The Alignment of the NAEP Grade 12 Mathematics Assessment and the ACT Mathematics Test

Based on the Program of Preparedness Research adopted by the National Assessment Governing Board, research is being conducted to study the relationship between grade 12 NAEP and other indicators of preparedness for college and job training. Five categories of research are currently included in this research program: content alignment studies, statistical relationship studies, judgmental standard-setting studies, surveys, and benchmarking studies.

This study brief focuses on one of the content alignment studies: The ACT Mathematics Test and the NAEP 12th Grade Mathematics Assessment. This content alignment study was designed and implemented by ACT.

Type of Study
Content Alignment

Purpose
Identify and describe content overlap and non-overlap between NAEP and the ACT Mathematics Test

Project Period
September 2008 to November 2009

Contractor and Principal Investigator
ACT, Nancy Petersen and Jay Happel

Background on the ACT Mathematics Test and the NAEP 12th Grade Mathematics Assessment
The ACT assessment is a widely recognized set of standardized tests used for college admissions. It yields individual student results; it does not produce nationally representative results for any defined group. Test items are all multiple choice. The content domain of the ACT Mathematics Test is composed of six broad subcategories: Pre-Algebra, Elementary Algebra, Intermediate Algebra, Coordinate Geometry, Plane Geometry, and Trigonometry. Cognitive skills measured include: Knowledge and Skills (50% of questions); Direct Application (28%); Understanding Concepts (22%, with Integrating); and Integrating Conceptual Understanding (22%, with Understanding).

The NAEP 12th grade mathematics assessment produces reports to the public on the achievement of 12th graders. Results are nationally representative,[1] and, as required by law, NAEP does not produce individual student results. Test item types include multiple choice and both short and extended constructed response, and involve both real-world contexts and pure mathematics. The assessment is comprised of five content areas: Number Properties and Operations (10%), Measurement and Geometry (30%), Data Analysis, Statistics, and Probability (25%), and Algebra (35%). There are three levels of item complexity: Low— recall or recognize concepts or procedures, carry out specified procedures (25%); Moderate— think flexibly in solving problems; make connections among concepts and processes from various domains (50%); and High— use reasoning, planning, analysis, judgment, and creative thought in solving problems (25%). Items from the 2009 12th grade mathematics assessment were reviewed for this study.

Methodology
The study had two major components. The first was a detailed comparison of the NAEP 12th grade mathematics framework with the ACT content and cognitive domains, test specifications, and the ACT College Readiness Standards in Mathematics. The second, called the Item

[1] In 2009, 11 states participated in a pilot to obtain state-representative NAEP 12th grade results.

Classification Study, attempted to categorize each item in the 2009 NAEP assessment in mathematics according to the ACT College Readiness Standards score ranges. The data for both studies were collected through the work of a panel of subject matter experts in mathematics. For each subject area, ACT convened a seven-member panel of educators, some at the high school level and some at the postsecondary level, some with prior experience with the NAEP and some with experience with the ACT. ACT staff developed the procedures for the studies, trained the panel members, and facilitated the panel meeting.

Key Conclusions

The panel found considerable overlap in what the ACT and the NAEP assess in mathematics:

- The elements of the ACT Mathematics Test content domain were reflected well in the NAEP domain, although a number of lower-level ACT topics were seen as addressing topics from the Grade 8, not the Grade 12, NAEP.
- The panel was able to find elements of the NAEP domain for nearly all of the ACT College Readiness Standards in Mathematics.
- All of the skills represented in the ACT College Readiness Standards are measured on the NAEP Mathematics assessment.

The panel also found differences in what NAEP and the ACT assess:

- In the NAEP-to-ACT comparison, the difference in specificity with which the domains are articulated in the assessment documents left the panel uncertain as to whether a number of NAEP content topics—those pertaining to transformations, probability, statistics, and data analysis—are assessed by the ACT.
- In the NAEP-to-ACT comparison, there was some uncertainty within the panel on the degree to which higher-order analytic skills were assessed, and it was the sense of the panel that the ACT Mathematics Test contained few items involving high mathematical complexity, at least as the NAEP defines it.

While the NAEP Mathematics framework and the ACT Mathematics domain, test specifications, and College Readiness Standards share similarities, important differences in what and how the assessments measure suggest caution when drawing comparisons between the assessments.

The Alignment of the NAEP Grade 12 Reading Assessment and the ACT Reading Test

Based on the Program of Preparedness Research adopted by the National Assessment Governing Board, research is being conducted to study the relationship between grade 12 NAEP and other indicators of preparedness for college and job training. Five categories of research are currently included in this research program: content alignment studies, statistical relationship studies, judgmental standard-setting studies, surveys, and benchmarking studies.

This study brief focuses on one of the content alignment studies: The ACT Reading Test and the NAEP 12th Grade Reading Assessment. This content alignment study was designed and implemented by ACT.

