards or determining that there were no suitable matches in the CCSS.

Findings

Overall, the review by expert panelists suggests that concordance between NAEP and the CCSS is reasonable at both grade levels. It makes sense that NAEP, which is required by its mission to be broad, will include some items that are outside the bounds of the CCSS and will not assess every standard in the CCSS.

However, the agreement between NAEP and the CCSS is uneven across the five NAEP content areas, and, at grade 8, 42% of middle-grade CCSS standards are not assessed by any items in the 2015 NAEP item pool. Highlights of the specific findings for each grade level follow.

Grade 4. The overall agreement between NAEP and the CCSS is reasonable and also fairly symmetrical at grade 4: panelists found that 79% of NAEP items were clearly matched to content that appears in the CCSS standards at or below grade 4 (RQ 1) and 77% of grade 3 and 4 CCSS standards were assessed by at least one NAEP item (RQ 2). However, the areas of divergence differ between the two analyses: The content area with the lowest percentage of NAEP items assessing mathematics covered in the CCSS by grade 4 is data analysis, statistics, and probability (47%); the fewest CCSS standards matched by NAEP are in the operations and algebraic thinking domain (57%). From the CCSS point of view, this domain is the centerpiece of coherence in the progression from arithmetic to algebra.

Grade 8. At grade 8, the overall alignment of NAEP items to the CCSS standards at or below grade 8 appears strong at 87%, although there is variation in fit across content areas that generally parallels the variation seen at grade 4 (RQ 1). However, there is more divergence when concordance is looked at from the opposite direction: 58% of grade 6, 7, and 8 CCSS standards are linked with at least one grade 8 NAEP item from the 2015 item pool (RQ 2). There appears to be a notable amount of middle-school mathematics content recommended by the CCSS that is not part of the current NAEP assessment.

Conclusion

The mission of NAEP is to provide a valid and reliable measure of progress for the nation over time, and NAEP results make the most sense over the longer timelines for which trend assessments are designed. Viewing results over a number of years establishes an important context for interpreting the present and for looking ahead.

Prioritizing trend means that NAEP must maintain a degree of independence from what might be seen as current fashions in instruction and curriculum. At the same time, NAEP must periodically review and update its frameworks because it cannot provide useful results if the content that it measures is too far from the content of contemporary instruction.

Given the historical variation in instructional standards across states, NAEP has sought to represent the union of all major curricula when developing and revising its frameworks. It may be that the landscape is changing now for mathematics and ELA, with many states having adopted either the CCSS or other college and career readiness standards. However, changing standards is always far simpler and faster than changing curriculum and instruction, and both time and research are needed to see how the CCSS actually play out in U.S. classrooms. At the same time, the immediate salience of the CCSS is raised by the fact that many states are right now basing accountability on CCSS-aligned state assessments.

It has been 10 years since NAEP conducted a major review of its mathematics framework. Despite some uncertainties concerning the way that the CCSS will influence instruction over time, we believe that this is an appropriate moment for NAEP's Governing Board to review the framework in light of the CCSS as well as other states' college and career standards. The findings from this study suggest particular areas where such a review should focus.

Acknowledgments

The study authors wish to thank the 18 mathematicians, classroom teachers (or recent classroom teachers), mathematics supervisors, and mathematics educators who graciously gave their time on very short notice to help with this study, and who participated with great integrity and enthusiasm. We also thank NVS panel members George Bohrnstedt and David Thissen, who consulted on the design of the study and, along with panel members Peter Behuniak and Scott Norton, reviewed and commented on our draft report.

CONTENTS

Introduction

For more than four decades, the National Assessment of Educational Progress (NAEP) has provided the best available information about the academic achievement and educational progress of the nation's students. NAEP results also are used to make comparisons among participating states and to track the progress of states over time. To support this mission, the National Assessment Governing Board periodically adjusts the NAEP content frameworks to reflect curricula commonly taught in the United States while also, where possible, avoiding abrupt changes that would require a break in the trend line.[2]

The Common Core State Standards (CCSS) for mathematics and English language arts (ELA) arrived on the education scene in 2010 and were adopted shortly thereafter by nearly all of the states, four territories, and the District of Columbia. Around the same time, two multistate consortia—the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC)—were formed and then funded by the federal government for the purpose of developing comprehensive assessment systems aligned with the CCSS, the results of which would be used by states to make decisions about educational and, in some cases, educator accountability. Although actual classroom implementation of the CCSS has so far varied substantially, there is evidence that concepts about mathematics instruction embedded in the standards are beginning to influence instruction, even in states that never adopted the CCSS or backed away from initial adoption (Santelises & Dabrowski, 2015; Heitin, 2015). The current and anticipated influence of the CCSS on instruction, as well as the administration of CCSS-aligned assessments by many states, suggest the need to examine the alignment between the content covered by NAEP assessments and the content covered by the CCSS and its associated assessments.

In spring 2011, the NAEP Validity Studies (NVS) Panel began an investigation of the validity and utility of NAEP in the context of the CCSS. Two interrelated studies comparing NAEP and the CCSS were carried out: one in reading and writing and the other in mathematics. At that time (and in the absence of CCSS-aligned assessments), the studies were conceived and designed to compare the content of the NAEP reading, writing, and mathematics frameworks for grades 4 and 8 with the CCSS in ELA and mathematics.[3] The intent was to follow up with additional studies examining actual assessment items, as these became available.

The 2011 study of mathematics focused primarily on the conceptual match between the mathematics subtopics and objectives in the *Mathematics Framework for the 2011 National Assessment of Educational Progress* (National Assessment Governing Board, 2010)[4] and the

[2] In 2005, NAEP did break the trend line for grade 12 mathematics as a result of revisions to the framework for that grade level.

[3] Because the NAEP frameworks for reading and writing do not have a level of specificity that would support detailed comparisons to the CCSS, the authors of the ELA study also included NAEP passages, items, and scoring guides in their analysis. For the mathematics study, analyses were strictly limited to a comparison of the NAEP framework and the CCSS.

[4] Although a 2010 version of the NAEP framework was used in the 2011 study, the framework has undergone only minimal changes between 2005 and 2015.

CCSS content standards for mathematics (Common Core State Standards Initiative, 2010). The 2011 study's findings were generated by the judgments of 14 expert panelists, who identified specific NAEP mathematics content that was not matched well or at all in the CCSS at or below the grade level tested by NAEP, as well as CCSS mathematics content that was not matched well or at all in the NAEP mathematics framework.

The study reached four major conclusions: (1) The CCSS have more rigorous content in eighth-grade algebra and geometry than NAEP; (2) the CCSS distribute the development of mathematical expertise and practices throughout the standards, whereas NAEP does not; (3) the CCSS attend to developing conceptual understandings in more topics than the NAEP frameworks; and (4) the CCSS introduce some mathematics content, such as probability and ratios and proportional reasoning, at higher grades than those assessed in NAEP (Hughes, Daro, Holtzman, & Middleton, 2013).

The next set of studies that the NVS Panel wanted to pursue was one that would compare NAEP reading, writing, and mathematics items and CCSS-aligned consortium items with one another and with their respective frameworks/standards. Upon consideration, however, the panel decided that such studies should be deferred until 2017, by which time NAEP would have largely transitioned to a digital-based assessment (DBA).[5] In the meantime, the NVS Panel asked Daro and Hughes to extend their earlier analysis in mathematics to include a comparison of NAEP items (specifically the 2015 item pools for grades 4 and 8) and the CCSS. [6]

The current study is built around two complementary research questions (RQs) that allow for bidirectional comparisons between NAEP and the CCSS:

RQ 1: To what extent does the NAEP item pool include content that is targeted by the CCSS for instruction at or below the grade level tested by NAEP? What is the alignment profile of NAEP to the CCSS across NAEP content areas?

RQ 2: To what extent do the CCSS target for instruction, at or below the grade level tested by NAEP, content that is assessed by NAEP? What is the alignment profile of the CCSS to NAEP across CCSS domains?

[5] The NAEP mathematics framework is not scheduled to be updated before 2017. However, there is an expectation that the new DBA items will allow for better measurement of some constructs already in the NAEP framework.
[6] Since the 2011 study of ELA already included an analysis of NAEP assessment items, no corresponding study was mounted for that subject area.

Methodology

Expert Panel

The study authors assembled a panel of 18 mathematicians, classroom teachers (or recent classroom teachers), mathematics supervisors, and mathematics educators to review the alignment between the 2015 NAEP mathematics item pool and the CCSS in mathematics. In order to ensure that the item review could be completed in a reasonable amount of time, we selected panelists who were already familiar with the CCSS through their work.[7] At the same time, we attempted to enlist panelists who held a variety of opinions on the value of the CCSS. The list of panelists, with their primary affiliations, is included in Appendix A.

Panelists participated in a two-hour webinar in which we introduced them to the study's purpose and methodology and then practiced the classification task (described below) using released NAEP items. Subsequently, panelists met for a one and a half day meeting in Washington, D.C., on September 16–17, 2015. Day 1 began with panelists once again practicing the classification task using released items. Panelists were then divided into grade-level groups (nine panelists per grade level) and given the task of classifying all of the items (150 items at each grade level) that were used in the 2015 NAEP mathematics assessment for grades 4 and 8.

Panelists worked independently to classify each NAEP item into a specific CCSS standard or determine that the item did not match any CCSS standard. Panelists were instructed to first work an item, noting the mathematical demands, and then search the CCSS for an appropriate standard, starting at the grade level where the item was administered by NAEP (grade 4 or 8). If no relevant standard was found at that grade level, they were to search one grade level lower, then one grade level higher, then lower and higher as made sense.[8]

It is typical for items to involve mathematics from several grade levels. In these cases, panelists were instructed to identify the CCSS standard at the grade level that matched the most demanding mathematics needed for the task. For example, if the task required subtraction from grade 2 and division from grade 5, panelists were to choose the division standard at grade 5.

Panelists also were told to use professional judgment in deciding whether the alignment between an item and a standard was close enough to warrant assigning the item to that standard. During the meeting, this guidance was refined to say that an item could be assigned to a standard if the panelist was reasonably confident that the knowledge and skills required to answer the item would be taught in classrooms that were teaching that standard, even if there was some degree of mismatch between the item and the specific wording of the standard. It was, of course, completely acceptable to classify the item as not matching any CCSS standard or matching a CCSS standard from a grade level higher than that being tested by NAEP.

---

[7] Familiarity with the NAEP framework was judged less important because we would not be matching items to that framework.

[8] This set of guidelines was intended to streamline the matching process and seemed sufficient given that the primary concern was whether or not the item represented content that would have been taught at or below the grade level tested.

Items were batched by NAEP content area, and panelists convened after completing each content area to discuss their classifications. Due to time constraints, discussion was limited to items on which there was disagreement as to whether the item was matched by the CCSS content at or below the grade level tested by NAEP (which is the critical distinction for RQ 1). With nine panelists rating each item, discussion was triggered if three or more panelists disagreed with the majority judgment on a given item (as indicated by a show of hands). Following discussion, panelists could change their classification if they wanted, but there was no effort to achieve consensus, given both the constraints of time and the complexity of the matching task, on which reasonable people could disagree. (Procedures for coding and reconciliation are described more fully in Appendix B, which reproduces the written guidance provided to panelists.)

The three NAEP content areas of measurement; data analysis, statistics, and probability; and geometry were classified on the first day of the meeting. On the second day, we determined that a change in procedure was necessary if we were to finish the two remaining content areas (algebra and number properties and operations) in the available time. Consequently, the panelists at each grade level were split, and items in algebra and number properties and operations were each classified by either four or five panel members. With fewer panelists per item, the threshold for triggering discussion of individual items also was lower. Items were discussed if two panelists differed from the others with regard to classifying the item as matched by standards at or below the grade level tested by NAEP.

Day 2 concluded with a general discussion among all 18 panelists about their experience with the alignment activity and the patterns they saw in the item classifications. Panelists also were asked to submit written notes on this topic. These notes provided context for the interpretation of the numerical findings.

Analysis

Three major analyses were carried out to address the RQs. All analyses were done separately for grade 4 and grade 8.

For RQ 1:

1. Overall and by NAEP content area, what number and percentage of NAEP mathematics items were judged to be matched to the CCSS in mathematics at or below the grade level tested by NAEP?[9]
2. Overall and by NAEP content area, among the items judged as not matched, how many were aligned with standards at higher grade levels and how many were not aligned with any CCSS standards?

For RQ 2:

3. Overall and by CCSS domain, what percentage of CCSS standards was matched by at least one NAEP item? (Analysis was restricted to items classified as being in the CCSS at or below the grade level tested by NAEP.)

[9] The question of alignment profile was addressed by considering how the match between NAEP and the CCSS varied across NAEP content areas (RQ1) or across CCSS domains (RQ2).

In carrying out these analyses, rules had to be established for classifying items on which the panelists disagreed. As described below, we used somewhat different rules for each of the analyses. In addition, if a panelist provided only partial information on a given item (which happened in a minority of cases), we retained the panelist's data in the analyses to the extent possible. For example, if a panelist specified the CCSS grade level, domain, and cluster (but not the specific standard) matched to a given item, then the panelist's classification was included in the first and second analyses, but not the third.[10]

In summarizing information for the first analysis (items matched at or below the grade level tested by NAEP), we decided to create three categories of items: "in the standards at or below the NAEP grade level," "not in the standards at or below the NAEP grade level," and "uncertain." Items with the most mixed classifications were assigned to the "uncertain" category. Table 1 shows our classification rules for items reviewed by different numbers of panelists.