Type of Study
Content Alignment

Purpose
Identify and describe content overlap and non-overlap between NAEP and the ACT Reading Test

Project Period
September 2008 to November 2009

Contractor and Principal Investigator
ACT, Nancy Petersen and Jay Happel

Background on the ACT Reading Test and the NAEP 12th Grade Reading Assessment

The ACT assessment is a widely recognized set of standardized tests used for college admissions. It yields individual student results; it does not produce nationally representative results for any defined group. Test items are all multiple choice. The types of passages in the ACT Reading Test are: Prose Fiction—passages from short stories or novels (25%); Humanities—passages from memoirs, personal essays, and nonfiction essays (25%); Social Science—nonfiction passages on topics such as anthropology, business, economics, history, political science, psychology, and sociology (25%); Natural Science—nonfiction passages in areas such as biology, chemistry, geology, and meteorology. Passage length is approximately 750 words. Cognitive targets are: (1) Referring, which includes recognizing the main idea, cause-effect and comparative relationships, and significant details; and (2) Reasoning, which includes drawing inferences from the text, demonstrating critical understanding of the text; and determining the meaning of words in context.

The NAEP 12th grade reading assessment produces reports to the public on the achievement of 12th graders. Results are nationally representative,[2] and, as required by law, NAEP does not produce individual student results. Reading passages are literary (30%) and informational (70%). Authentic texts are used that high school students could encounter. Literary texts include fiction, literary non-fiction, and poetry. Informational text includes exposition, argumentation or persuasive text, and procedural texts or documentation. Cognitive targets are (1) Locate/Recall, (2) Integrate/Interpret, and (3) Critique/Evaluate. Passage length is 500-1500 words. Passage difficulty is grade 12 appropriate. Item types include multiple-choice and both short and extended constructed response. Items reviewed were from the 2009 12th grade reading assessment, 131 in all.

Methodology

[2] In 2009, 11 states participated in a pilot to obtain state-representative NAEP 12th grade results.

The study had two major components. The first was a detailed comparison of the NAEP 12th grade reading framework with the ACT content and cognitive domains, test specifications, and the ACT College Readiness Standards in Reading. The second, called the Item Classification Study, attempted to categorize each item in the 2009 NAEP assessment in reading according to the ACT College Readiness Standards score ranges. The data for both studies were collected through the work of a panel of subject matter experts in reading. For each subject area, ACT convened a seven-member panel of educators, some at the high school level and some at the postsecondary level, some with prior experience with the NAEP and some with experience with the ACT. ACT staff developed the procedures for the studies, trained the panel members, and facilitated the panel meeting.

Key Conclusions

The panel found considerable overlap in what the ACT and the NAEP assess in reading:

- All of the ACT Reading Test cognitive skills—the skills around which ACT items are written-were reflected in the NAEP Reading domain.
- All of the skills represented in the ACT College Readiness Standards are measured on the NAEP Reading assessment.

The panel also found differences in what NAEP and the ACT assess:

- There was less evidence of the higher-level analytical and evaluative "Critique/Evaluate" skills in the ACT domain than in NAEP, and this was attributed by the panel to the multiple-choice format of the ACT, whereas NAEP employs constructed response items as well as multiple choice items.
- Another difference is that NAEP includes items and texts measuring how well an examinee can apply reading skills across texts, whereas the paired passage format in NAEP is not a feature of the ACT.

While the NAEP Reading framework and the ACT Reading domain, test specifications, and College Readiness Standards share similarities, important differences in what and how the assessments measure suggest caution when drawing comparisons between the assessments.

The Alignment of the NAEP Grade 12 Mathematics Assessment and the SAT Mathematics Test

Based on the Program of Preparedness Research adopted by the National Assessment Governing Board, research is being conducted to study the relationship between grade 12 NAEP and other indicators of preparedness for college and job training. Five categories of research are currently included in this research program: content alignment studies, statistical relationship studies, judgmental standard-setting studies, surveys, and benchmarking studies. This study brief focuses on one of the content alignment studies. This content alignment study was implemented based on a design developed by Dr. Norman Webb specifically for addressing the Board's preparedness research objective to compare the NAEP assessment with other assessments regarded as indicators of preparedness.[3]

Type of Study
Content Alignment

Purpose
Identify and describe content overlap and non-overlap between NAEP and the SAT Mathematics Assessments

Project Period
September 2009 to November 2010

Contractor and Principal Investigator
WestEd, Stuart Rabinowitz (Principal Investigator), Peter Worth (Project Director), and Norman Webb (Senior Technical Advisor).

Background on SAT Mathematics Test and the NAEP 12th Grade Mathematics Assessment

The SAT is a widely recognized standardized test used for college admissions. The College Board produces the SAT, which is now officially called the SAT Reasoning Test. SAT scores are produced for three major sections: mathematics, critical reading and writing. It yields individual student results, but does not produce nationally representative results for any defined group. The mathematics content covered on the assessment includes Numbers and Operations (20-25% of the items); Algebra and Functions (35-40%); Geometry and Measurement (25-30%); and Data Analysis, Statistics, and Probability (10-15%). Within each content category is a list of more detailed specifications. Cognitive complexity of the mathematics items in the SAT is classified as (1) Routine, (2) Comprehension, and (3) Nonroutine/Insightful. While this cognitive complexity classification of SAT items generally increases with item difficulty, there is not total correspondence. The SAT mathematics assessment is a fixed-form test comprising 54 items, and two forms were used for this study—a total of 108 unique items. Each SAT Mathematics form contains 44 standard 5-choice items and 10 Student-produced Response (SPR) items. For the SPR questions, students must arrive at numerical answers and enter them into a grid.

The NAEP 12th grade mathematics assessment produces reports to the public on the achievement of 12th graders. Results are nationally representative,[4] and, as required by law, NAEP does not produce individual student results. Test item types include multiple choice and both short and extended constructed response, and involve both real-world contexts and pure mathematics. The assessment is comprised of five content areas: Number Properties and Operations (10%), Measurement and Geometry (30%), Data Analysis, Statistics, and Probability (25%), and Algebra (35%). There are three levels of item complexity:

[3] The design document—*Design of Content Alignment Studies in Mathematics and Reading for 12th Grade NAEP and Other Assessments* is available at http://www.nagb.org/publications/design-document-final.pdf.

[4] In 2009, 11 states participated in a pilot to obtain state-representative NAEP 12th grade results.

Low— recall or recognize concepts or procedures, carry out specified procedures (25%); Moderate— think flexibly in solving problems; make connections among concepts and processes from various domains (50%); and High— use reasoning, planning, analysis, judgment, and creative thought in solving problems (25%). Items from the 2009 12th grade mathematics assessment were reviewed for this study.

Methodology

The study used two panels of subject matter experts. Each panel operated independently to permit an examination of the comparability of the respective results, to serve as an indication of the replicability and reliability of the study outcomes. The study produced within-panel reliability measures, but not cross-panel reliability measures. Sixteen subject matter experts, approximately equally balanced in representation of secondary and post-secondary educators, were engaged as panelists in reviewing and analyzing the NAEP and SAT mathematics assessment frameworks and standards documents as well as the test questions for each assessment. In order to ensure that the panelists did not hold biases toward either of the assessments included in the study, panelists with substantial involvement in the development of either NAEP or SAT were disqualified from participation in the alignment study. Items from each of the two assessments were examined relative to its own academic content standards and the content standards of the other assessment. Based on evidence from the pilot alignment study and feedback from the ACT-led studies using the same design, the decision was made to use a reduced set of items for the alignment of each assessment to its own framework. This decision was made to ensure that panelists could complete the study within the scheduled five days. A sample of 42 items were selected for aligning NAEP items to the NAEP framework and a sample of 40 SAT items were selected for the alignment of SAT items to the SAT framework/specifications. The selection was made to be representative of key features of each framework. Panelists also classified the depth of knowledge (DOK) associated with each item and each content standard.[5]

Key Conclusions

- NAEP and the SAT assess almost the same content areas of mathematics with similar emphasis for each.
- With regard to alignment to the NAEP framework, both SAT items (around 38%) and NAEP items (around 34%) had the highest percentage of their overall alignments to the Algebra standard. The smallest percentage of total NAEP items were found to align to Data Analysis, Statistics, and Probability (around 11%) while the lowest percentages of SAT items were aligned to the Measurement standard (around 8%).
- The percentages of alignments to each SAT standard were relatively evenly distributed in both assessments and similar in distribution across assessments. The SAT Algebra standard had the highest number of items from each assessment aligned: around 36% of the SAT items and 32% of the NAEP items. For both assessments, the next highest rate of alignment was for the Geometry standard with around 34% of the SAT items and 27 of the NAEP items aligned to that standard. The largest difference in emphasis was found for alignment to the SAT Data, Statistics, and Probability content standard: around 21% of the NAEP items and only 9% of the SAT items were aligned to this standard.
- The within-panel reliability measures were reasonably similar across the two panels.
- The two tests are also similar in the average depth of knowledge (DOK) levels of items. For the

[5] Depth of knowledge categories are defined in the study design document (see Footnote 1)

complete set of NAEP items, the average DOK of items was judged to be about 1.7; the average DOK of the SAT items was about 1.8. While most items in both tests were found to be DOK Level 2, NAEP items had a wider range of DOK, with more items coded to Level 1 and Level 3. The Level 3 items often involved application of concepts through short or extended constructed-response items.

The Alignment of the NAEP Grade 12 Reading Assessment and the SAT Critical Reading Test

Based on the Program of Preparedness Research adopted by the National Assessment Governing Board, research is being conducted to study the relationship between grade 12 NAEP and other indicators of preparedness for college and job training. Five categories of research are currently included in this research program: content alignment studies, statistical relationship studies, judgmental standard-setting studies, surveys, and benchmarking studies. This study brief focuses on one of the content alignment studies. This content alignment study was implemented based on a design developed by Dr. Norman Webb specifically for addressing the Board's preparedness research objective to compare the NAEP assessment with other assessments regarded as indicators of preparedness.[6]

Type of Study
Content Alignment

Purpose
Identify and describe content overlap and non-overlap between NAEP and the SAT Critical Reading Test

Project Period
September 2009 to November 2010

Contractor and Principal Investigator
WestEd, Stuart Rabinowitz (Principal Investigator), Peter Worth (Project Director), and Norman Webb (Senior Technical Advisor)

Background on SAT Critical Reading Test and the NAEP 12th Grade Reading Assessment

The SAT is a widely recognized standardized test used for college admissions. The College Board produces the SAT, which is now officially called the SAT Reasoning Test. SAT scores are produced for three major sections: Mathematics, Critical Reading, and Writing. It yields individual student results, but does not produce nationally representative results for any defined group. The types of passages in the SAT Critical Reading Test are passage-based (approximately 72% of the assessment) and sentence completion (approximately 28%). Passage-based reading assesses extended reasoning, literal comprehension, and vocabulary in context. Passages are taken from the natural sciences, humanities, social sciences and literary fiction. Passage length ranges from 100-850 words. Skills assessed include Identifying main and supporting ideas, determining the meaning of words in context, understanding authors' purposes, and understanding the structure and function of sentences. The SAT Critical Reading test is a fixed-form test comprising 67 items, and two forms were used for this study—a total of 134 SAT items. All items are multiple choice.