Table 1. Rules for classifying items

Total number of panelists rating item	Item classification based on number of panelists rating item as in standards at or below grade level tested		
	Item classified as *in* if rated "in" by:	Item classified as *uncertain* if rated "in" by:	Item classified as *not in* if rated "in" by:
9	7–9 panelists	3–6 panelists	0–2 panelists
5	4–5 panelists	2–3 panelists	0–1 panelists
4	3–4 panelists	2 panelists	0–1 panelists

NOTE: Although each item was intended to be reviewed by either nine panelists (Day 1) or four or five panelists (Day 2), a few items were actually classified by seven or eight panelists. Comparable rules were created for groups of these sizes.

For the remaining analyses, we classified an item based on the modal response among panelists rating the item. For the second analysis (items matching standards at higher grade levels versus items not matching any standards) , all but two of the (grade 8) items had a single mode that could be used for classification; the two items for which there was no single mode are noted separately when the grade 8 results are presented in Table 6.

For the third analysis (covering RQ 2), our approach to assigning items for which there was no single mode was to assign fractional item weights to each of the tied modes, with the constraint that at least two panelists had to have associated an item with that standard for the standard to receive any fractional weight for that item.

For example, if the distribution of panelists' ratings for a particular item was Std A, Std A, Std A, Std B, Std B, Std B, Std C, Std C, and Std D, then 0.5 of the item weight would be assigned to Std A and 0.5 to Std B, these being the two tied modes. On the other hand, if the distribution of panelists' ratings was Std A, Std B, Std C, Std D (across four panelists), then no weight for that item would be assigned to any standard.

[10] In some cases, information was missing because a panelist had run out of time for that batch of items. More often, it reflected the fact that the panelist did not feel that the item could be strictly classified into a specific standard, but still thought that the item was generally related to content covered by the CCSS.

Appendix C includes tables summarizing agreement among panelists at the levels of standards and with regard to the question of whether an item was matched to a standard above the grade level tested or was not in the CCSS at all. Appendix D provides information on the frequency of each possible pattern of ratings used to classify items, following the rules shown in Table 1.

Findings

In interpreting the findings of this study, it is important to note that changing standards is always far simpler and faster than changing curriculum and instruction. This report is limited to matching items and standards; the extent to which there have been changes in what is actually taught, as a result of introducing the CCSS standards, is surely quite variable across schools and jurisdictions and was not explored in this study.

Furthermore, differences between standards and test items create complexities even when attempting to answer this study's relatively narrow research questions. Standards, in general, and the CCSS, in particular, are not taxonomies of mathematics items. Items do not sort easily into standards. Good standards describe the mathematical knowledge and expertise students are expected to acquire. Test items provide occasions to use that knowledge and expertise to solve mostly short problems or answer short questions.

This difference between standards and test items gives rise to several difficulties when attempting classification. One difficulty is that even simple items typically relate to many standards, often an interdependent web of standards. For example, take a simple item such as finding the area of a rectangle at grade 3. This ties to a geometry standard, a measurement and data standard (lengths of sides), and an operations and algebraic thinking standard (multiplication). When asking experts to map such an item to standards, which standard should they chose? As noted in the methodology section, this study used the rule that, of the standards required to respond to the item correctly, panelists should chose the standard that matches the most demanding mathematics needed for the task. As a further complication, panelists observed that many items depend on a progression of standards across several grades. This makes sense in a subject such as mathematics, but it makes it challenging to pinpoint the specific grade and standard that should be assigned.

A second difficulty is that items can typically be solved in more than one way, and different students might solve the same item using different mathematics. A multiplication item, for example, can be solved using multiplication, repeated additions, or, in the case of multiple-choice items (where the correct answer is among the answer choices), estimation and guess and check. This study directed judges to rate the intended mathematics rather than speculate on what clever students might do to work around the intentions.

A third difficulty is that many standards require many items to fully assess the content of the standard. A test has a very limited number of items (although NAEP, with its matrix sampling approach, includes substantially more items than a typical state assessment). Some standards or problem types within standards may be oversampled in the item pool and others undersampled. Findings that show that high percentages of items can be matched to a set of standards are not necessarily findings that show that the test is well aligned with the standards. RQ 2 gives a first-level look at alignment by calculating the number of standards in each CCSS domain that are assessed by at least one item in the 2015 NAEP item pool. The broader question of how well the items sample the standards (balance and focus) is beyond the scope of this study.

With these considerations in mind, the findings are presented in the following sections.

Grade 4

RQ 1: *To what extent does the NAEP item pool include content that is targeted by the CCSS for instruction at or below the grade level tested by NAEP? What is the alignment profile of NAEP to the CCSS across NAEP content areas?*

Our analysis reveals that 79% of NAEP items were judged to assess mathematics that is included in the CCSS at grade 4 or below. Twelve percent of the NAEP items were judged to assess something that was either not found in the CCSS or was found at grade 5 or above, while an additional 9% of items were classified as "uncertain" due to substantial disagreement among panelists as to whether the item mapped to a standard at an appropriate grade level.

As shown in Table 2, the preponderance of grade 4 NAEP items from measurement and number properties and operations assessed content that is included in the CCSS standards at or below grade 4. The items that are not congruent with the CCSS are concentrated in algebra; data analysis, statistics, and probability; and geometry—especially data analysis, statistics, and probability, where only 47% of the items were classified as certain matches.

Table 2. Number and percentage of grade 4 NAEP items matching the CCSS at or below grade 4

Content area	In stds ≤ NAEP grade	Uncertain	Not in stds ≤ NAEP grade
Algebra (*N*=21)	13 62%	5 24%	3 14%
Data analysis (*N*=19)	9 47%	4 21%	6 32%
Geometry (*N*=22)	15 68%	3 14%	4 18%
Measurement (*N*=27)	26 96%	0 0%	1 4%
Number properties (*N*=61)	55 90%	2 3%	4 7%
Total (*N*=150)	118 79%	14 9%	18 12%

NOTE: See the methodology section for an explanation of how panelists' judgments were summarized to classify items. Percentages may not sum to 100 due to rounding.

Table 3 shows the number of grade 4 items judged to assess content that is in the CCSS at higher grade levels versus those judged to assess content not covered by the CCSS at all. In all content areas except measurement (which only had one item in this set), the number of items that were assigned to standards at higher grade levels was much greater than the number judged not to match the CCSS at all.

Table 3. Numbers of items judged to match CCSS standards at a higher grade and items judged not to match any CCSS standards, grade 4

Content area	Matched by the CCSS at a higher grade	Not matched by the CCSS
Algebra	7	1
Data analysis	9	1
Geometry	7	0
Measurement	0	1
Number properties	5	1
Total	28	4

NOTE: Based on modal response from panelists. Numbers in cells represent number of items. Includes items classified as "not in standards at or below grade level tested" and those classified as "uncertain."