The NAEP 12th grade reading assessment produces reports to the public on the achievement of 12th graders. Results are nationally representative,[7] and, as required by law, NAEP does not produce individual student results. Reading passages are literary (30%) and informational (70%). Authentic texts are used that high school students could encounter. Literary texts include fiction, literary non-fiction, and poetry. Informational text includes exposition, argumentation or persuasive text, and procedural texts or documentation. Cognitive targets are (1) Locate/Recall, (2) Integrate/Interpret, and (3) Critique/Evaluate. Passage length is 500-1500 words. Passage difficulty is grade 12 appropriate. Item types include multiple-choice and both short and extended constructed response. Items reviewed were from the 2009 12th grade

[6] The design document—*Design of Content Alignment Studies in Mathematics and Reading for 12th Grade NAEP and Other Assessments* is available at http://www.nagb.org/publications/design-document-final.pdf

[7] In 2009, 11 states participated in a pilot to obtain state-representative NAEP 12th grade results.

reading assessment, 131 in all.

Methodology

The study used two panels of subject matter experts. Each panel operated independently to permit an examination of the comparability of the respective results, to serve as an indication of the replicability and reliability of the study outcomes. The study produced within-panel reliability measures, but not cross-panel reliability measures. Fourteen subject matter experts, approximately equally balanced in representation of secondary and post-secondary educators, were engaged as panelists in reviewing and analyzing the NAEP and SAT reading assessment frameworks and standards documents as well as the test questions for each assessment. In order to ensure that the panelists did not hold biases toward either of the assessments included in the study, panelists with substantial involvement in the development of either NAEP or SAT were disqualified from participation in the alignment study. Items from each of the two assessments were examined relative to its own academic content standards and the content standards of the other assessment. Based on evidence from the pilot alignment study and feedback from the ACT-led studies using the same design, the decision was made to use a reduced set of items for the alignment of each assessment to its own framework. This decision was made to ensure that panelists could complete the study within the scheduled five days. A sample of 40 items were selected for aligning NAEP items to the NAEP framework and a sample of 46 SAT items were selected for the alignment of SAT items to the SAT framework/specifications. The selection was made to be representative of key features of each framework, and all items associated with reading passages were kept intact. Panelists also classified the depth of knowledge (DOK) associated with each item and each content standard.[8]

Key Conclusions

- The greatest commonality between the two tests is in their shared emphasis on the broad skills of integrating and interpreting informational and literary texts.
- The results of the item alignments show a broader range of cognitive complexity in the NAEP items.
- Approximately two-thirds of the NAEP items and over 90% of the SAT items aligned to one of the three NAEP standards: "Integrate/Interpret."
- All of the NAEP items and about three-quarters of the SAT items were aligned to the passage-based reading comprehension portion of the SAT. The NAEP Grade 12 Reading assessment does not align with the sentence completion portion of the SAT assessment.
- Both tests emphasize many of the same or closely related specific skills including inferring/analyzing the main idea and author's purpose; the tone and attitude of an author or character; the use of rhetorical strategies; and connections between ideas, perspectives, or problems.
- The within-panel reliability measures were reasonably similar across the two panels.
- The average depth of knowledge (DOK) level for NAEP content standards (2.3) was lower on a four-point scale than the DOK level for SAT (2.5). The average DOK level for NAEP items (2.4) was also lower than for SAT items (2.5). No SAT items were coded to DOK level 1.

[8] Depth of knowledge categories are defined in the study design document (see Footnote 1).

The Alignment of the NAEP Grade 12 Mathematics Assessment and the ACCUPLACER Mathematics Test

Based on the Program of Preparedness Research adopted by the National Assessment Governing Board, research is being conducted to study the relationship between grade 12 NAEP and other indicators of preparedness for college and job training. Five categories of research are currently included in this research program: content alignment studies, statistical relationship studies, judgmental standard-setting studies, surveys, and benchmarking studies. This study brief focuses on one of the content alignment studies. This content alignment study was implemented based on a design developed by Dr. Norman Webb specifically for addressing the Board's preparedness research objective to compare the NAEP assessment with other assessments regarded as indicators of preparedness.[9]

Type of Study
Content Alignment

Purpose
Identify and describe content overlap and non-overlap between NAEP and the ACCUPLACER Mathematics

Project Period
September 2009 to November 2010

Contractor and Principal Investigator
WestEd, Stuart Rabinowitz (Principal Investigator), Peter Worth (Project Director), and Norman Webb (Senior Technical Advisor).

Background on ACCUPLACER Mathematics Test and the NAEP 12th Grade Mathematics Assessment

The ACCUPLACER is a widely recognized set of standardized tests used for college course placement. It yields individual student results, but does not produce nationally representative results for any defined group. The ACCUPLACER mathematics assessment is a computer-adaptive test, consisting of a large pool of items from which a test-generation algorithm selects items for a student. The College Board provided two paper-based forms that are an alternative format to the computer-adaptive administration and are used in test centers that lack the computer-based delivery mode. These forms have been determined by the College Board to be representative of the ACCUPLACER item pool and have been used in other ACCUPLACER alignment studies. ACCUPLACER mathematics examines students in three core test areas: Arithmetic, Elementary Algebra, and College Level Math that are further delineated into 87 objectives. The Arithmetic Test assesses basic computation skills. The Elementary Algebra Test assesses basic computation skills and the basic skills of algebraic manipulation that may be acquired in typical high school Algebra I and II courses. The College-Level Math Test assesses more advanced algebra skills typically required at the end of high school and beginning of college, as well as geometry and trigonometry. Each paper-based form consists of 35 items—20 items specific to that form (variable items) and 15 items common to both forms (common items) for each of the three core tests included on each form—a total of 105 items per form. For purposes of efficiency and balance of workload across assessments and content areas, one form was selected for use in this study, for a total of 105 items. All items are multiple choice.

The NAEP 12th grade mathematics assessment produces reports to the public on the achievement of 12th graders. Results are nationally representative,[10] and, as required by law, NAEP does not produce

[9] The design document—*Design of Content Alignment Studies in Mathematics and Reading for 12th Grade NAEP and Other Assessments* is available at http://www.nagb.org/publications/design-document-final.pdf

[10] In 2009, 11 states participated in a pilot to obtain state-representative NAEP 12th grade results.

individual student results. Test item types include multiple choice and both short and extended constructed response, and involve both real-world contexts and pure mathematics. The assessment is comprised of five content areas: Number Properties and Operations (10%), Measurement and Geometry (30%), Data Analysis, Statistics, and Probability (25%), and Algebra (35%). There are three levels of item complexity: Low— recall or recognize concepts or procedures, carry out specified procedures (25%); Moderate— think flexibly in solving problems; make connections among concepts and processes from various domains (50%); and High— use reasoning, planning, analysis, judgment, and creative thought in solving problems (25%). Items from the 2009 12th grade mathematics assessment were used for this study.

Methodology

The study used two panels of subject matter experts. Each panel operated independently to permit an examination of the comparability of the respective results, to serve as an indication of the replicability and reliability of the study outcomes. The study produced within-panel reliability measures, but not cross-panel reliability measures. Sixteen subject matter experts, approximately equally balanced in representation of secondary and post-secondary educators, were engaged as panelists in reviewing and analyzing the NAEP and ACCUPLACER mathematics assessment frameworks and standards documents as well as the test questions for each assessment. In order to ensure that the panelists did not hold biases toward either of the assessments included in the study, panelists with substantial involvement in the development of either NAEP or ACCUPLACER were disqualified from participation in the alignment study. Items from each of the two assessments were examined relative to its own academic content standards and the content standards of the other assessment. Based on evidence from the pilot alignment study and feedback from the ACT-led studies using the same design, the decision was made to use a reduced set of items for the alignment of each assessment to its own framework. This decision was made to ensure that panelists could complete the study within the scheduled five days. A sample of 42 items were selected for aligning NAEP items to the NAEP framework and a sample of 45 ACCUPLACER items were selected for the alignment of ACCUPLACER items to the ACCUPLACER framework/specifications. The selection was made to be representative of key features of each framework. Panelists also classified the depth of knowledge (DOK) associated with each item and each content standard.[11]

Key Conclusions

- The greatest area of commonality between the assessments is in the areas of Number Properties and Operations and Algebra.
- Most (90%) of the ACCUPLACER items were aligned to NAEP objectives in either Number Properties and Operations or Algebra, and the items aligned to a limited number of objectives within those content areas.
- NAEP items were aligned to 75 of the 87 ACCUPLACER objectives and were distributed fairly evenly across the three ACCUPLACER content areas.
- The greatest areas of difference are in measurement, geometry, and data analysis. Many NAEP items assessing these content areas could not be aligned to the ACCUPLACER.
- NAEP items were more evenly and more broadly distributed across the NAEP objectives than were the ACCUPLACER items.

[11] Depth of knowledge categories are defined in the study design document (see Footnote 1).

- The within-panel reliability measures were reasonably similar across the two panels.
- For alignment to the NAEP framework, the NAEP items were found to meet depth-of-knowledge (DOK) consistency in all standards. For alignment to the ACCUPLACER framework, DOK was analyzed as range of depth of knowledge. NAEP items aligned to the ACCUPLACER framework were coded at DOK Levels 1–3, with roughly two-thirds of the NAEP items at DOK Level 2. Almost all ACCUPLACER items aligned to DOK Level 1 or Level 2. The ACCUPLACER items aligned to the arithmetic content area were about evenly split between DOK levels 1 and 2; nearly three-quarters of the items aligned to the Elementary Algebra content were coded as DOK level 1; and a little over two-thirds of the ACCUPLACER items aligned to College Math were coded at DOK level 2.