Together, the results shown in Tables 2 and 3 are consistent with the earlier NVS study of alignment between the NAEP framework and the CCSS (Hughes, Daro, Holtzman, & Middleton, 2013), which found that:

> The clearest difference between the NAEP Grade 4 framework and the CCSS-M is in Data Analysis, Statistics, and Probability. The NAEP framework has substantially more emphasis on data and probability by Grade 4 than do the CCSS-M. … The CCSS-M concentrate data and probability in fewer and later grades (particularly in Grade 7) than does the NAEP framework (p. 50).

RQ 2: *To what extent do the CCSS target for instruction, at or below the grade level tested by NAEP, content that is assessed by NAEP? What is the alignment profile of the CCSS to NAEP across CCSS domains?*

RQ 2 was operationalized as the percentage (and allocation across domains) of CCSS standards from grades 3 and 4 that panelists associated with at least one NAEP item. (See the methodology section for details on how these percentages were computed.)

As shown in Table 4, there are 55 CCSS standards in grades 3 and 4. Of these, 41 (or 77%) had at least one item linked to them by the panelists. Conversely, 23% of the CCSS standards had no 2015 grade 4 NAEP items linked to them.[11] (Matching to K–2 standards is shown in Appendix E.)

[11] The percentage of CCSS standards matched by NAEP is about the same for grades 3 and 4 (76% for grade 3 and 79% for grade 4).

Table 4. Matching of grade 3 and 4 CCSS standards to the 2015 NAEP item pool at grade 4

CCSS domain	Number of grade 3 and 4 standards	Number (percentage) of standards with at least one matched item
Geometry	5	5 (100%)
Measurement and data	15	12 (80%)
Number and operations in base ten	9	8 (89%)
Fractions	10	8 (80%)
Operations and algebraic thinking	14	8 (57%)
Total	53	41 (77%)

The biggest gap between the CCSS and NAEP lies in the operations and algebraic thinking domain, where 6 of the 14 standards have no matching NAEP items. This CCSS domain contains the buildup within elementary arithmetic toward algebra. It goes beyond executing calculations to formulating numerical expressions that correspond to problem situations. From the CCSS point of view, this domain is the centerpiece of coherence in the progression from arithmetic to algebra; it attends to the mathematical structure of problems and the mathematical expression (number equations) of these structures.

In summary, the overall agreement between NAEP and the CCSS is reasonable and also fairly symmetrical at grade 4: 79% of NAEP items are clearly matched by the CCSS at or below grade 4 (RQ 1), and 77% of grade 3 and 4 CCSS standards are matched by at least one NAEP item (RQ 2). However, the areas of divergence are different between the two analyses: The fewest NAEP items assessing content covered by the CCSS are in the content area of data analysis, statistics, and probability; the fewest CCSS standards matched by NAEP are in the operations and algebraic thinking domain.

Grade 8

RQ 1: *To what extent does the NAEP item pool include content that is targeted by the CCSS for instruction at or below the grade level tested by NAEP? What is the alignment profile of NAEP to the CCSS across NAEP content areas?*

For eighth grade, 87% of NAEP items were judged to assess mathematics that is included in the CCSS at or below grade 8. Seven percent of the NAEP items were judged to assess something that was either not found in the CCSS or was found in high school, while an additional 5% of items were classified as “uncertain” due to substantial disagreement among panelists as to whether the item mapped to a standard at an appropriate grade level. The overall alignment at grade 8 is therefore strong.

As with grade 4, almost all measurement and number properties and operations items assessed content that is found in the CCSS at or below grade 8 (Table 5); there is less agreement for algebra; data analysis, statistics, and probability; and geometry. However, the differences are modest, even in these content areas. Again, as with grade 4, the content area of data analysis, statistics, and probability had the greatest proportion of items that were judged to assess content not covered in the CCSS by the grade level tested.

Table 5. Number and percentage of grade 8 NAEP items matching the CCSS at or below grade 8

Content area	In stds ≤ NAEP grade	Uncertain	Not in stds ≤ NAEP grade
Algebra (*N*=45)	38 84%	2 4%	5 11%
Data analysis (*N*=23)	17 74%	2 9%	4 17%
Geometry (*N*=25)	20 80%	3 12%	2 8%
Measurement (*N*=26)	26 100%	0 0%	0 0%
Number properties (*N*=31)	30 97%	1 3%	0 0%
Total (*N*=150)	131 87%	8 5%	11 7%

NOTE: See the methodology section for an explanation of how panelists' judgments were summarized to classify items. Percentages may not sum to 100 due to rounding.

Table 6 shows the number of grade 8 items judged to assess content that is in the CCSS standards for high school versus those judged to assess content not covered by the CCSS. In four of the five content areas, almost all of the items that were not rated as being in the CCSS at or below the grade level tested by NAEP were matched to standards at a higher grade level. The exception is the data analysis, statistics, and probability content area, in which five of the six items that were not matched to grade-appropriate standards were not found anywhere in the CCSS. This was associated with the use of data representations (such as stem and leaf) that are intentionally excluded from the CCSS.

Table 6. Numbers of items judged to match CCSS standards at a higher grade and items judged not to match any CCSS standards, grade 8

Content area	Matched by the CCSS at a higher grade	Not matched by the CCSS
Algebra	6	0
Data analysis	1	5
Geometry	4	1
Measurement	0	0
Number properties	0	0
Total	11	6

NOTE: Based on modal response from panelists. Numbers in cells represent number of items.
For two additional items (one in algebra and one in number properties and operations), panelists were evenly split between classifying the item as associated with a standard at a higher grade level or not in the standards at all.

RQ 2: *To what extent do the CCSS target for instruction, at or below the grade level tested by NAEP, content that is assessed by NAEP? What is the alignment profile of the CCSS to NAEP across CCSS domains?*

RQ 2 was operationalized as the percentage (and allocation across domains) of CCSS standards from grades 6, 7, and 8 that panelists associated with at least one NAEP item.

As shown in Table 7, there are 81 standards in grades 6, 7, and 8. Of these, 47 (58%) had at least one 2015 grade 8 NAEP item linked to them by the panelists. Conversely, 42% of the CCSS standards had no NAEP items linked to them.[12] (Matching to K–5 standards is shown in Appendix E.)

[12] Percentages of standards linked to at least one NAEP item were similar across the three grades: 59% for grade 6, 54% for grade 7, and 61% for grade 8.

Table 7: Matching of grade 6, 7, and 8 CCSS standards to the 2015 NAEP item pool at grade 8

CCSS domain	Number of grade 6, 7, and 8 standards	Number (percentage) of standards with at least one matched item
Expressions and equations	21	13 (62%)
Functions	5	3 (60%)
Geometry	19	14 (74%)
Ratios and proportional relationships	6	3 (50%)
Statistics and probability	17	7 (41%)
The number system	13	7 (54%)
Total	81	47 (58%)

Geometry had the highest percentage of standards matched by at least one item in the 2015 NAEP item pool (74%) despite the fact that the CCSS made a major shift in geometry toward emphasis on transformations. A review of the NAEP framework, however, shows that NAEP made the same shift more than a decade ago.

In summary, although 87% of grade 8 NAEP items assessed content that is found in the CCSS at or below grade 8 (RQ 1), 58% of CCSS standards were linked with at least one grade 8 NAEP item (RQ 2). There appears to be a notable amount of middle-school content recommended by the CCSS that is not part of the current NAEP assessment.

Overall Summary and Conclusion

Overall, the review by expert panelists suggests that concordance between NAEP and the CCSS is reasonable at both grade levels. It makes sense that NAEP, which is required by its mission to be broad, will include some items that are outside the bounds of the CCSS and will not assess every standard in the CCSS.

However, the agreement between NAEP and the CCSS is uneven across the five NAEP content areas, and, at grade 8, 42% of middle-grade CCSS standards are not assessed by any items in the 2015 NAEP item pool. Highlights of the specific findings for each grade level follow.

Grade 4

Although the overall alignment between NAEP and the CCSS is reasonable at grade 4, with 79% of NAEP items clearly matched to the CCSS standards at or below grade 4 and 77% of grade 3 and 4 CCSS standards assessed by at least one NAEP item, there are substantial differences in some content areas, particularly areas where NAEP assesses content that the CCSS has shifted to later grade levels.

One of the more notable differences between NAEP and the CCSS at grade 4 is the two-way divergence in the way *algebra* is viewed in the elementary grades. Only 62% of NAEP grade 4 algebra items were associated with a standard in the CCSS at or below grade 4; only 57% of grade 3 and 4 CCSS standards within the operations and algebraic thinking domain had at least one NAEP item associated with them.

By contrast, the differences in *geometry* were mostly in one direction: Only 68% of NAEP geometry items matched the CCSS at or below grade 4, but 100% of grade 3 and 4 CCSS geometry standards had at least one item matched to them.

Another asymmetrical divergence is seen in *data analysis, statistics, and probability*. This content area had the lowest percentage of items (42%) matched to a standard in the CCSS at or below grade 4. However, 80% of the CCSS measurement and data standards in grades 3 and 4 had at least one NAEP item linked to them.

All the unmatched grade 4 items in geometry, and 90% of those in data analysis, statistics, and probability, were matched to standards at a higher grade level in the CCSS.

Grade 8

The overall match of grade 8 NAEP items to the CCSS at or below grade 8 (87%) is strong. Measurement (100%) and number properties and operations (97%) are highly concordant. Algebra (84%); data analysis, statistics, and probability (74%); and geometry (80%) are somewhat more divergent but still reasonable.

Looking from the perspective of RQ 2, another picture of grade 8 alignment comes into view: 42% of the grade 6, 7, and 8 CCSS have no NAEP items linked to them. Considering the CCSS intent to focus more deeply on fewer topics, the percentage of standards with no clear link to a NAEP item was unexpected.

Given the two-way discrepancies that our panelists found between NAEP and the CCSS in algebra in the elementary grades, it is worth looking more closely at algebra as it appears in the CCSS at grades 6, 7, and 8. An excerpt of Table 7 is shown in Table 8 below. The domains excerpted match the middle-grade content associated with "algebra." Forty-one percent of the CCSS algebra content (13 of 32 standards) has no matching items in the 2015 NAEP item pool. This is consistent with the extent of matching seen across all the middle-grade standards.

Table 8. Matching of grade 6, 7, and 8 "algebra" standards to the 2015 NAEP item pool at grade 8

CCSS domain	Number of grade 6, 7, and 8 standards	Number (percentage) of standards with a matched item
Expressions and equations	21	13 (62%)
Functions	5	3 (60%)
Ratios and proportional relationships	6	3 (50%)
Total	32	19 (59%)

Standards for Mathematical Practice

Another important aspect of the CCSS is the eight standards for mathematical practice that are deliberately threaded throughout the CCSS content standards. This limited study made no attempt to determine if NAEP assesses the practice standards. Items can be rated against goals like the practices, but it takes more training and rater calibration than this study allowed.[13] We hope to include such an analysis when we compare both NAEP and CCSS-consortium items with one another and with their respective frameworks/standards.

Conclusion

The mission of NAEP is to provide a valid and reliable measure of progress for the nation over time, and NAEP results make the most sense over the longer timelines for which trend assessments are designed. Viewing results over a number of years establishes an important context for interpreting the present and for looking ahead.

Prioritizing trend means that NAEP must maintain a degree of independence from what might be seen as current fashions in instruction and curriculum. At the same time, NAEP must periodically review and update its frameworks because it cannot provide useful results if the content that it measures is too far from the content of contemporary instruction.

[13] Examples where such work has been done include analyses of the Program for International Student Assessment (PISA), depth of knowledge analysis, and NAEP's own item complexity ratings (Herman et al., in development; Webb, 2007; National Assessment Governing Board, 2010).

Given the historical variation in instructional standards across states, NAEP has sought to represent the union of all major curricula when developing and revising its frameworks. It may be that the landscape is changing now for mathematics and ELA, with many states having adopted either the CCSS or other college and career readiness standards. However, changing standards is always far simpler and faster than changing curriculum and instruction, and both time and research are needed to see how the CCSS actually play out in U.S. classrooms. At the same time, the immediate salience of the CCSS is raised by the fact that many states are right now basing accountability on CCSS-aligned state assessments.

It has been 10 years since NAEP conducted a major review of its mathematics framework. Despite some uncertainties concerning the way that the CCSS will influence instruction over time, we believe that this is an appropriate moment for NAEP's Governing Board to review the framework in light of the CCSS as well as other states' college and career standards. The findings from this study suggest particular areas where such a review should focus.

References

Common Core State Standards Initiative. (2010). *Common Core State Standards for mathematics.* Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Heitin, L. (2015, July 7). Common-Core materials penetrate every state. *Education Week.* Retrieved from http://www.edweek.org/ew/articles/2015/07/08/common-core-materials-penetrate-every-state.html?tkn=VLVFT77bcIGGzdl8aKcrAr2HjOW6pFN%2B4KGp&intc=es

Herman et al. (in development). *Benchmarks for deeper learning on next generation tests: A study of PISA.* Los Angeles, CA: CRESST.

Hughes, G. B., Daro, P., Holtzman, D., & Middleton, K. (2013). A study of the alignment between the NAEP mathematics framework and the Common Core State Standards for mathematics (CCSS-M). In F. B. Stancavage and G. W. Bohrnstedt (Eds.), *Examining the content and context of the Common Core State Standards: A first look at implications for the National Assessment of Educational Progress* (pp. 9–86). San Mateo, CA: American Institutes for Research.

National Assessment Governing Board. (2010). *Mathematics framework for the 2011 National Assessment of Educational Progress.* Washington, DC: Author.