The Alignment of the NAEP Grade 12 Reading Assessment and the ACCUPLACER Reading Test

Based on the Program of Preparedness Research adopted by the National Assessment Governing Board, research is being conducted to study the relationship between grade 12 NAEP and other indicators of preparedness for college and job training. Five categories of research are currently included in this research program: content alignment studies, statistical relationship studies, judgmental standard-setting studies, surveys, and benchmarking studies. This study brief focuses on one of the content alignment studies. This content alignment study was implemented based on a design developed by Dr. Norman Webb specifically for addressing the Board's preparedness research objective to compare the NAEP assessment with other assessments regarded as indicators of preparedness.[12]

Type of Study
Content Alignment

Purpose
Identify and describe content overlap and non-overlap between NAEP and the ACCUPLACER Reading Test

Project Period
September 2009 to November 2010

Contractor and Principal Investigator
WestEd, Stuart Rabinowitz (Principal Investigator), Peter Worth (Project Director), and Norman Webb (Senior Technical Advisor)

Background on ACCUPLACER Reading Test and the NAEP 12th Grade Reading Assessment

The ACCUPLACER is a widely recognized set of standardized tests used for college course placement. It yields individual student results, but does not produce nationally representative results for any defined group. The ACCUPLACER Reading assessment is a computer-adaptive test, consisting of a large pool of items from which a test-generation algorithm selects items for a student. The College Board provided two paper-based forms that are an alternative format to the computer-adaptive administration and are used in test centers that lack the computer-based delivery mode. These forms have been determined by the College Board to be representative of the ACCUPLACER item pool and have been used in other ACCUPLACER alignment studies. Each paper-based form consists of 35 items—20 items specific to that form (variable items) and 15 items common to both forms (common items)—for a total of 70 items. All items are multiple choice.

The ACCUPLACER for reading assesses Reading Comprehension (75%) and Sentence Relationships (25%). Reading Comprehension passages may be long or short. Five objectives were specified for the ACCUPLACER in reading: Identifying Main Ideas, Direct Statements/Secondary Ideas, Inferences, Applications, and Sentence Relationships. The assessment of Sentence Relationships focuses on two sentences followed by a question about the relationship between the two sentences.

The NAEP 12th grade reading assessment produces reports to the public on the achievement of 12th graders. Results are nationally representative,[13] and, as required by law, NAEP does not produce individual student results. Reading passages are literary (30%) and informational (70%). Authentic texts are used that high school students could encounter. Literary texts include fiction, literary non-fiction, and poetry. Informational text includes exposition, argumentation or persuasive text, and procedural texts or

[12] The design document—*Design of Content Alignment Studies in Mathematics and Reading for 12th Grade NAEP and Other Assessments* is available at http://www.nagb.org/publications/design-document-final.pdf

[13] In 2009, 11 states participated in a pilot to obtain state-representative NAEP 12th grade results.

documentation. Cognitive targets are (1) Locate/Recall, (2) Integrate/Interpret, and (3) Critique/Evaluate.

Passage length is 500-1500 words. Passage difficulty is grade 12 appropriate. Item types include multiple-choice and both short and extended constructed response. Items reviewed were from the 2009 12th grade reading assessment, 131 in all.

Methodology

The study used two panels of subject matter experts. Each panel operated independently to permit an examination of the comparability of the respective results, to serve as an indication of the replicability and reliability of the study outcomes. The study produced within-panel reliability measures, but not cross-panel reliability measures. Fourteen subject matter experts, approximately equally balanced in representation of secondary and post-secondary educators, were engaged as panelists in reviewing and analyzing the NAEP and ACCUPLACER Reading assessment frameworks and standards documents as well as the test questions for each assessment. In order to ensure that the panelists did not hold biases toward either of the assessments included in the study, panelists with substantial involvement in the development of either NAEP or ACCUPLACER were disqualified from participation in the alignment study. Items from each of the two assessments were examined relative to its own academic content standards and the content standards of the other assessment. Based on evidence from the pilot alignment study and feedback from the ACT-led studies using the same design, the decision was made to use a reduced set of items for the alignment of NAEP items to the NAEP framework. This decision was made to ensure that panelists could complete the study within the scheduled five days. A sample of 40 items was selected for aligning NAEP items to the NAEP framework. The selection was made to be representative of key features of the NAEP framework. Panelists also classified the depth of knowledge (DOK) associated with each item and each content standard.[14]

Key Conclusions

- The greatest commonality between the two tests is in their shared emphasis on the broad skills of comprehending and interpreting informational text, primarily through inferential reasoning.
- A majority of both ACCUPLACER and NAEP items were aligned to the NAEP "Integrate/Interpret" objective.
- Items from each assessment had the highest rates of alignment to the ACCUPLACER "Inferences" objective, and a higher percentage of NAEP items than ACCUPLACER items were aligned to this objective.
- NAEP items aligned only to the ACCUPLACER Reading Comprehension objectives; NAEP items did not align to Sentence Relationships.
- NAEP addresses literary and informational text, while ACCUPLACER does not address reading skills specific to literary text.
- Up to 15% of the items in the NAEP item pool were unaligned to the ACCUPLACER.
- ACCUPLACER items did not align to the NAEP "Critique/Evaluate" reading objective.