Santelises, S. B., & Dabrowski, J. (2015). *Checking in: Do classroom assignments reflect today's higher standards?* Washington, DC: The Education Trust.

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Measurement in Education, 20*(1), 7–25. Retrieved from http://www.cehd.umn.edu/edpsych/C-BAS-R/Docs/Webb2007.pdf

Appendix A: Expert Panelists

Harold Asturias	Director, Center for Mathematics Excellence and Equity at Lawrence Hall of Science, University of California, Berkeley
Chris Avila	Coordinator, Mathematics Programs, State Department of Education, Idaho
Hy Bass	Professor, School of Education and Department of Mathematics, University of Michigan
Wade Ellis	Senior Mathematics Advisor, Texas Instruments, Educational Technology Division
Brad Findell	Associate Director of Mathematics Programs for Teachers, Ohio State University
Larry Gray	Professor, School of Mathematics, University of Minnesota
Jill Griffin	Urban Education Consultant (Curriculum & Instruction), State Department of Education, Michigan
Anton Jackson	Mathematics Assessment Specialist, Office of the Superintendent of Public Instruction, Washington
Diana Kasbaum	Mathematics Education Consultant, Washington Department of Public Instruction and Association of State Supervisors of Mathematics
Deena Khalil	Assistant Professor of Mathematics Education, Howard University
Nicole Kim	Middle and High School Mathematics Teacher, William S. Hart Union High School District, California
Crystal Lancour	Supervisor of Curriculum and Instruction for K-12 Mathematics, Colonial School District, Delaware
Life LeGeros	Professional Development Coordinator, Tarrant Institute for Innovative Education, University of Vermont
Jane Porath	Middle School Mathematics Teacher, Traverse City Area Public Schools, Michigan
Jim Ryan	STEM Executive Director, San Francisco Unified School District, California
Mary Jo Tavormina	Visiting Project Director, University of Illinois at Chicago
Denise Walston	Director of Mathematics, The Council of the Great City Schools, Washington, DC
Tad Watanabe	Professor of Mathematics Education, Kennesaw State University, Georgia

Appendix B: Directions for Coding NAEP Items to CCSS Mathematics Standards

Overview

The focus of our workshop is on providing the item classification data that will allow us to address the study's two research questions. Our task is not to provide an evaluation of NAEP or the CCSS. You will, no doubt, find things worthy of criticism in both NAEP and the CCSS. We have a very tight schedule and a lot of items to classify, so please don't take time to contribute these critiques during coding unless they facilitate completion of the task. There will be time on Thursday when we ask you to discuss and put into writing your overall analysis of how NAEP items relate to the CCSS.

NAEP Items and Codes

1. NAEP has five subscales: number properties and operations; measurement; geometry; data analysis, statistics, and probability; and algebra. The copies of NAEP items that you work with will have information identifying the subscale to which each item belongs. Also provided will be the answer key (if multiple choice) or the scoring guide (if constructed response), as well as an indication as to whether the item appeared in a block in which students were allowed to use calculators or one for which a tool (such as a ruler) was provided.[14]
2. Each item has a unique identifying code in the upper right-hand corner. These codes take the form of 8.N.12, where 8 represents the grade level, N represents the NAEP subscale, and 12 represents an (arbitrary) sequential order of the item within grade and subscale.

The Procedure

1. After an initial round of practice, panelists for each grade level will work separately to classify items for that grade level.
2. Working with batches of items that we will assign, work independently to classify each item in the 2015 NAEP item pool for your grade level (approximately 150 items in total).
3. After each batch, we will conduct a group resolution discussion. You will then have the opportunity to change your original classification, if desired.
    a. In order to identify items needing resolution: For each item, we will ask for a "show of thumbs" to determine whether there is a critical level of disagreement. Each panelist will hold thumbs up for items that they have coded "yes" in column G (item in standard at appropriate grade level?) and thumbs down for items coded "no." **This is the critical distinction needed to answer the main study questions.**
    b. If three or more judges disagree with the majority on this code for a given item, we will conduct a resolution discussion, alternating panelists from the minority and the

[14] Copies of the NAEP mathematics framework will be available at the meeting, but they are not needed for the work of the panel. The panel will focus on classifying each item into the Common Core State Standards, and the NAEP coding is provided solely for your information and convenience.

majority as they explain their reasons. During this process, it will be important to review the relevant guidance for coding items, as shown below.

   c. Revise coding as desired.

4. As time permits, we also will discuss differences in opinion in classifying items at the grade-domain-cluster level. (That is, we won't drill down to differences at the standard level within cluster.)

Guidance: Making Judgments

1. Begin by working an item, noting the mathematical demands. Then search the CCSS for a standard that targets the mathematical demands for getting the item correct (multiple choice or dichotomous CR items) or getting full credit on the item (partial credit CR items).
2. Start at the grade level where the item was administered by NAEP (grade 4 or 8). If no relevant standard is found in the CCSS at that grade level, search one grade level lower, then one higher, then lower and higher as makes sense.
   a. Note that as you proceed to lower grades, the referent of words such as "multiply" can change, referring to numbers with fewer digits in simpler problems. For fourth-grade NAEP, search no lower than grade 2.
3. Choosing among standards:
   a. Many items involve mathematics from several grade levels. Identify the CCSS at the grade level that matches the most demanding mathematics needed for the task. For example, if the task requires subtraction from grade 2 and division from grade 5, choose the division standard at grade 5.
   b. If an item requires skills or knowledge associated with more than one standard at a given grade level, choose the standard associated with the **most demanding skill or knowledge** required in order to answer the item correctly.
   c. If an item demands a term or representation that appears at a higher grade level in the CCSS, then classify the item to the standard at that higher grade level.
      i. The grade level where specific geometric terms are introduced can, for example, be somewhat arbitrary (e.g., the term "mid-point" does not occur in the CCSS until high school).
   d. The CCSS set limits on the size of numbers involved in calculations in the number and base ten domain (NBT) at particular grade levels; items should be coded to a grade level that is inclusive of the number range used in the item.
4. Many mathematics items in grades 4 and 8 can be answered correctly using clever strategies that work around the mathematics being assessed. For the purposes of this study, do not consider such cleverness as part of your judgment. To do so would open too wide a door. Focus on the intended mathematical target of the item.
5. You should not identify standards as a match when the item is off target. You may classify an item as not matching well to any standard. **This is a perfectly acceptable choice**—do not force an item to fit into a standard. On the other hand, you should not conclude "not targeted by CCSS" merely because the fit is imperfect. A degree of professional common sense will be needed to make these judgments.
   a. If the **most demanding skill or knowledge** required to answer the item correctly does not match any standard in the CCSS, then classify the item as not matching, even if incidental knowledge needed matches a standard.

6. The CCSS explicitly aim to focus the curriculum on fewer topics to remedy the "mile-wide, inch deep" character of American curricula. Therefore, omissions from traditional U.S. curricula in the CCSS are to be interpreted as deliberate.
   a. If a NAEP item demands comprehension of a mathematical term or representation not targeted in the CCSS, then the item does not match the CCSS.
   b. If an item requires comprehension of a particular type of chart not required by the CCSS, then it is not in the CCSS. For example, pie charts and stem and leaf graphs are not included in the CCSS at any grade level.

Coding

1. Use the coding sheet provided to record your classifications. The sheet is prepopulated with the item-identifying codes.
2. Enter the CCSS grade level, domain, cluster, and standard for each item.[15] Alternatively, mark the box that says that the item is not matched with any standard.
   a. Under grade level, include the topic area for high school as is done in the standards themselves—for example, HSS for high school statistics and probability.
   b. For the domain, do not repeat the grade-level code—for example, only enter OA for operations and algebraic thinking, at whatever grade level.
   c. For cluster, use the capital letters that appear in the online version of the standards.
   d. For the standard, use the Arabic number that appears in the standards document. Note that from K through 8, the numbering of standards is sequential within domain; in high school, the numbers start over for each cluster.
   e. We will **not** code at the level of substandards.
   f. In the next-to-last column on the coding sheet (labeled "Item in standard at appropriate grade level?"), mark your summary judgment (Y or N) as to whether or not the item can be classified into a standard at the NAEP assessment grade level (4 or 8) or below.
   g. The last column provides space for you to enter a brief comment, if needed.
3. For each item, use the first (unshaded) line to enter your codes. If you decide to change your codes after the group discussion, enter the new codes on the second (shaded) line.
4. **Be sure to proofread your codes**—It will be very hard for us to identify any errors in your codes. You will be reminded to do this after each coding block, before we gather to discuss the codes in that block.

[15] For convenience, you have been provided with a reference document ("Summary of CCSS math grade level, domain, and cluster codes") that lists all the domains and clusters by grade level and shows the correct codes to enter for each.

Appendix C: Agreement Among Panelists on Selected Variables

Agreement Among Panelists at the Standard Level

During the classification process, individual panel members were only allowed to assign one standard to an item (the one that they felt was the best fit). Therefore, the greater the number of different standards assigned to an item, the lower the agreement—at the standard level—among the panelists who classified the item. Tables C-1 and C-2 show counts of items in each NAEP content area that were assigned one standard, two standards, three standards, or four or more standards.

For example, in algebra at grade 4, there were four items where all panelists agreed on a single standard. There were five items where panelists were split between two standards, three items for which there was a three-way split, and one item for which the pool of panelists classifying the item was divided among four or more standards.

Table C-1. Number of items for which different numbers of standards were assigned across panelists, grade 4

	Number of different standards assigned to the same item among the pool of panelists classifying the item			
Content area	**1 standard**	**2 standards**	**3 standards**	**≤4 standards**
Algebra	4	5	3	1
Data analysis	1	3	2	3
Geometry	2	5	4	4
Measurement	1	10	6	9
Number properties	9	21	20	5
Total	17	44	35	22

NOTE: Numbers in the table cells represent the number of items. Analysis restricted to items classified as being in the CCSS at or below the grade tested by NAEP.

Table C-2. Number of items for which different numbers of standards were assigned across panelists, grade 8

Content area	Number of different standards assigned to the same item among the pool of panelists classifying the item			
	1 standard	2 standards	3 standards	≤4 standards
Algebra	3	14	12	9
Data analysis	2	6	3	6
Geometry	3	4	3	10
Measurement	0	4	4	18
Number properties	5	13	8	4
Total	13	41	30	47

NOTE: Numbers in the table cells represent the numbers of items. Analysis restricted to items classified as being in the CCSS at or below the grade tested by NAEP.

Agreement Among Panelists on Disposition of Items Not Matched by the CCSS at or Below the Grade Tested by NAEP

Another area of potential disagreement among panelists judging the same item concerned items that were not classified as matched by the CCSS at or below the grade tested. In these cases, panelists could indicate that the item was matched by a CCSS standard at a higher grade level or that the item was not matched by the CCSS at all.

Table C-3 shows that, for grade 4, panelists agreed on the disposition of these items more than two thirds of the time. For example, in algebra, six items were rated by all panelists as matched to a higher grade level, and one item was rated by all panelists as not found in the CCSS. Panelists' judgments were split on one item.

Table C-4 shows that, for grade 8, items were about equally split between those on which all panelists were in agreement and those on which the panelists were split.

Table C-3. Variation in panelists’ judgments of items as matched by the CCSS at a higher grade versus not matched at all, grade 4

Content area	All panelists rated item as matched to a higher grade	All panelists rated item as not found in CCSS	Some rated as higher grade and some as not found
Algebra	6	1	1
Data analysis	7	0	3
Geometry	2	0	5
Measurement	0	0	1
Number properties	5	1	0
Total	20	2	10

NOTE: Numbers in the table cells represent the number of items. Analysis restricted to items other than those classified as being in the CCSS at or below the grade tested by NAEP.

Table C-4. Variation in panelists’ judgments of items as matched by the CCSS at a higher grade versus not matched at all, grade 8

Content area	All panelists rated item as matched to a higher grade	All panelists rated item as not found in CCSS	Some rated as higher grade and some as not found
Algebra	6	0	1
Data analysis	0	3	3
Geometry	2	0	3
Measurement	0	0	0
Number properties	0	0	1
Total	8	3	8

NOTE: Numbers in the table cells represent the number of items. Analysis restricted to items other than those classified as being in the CCSS at or below the grade tested by NAEP.

Appendix D: Application of Rules for Classifying Items as Matched by the CCSS at or Below the Grade Tested by NAEP

Tables D-1 and D-2 show how the rules for classifying items into three groups—in the CCSS at or below the grade level tested by NAEP, not in the CCSS at or below the grade level tested by NAEP, or uncertain—played out across the pool of items analyzed in this study.

For example, as shown in the top panel of Table D-1, with four panelists rating the item, there were two patterns of responses that would result in the item being classified as not in the CCSS at or below the grade level tested by NAEP: zero panelists rating the item as "in" and four panelists rating it as "out," or one panelist rating the item as "in" and three panelists rating it as "out." At grade 4, there were five items that fit the first pattern and none that fit the second. There was only a single pattern that would result in an item with four panelists being classified as uncertain: two panelists rating the item as "in" and two rating it as "out." At grade 4, one item fit this pattern. Finally, for groups of four panelists, there were two patterns that would result in the item being classified as in the CCSS at or below the grade level tested by NAEP: three panelists rating the item as "in" and one panelist rating it as "out," or four panelists rating the item as "in" and zero panelists rating it as "out." At grade 4, there were three items that fit the first pattern and 26 that fit the second.