[14] Depth of knowledge categories are defined in the study design document (see Footnote 1).

- The within-panel reliability measures were reasonably similar across the two panels.
- For alignment to the NAEP framework, the NAEP items were found to meet depth-of-knowledge (DOK) consistency in all standards. The ACCUPLACER items met depth-of-knowledge consistency only for items in the NAEP “Locate/Recall” standard. Both panels found that the majority of ACCUPLACER items aligned to the “Interpret/Integrate” standard had a lower DOK level than that of the standard.
- For alignment to the ACCUPLACER specifications, the ACCUPLACER items were found to meet depth-of-knowledge consistency in all objectives. The NAEP items met depth-of-knowledge consistency in the four ACCUPLACER objectives to which there were alignments, but not for “Sentence Relationships” to which no NAEP items were aligned.

The Alignment of the NAEP Grade 12 Mathematics Assessment and the WorkKeys Applied Mathematics Assessment

Based on the Program of Preparedness Research adopted by the National Assessment Governing Board, research is being conducted to study the relationship between grade 12 NAEP and other indicators of preparedness for college and job training. Five categories of research are currently included in this research program: content alignment studies, statistical relationship studies, judgmental standard-setting studies, surveys, and benchmarking studies. This study brief focuses on one of the content alignment studies. This content alignment study was implemented based on a design developed by Dr. Norman Webb specifically for addressing the Board's preparedness research objective to compare the NAEP assessment with other assessments regarded as indicators of preparedness.[15]

Type of Study
Content Alignment

Purpose
Identify and describe content overlap and non-overlap between NAEP and the ACT WorkKeys Assessment

Project Period
September 2009 to October 2010

Contractor and Principal Investigator
ACT, Oliver Cummings and Jennifer Horn-Frasier

Background on ACT WorkKeys Applied Mathematics Assessment and the NAEP 12th Grade Mathematics Assessment

The WorkKeys assessment is a widely recognized set of standardized tests related to the workplace. It yields individual student results; it does not produce nationally representative results for any defined group. All items are multiple choice and involve real-world application of mathematics content in a workplace context. The mathematics content is comprised of the following five areas: number properties and operations (62%), measurement and geometry (32%), data analysis, statistics and probability (0 – 6%), and algebra (0 – 6 %). There are five levels of item complexity, related to the WorkKeys mathematics scale of 3-7: Level 3 – basic mathematics operations (20%), Level 4 – multiple mathematical operations (20%), Level 5 – application of logic and calculation (20%), Level 6 – complex and multiple-step calculation (20%), and Level 7 – nonlinear functions, conversions, and complex calculations. There are 30 items per form, and the two WorkKeys forms used in the study shared two common items, for a total of 56 unique items plus 2 common items in the study.

The NAEP 12th grade mathematics assessment produces reports to the public on the achievement of 12th graders. Results are nationally representative,[16] and, as required by law, NAEP does not produce individual student results. Test item types include multiple choice and both short and extended constructed response, and involve both real-world contexts and pure mathematics. The assessment is comprised of five content areas: Number Properties and

[15] The design document—*Design of Content Alignment Studies in Mathematics and Reading for 12th Grade NAEP and Other Assessments* is available at http://www.nagb.org/publications/design-document-final.pdf

[16] In 2009, 11 states participated in a pilot to obtain state-representative NAEP 12th grade results.

Operations (10%), Measurement and Geometry (30%), Data Analysis, Statistics, and Probability (25%), and Algebra (35%). There are three levels of item complexity: Low— recall or recognize concepts or procedures, carry out specified procedures (25%); Moderate— think flexibly in solving problems; make connections among concepts and processes from various domains (50%); and High— use reasoning, planning, analysis, judgment, and creative thought in solving problems (25%). Items from the 2009 12th grade mathematics assessment were used for this study.

Methodology

The study used two panels of subject matter experts. Each panel operated independently to permit an examination of the comparability of the respective results, to serve as an indication of the replicability and reliability of the study outcomes. The study produced within-panel reliability measures, but not cross-panel reliability measures. Fifteen subject matter experts from universities, professional organizations, school districts, and schools were engaged as panelists in reviewing and analyzing the NAEP and WorkKeys mathematics assessment frameworks and standards documents, as well as the test questions for each assessment. Some of the panelists had previous experience or were familiar with NAEP and/or WorkKeys. Items from each of the two assessments were examined relative to its own academic content standards and the content standards of the other assessment. Panelists also classified the depth of knowledge (DOK) associated with each item and each content standard.[17]

Key Conclusions

- The NAEP Grade 12 Mathematics assessment covers a broader range of math skills than does the WorkKeys Applied Mathematics assessment, particularly in Geometry; Data Analysis, Statistics, and Probability; and Algebra.
- The WorkKeys Applied Mathematics assessment focuses on a narrower range of math skills than does the NAEP assessment. Specifically, the WorkKeys assessment focuses on the application of foundational math skills in workplace situations.
- Most of the WorkKeys items aligned with NAEP objectives related to number operations and measurement.
- WorkKeys objectives that are not assessed by the NAEP items include conversions, determining the best deal, finding errors, and calculating discounts or markups. Thus, even though the math content included in the NAEP framework is significantly broader than that included in the WorkKeys framework, there are numerous WorkKeys objectives that the NAEP items used in this study did not target.
- The within-panel reliability measures were reasonably similar across the two panels.
- The average DOK level for NAEP content standards (2.0) was higher on a four-point scale than the DOK level for WorkKeys (1.6).