Table D-1. Application of rules for classifying items as matched by the CCSS at or below the grade tested by NAEP, grade 4

Total number of panelists	Number of panelists rating as "in"	Number of panelists rating as "out"	Item classification	Number of items fitting this pattern
4	0	4	Not in	5
4	1	3	Not in	0
4	2	2	Uncertain	1
4	3	1	In	3
4	4	0	In	26
5	0	5	Not in	2
5	1	4	Not in	0
5	2	3	Uncertain	3
5	3	2	Uncertain	0
5	4	1	In	2
5	5	0	In	32
7	0	7	Not in	0
7	1	6	Not in	0
7	2	5	Not in	0
7	3	4	Uncertain	0
7	4	3	Uncertain	1

Total number of panelists	Number of panelists rating as "in"	Number of panelists rating as "out"	Item classification	Number of items fitting this pattern
7	5	2	In	0
7	6	1	In	0
7	7	0	In	1
8	0	8	Not in	0
8	1	7	Not in	0
8	2	6	Not in	0
8	3	5	Uncertain	0
8	4	4	Uncertain	0
8	5	3	Uncertain	0
8	6	2	In	0
8	7	1	In	1
8	8	0	In	0
9	0	9	Not in	6
9	1	8	Not in	4
9	2	7	Not in	1
9	3	6	Uncertain	2
9	4	5	Uncertain	1
9	5	4	Uncertain	2
9	6	3	Uncertain	4
9	7	2	In	5
9	8	1	In	13
9	9	0	In	35

Table D-2. Application of rules for classifying items as matched by the CCSS at or below the grade tested by NAEP, grade 8

Total number of panelists	Number of panelists rating as "in"	Number of panelists rating as "out"	Item classification	Number of items fitting this pattern
4	0	4	Not in	2
4	1	3	Not in	2
4	2	2	Uncertain	1
4	4	0	In	34
5	0	5	Not in	1
5	1	4	Not in	0
5	2	3	Uncertain	1
5	3	2	Uncertain	0
5	4	1	In	3
5	5	0	In	26
8	0	8	Not in	0
8	1	7	Not in	0
8	2	6	Not in	0
8	3	5	Uncertain	0
8	4	4	Uncertain	0
8	5	3	Uncertain	0
8	6	2	In	1
8	7	1	In	3
8	8	0	In	4
9	0	9	Not in	4
9	1	8	Not in	1
9	2	7	Not in	1
9	3	6	Uncertain	1
9	4	5	Uncertain	1
9	5	4	Uncertain	3
9	6	3	Uncertain	1
9	7	2	In	7
9	8	1	In	18
9	9	0	In	35

Appendix E: Matching of CCSS Standards at or Below the Grade Tested by NAEP

Table E-1. Matching of CCSS standards at or below the grade tested by NAEP, grade 4

CCSS domain, by grade level	Number of standards	Number of standards with at least one matched item
GR K -Counting And Cardinality	7	0
_ Total Counting And Cardinality	7	0
GR K –Geometry	6	1
GR 1 –Geometry	3	0
GR 2 -Geometry	3	0
GR 3 -Geometry	2	2
GR 4 -Geometry	3	3
_ Total Geometry	17	6
GR K -Measurement And Data	3	0
GR 1 -Measurement And Data	4	1
GR 2 -Measurement And Data	10	4
GR 3 -Measurement And Data	8	8
GR 4 -Measurement And Data	7	4
_ Total Measurement And Data	32	17
GR K -Number And Operations In Base Ten	1	0
GR 1 -Number And Operations In Base Ten	6	0
GR 2 -Number And Operations In Base Ten	9	0
GR 3 -Number And Operations In Base Ten	3	2
GR 4 -Number And Operations In Base Ten	6	6
_ Total Number And Operations In Base Ten	25	8
GR 3 -Number And Operations/Fractions	3	3
GR 4 -Number And Operations/Fractions	7	5
_ Total Number And Operations/Fractions	10	8
GR K -Operations And Algebraic Thinking	5	0
GR 1 -Operations And Algebraic Thinking	8	0
GR 2 -Operations And Algebraic Thinking	4	0
GR 3 -Operations And Algebraic Thinking	9	4
GR 4 -Operations And Algebraic Thinking	5	4
_ Total Operations And Algebraic Thinking	31	8
Overall Total	122	47

Table E-2. Matching of CCSS standards at or below the grade tested by NAEP, grade 8

CCSS domain, by grade level	Number of standards	Number of standards with at least one matched item
GR K -Counting And Cardinality	7	0
_ Total Counting And Cardinality	7	0
GR 6 -Expressions And Equations	9	5
GR 7 -Expressions And Equations	4	2
GR 8 -Expressions And Equations	8	6
_ Total Expressions And Equations	21	13
GR 8 -Functions	5	3
_ Total Functions	5	3
GR K -Geometry	6	0
GR 1 -Geometry	3	0
GR 2 -Geometry	3	0
GR 3 -Geometry	2	0
GR 4 -Geometry	3	0
GR 5 -Geometry	4	1
GR 6 -Geometry	4	4
GR 7 -Geometry	6	4
GR 8 -Geometry	9	6
_ Total Geometry	40	15
GR K -Measurement And Data	3	0
GR 1 -Measurement And Data	4	0
GR 2 -Measurement And Data	10	0
GR 3 -Measurement And Data	8	1
GR 4 -Measurement And Data	7	2
GR 5 -Measurement And Data	5	1
_ Total Measurement And Data	37	4
GR K -Number And Operations In Base Ten	1	0
GR 1 -Number And Operations In Base Ten	6	0
GR 2 -Number And Operations In Base Ten	9	0
GR 3 -Number And Operations In Base Ten	3	0
GR 4 -Number And Operations In Base Ten	6	2
GR 5 -Number And Operations In Base Ten	7	2
_ Total Number And Operations In Base Ten	32	4

CCSS domain, by grade level	Number of standards	Number of standards with at least one matched item
GR 3 -Number And Operations/Fractions	3	0
GR 4 -Number And Operations/Fractions	7	0
GR 5 -Number And Operations/Fractions	7	3
_ Total Number And Operations/Fractions	17	3
GR K -Operations And Algebraic Thinking	5	0
GR 1 -Operations And Algebraic Thinking	8	0
GR 2 -Operations And Algebraic Thinking	4	0
GR 3 -Operations And Algebraic Thinking	9	0
GR 4 -Operations And Algebraic Thinking	5	1
GR 5 -Operations And Algebraic Thinking	3	0
_ Total Operations And Algebraic Thinking	34	1
GR 6 -Ratios And Proportional Relationships	3	1
GR 7 -Ratios And Proportional Relationships	3	2
_ Total Ratios And Proportional Relationships	6	3
GR 6 -Statistics And Probability	5	2
GR 7 -Statistics And Probability	8	4
GR 8 -Statistics And Probability	4	1
_ Total Statistics And Probability	17	7
GR 6 -The Number System	8	5
GR 7 -The Number System	3	1
GR 8 -The Number System	2	1
_ Total The Number System	13	7
Overall Total	229	60