[17] Depth of knowledge categories are defined in the study design document (see Footnote 1)

The Alignment of the NAEP Grade 12 Reading Assessment and the WorkKeys Reading for Information Assessment

Based on the Program of Preparedness Research adopted by the National Assessment Governing Board, research is being conducted to study the relationship between grade 12 NAEP and other indicators of preparedness for college and job training. Five categories of research are currently included in this research program: content alignment studies, statistical relationship studies, judgmental standard-setting studies, surveys, and benchmarking studies. This study brief focuses on one of the content alignment studies. This content alignment study was implemented based on a design developed by Dr. Norman Webb specifically for addressing the Board's preparedness research objective to compare the NAEP assessment with other assessments regarded as indicators of preparedness.[18]

Type of Study
Content Alignment

Purpose
Identify and describe content overlap and non-overlap between NAEP and the ACT WorkKeys Assessment

Project Period
September 2009 to October 2010

Contractor and Principal Investigator
ACT, Oliver Cummings and Jennifer Horn-Frasier

Background on ACT WorkKeys Reading for Information Assessment and the NAEP 12th Grade Reading Assessment

The WorkKeys assessment is a widely recognized set of standardized tests related to the workplace. It yields individual student results; it does not produce nationally representative results for any defined group. Reading passages are informational procedural texts and documents, drawn from authentic workplace settings, including memoranda, policies, instructions, bulletins, signs, and legal documents. Passage length is 70- 500 words. Passage difficulty ranges from 6th grade through postsecondary. Cognitive targets are: (1) choosing main ideas or details, (2) applying instructions, (3) applying information, and (4) applying reasoning. All items are multiple choice. There are 30 items per form, and two WorkKeys forms, totaling 60 items, were used in the study.

The NAEP 12th grade reading assessment produces reports to the public on the achievement of 12th graders. Results are nationally representative,[19] and, as required by law, NAEP does not produce individual student results. Reading passages are literary (30%) and informational (70%). Authentic texts are used that high school students could encounter. Literary texts include fiction, literary non-fiction, and poetry. Informational text includes exposition, argumentation or persuasive text, and procedural texts or documentation. Cognitive targets are (1) Locate/Recall, (2) Integrate/Interpret, and (3) Critique/Evaluate. Passage length is 500-1500 words. Passage difficulty is grade 12 appropriate. Item types include multiple-choice and both

[18] The design document—*Design of Content Alignment Studies in Mathematics and Reading for 12th Grade NAEP and Other Assessments* is available at http://www.nagb.org/publications/design-document-final.pdf

[19] In 2009, 11 states participated in a pilot to obtain state-representative NAEP 12th grade results.

short and extended constructed response. Items reviewed were from the 2009 12th grade reading assessment, 131 in all.

Methodology

The study used two panels of subject matter experts. Each panel operated independently to permit an examination of the comparability of the respective results, to serve as an indication of the replicability and reliability of the study outcomes. The study produced within-panel reliability measures, but not cross-panel reliability measures. Thirteen subject matter experts from universities, professional organizations, school districts, and schools were engaged as panelists in reviewing and analyzing the NAEP and WorkKeys reading assessment frameworks and standards documents as well as the test questions for each assessment. Some of the panelists had previous experience or were familiar with NAEP and/or WorkKeys. Items from each of the two assessments were examined relative to its own academic content standards and the content standards of the other assessment. Panelists also classified the depth of knowledge (DOK) associated with each item and each content standard.[20]

Key Conclusions

- The NAEP Grade 12 Reading assessment covers a broader range of reading skills than does the WorkKeys Reading for Information assessment, particularly in the literary genre and in requiring examinees to critique and evaluate reading materials.
- The WorkKeys Reading for Information assessment focuses on a narrower range of reading skills than does the NAEP assessment. Specifically, the WorkKeys assessment focuses on workplace communications, especially policies and instructions, and their application to workplace situations.
- Most of the WorkKeys items aligned with NAEP objectives were related to locating/recalling information and causal relations. WorkKeys items that aligned under Standard 2, Integrate/Interpret, targeted objectives that require the examinee to connect ideas, draw conclusions and provide supporting information, and to determine word meaning in context. No WorkKeys items included in this study require the examinee to critique or evaluate the reading passage.
- WorkKeys objectives that are not assessed by the NAEP items include applying complex, multistep, conditional instructions to similar and new workplace situations; determining the meaning of work-related acronyms, jargon, and technical terms; and figuring out and applying general principles contained in informational documents to similar and new workplace situations.
- The within-panel reliability measures were reasonably similar across the two panels.
- The average DOK level for NAEP content standards (2.5) was higher on a four-point scale than the DOK level for WorkKeys (1.9). The average DOK level for NAEP items (2.1) was also higher than for WorkKeys items (1.5).

[20] Depth of knowledge categories are defined in the study design document (see Footnote 1).

Exemplar Occupations for NAEP Reporting

Type of Study
Judgmental Standard-Setting

Purpose
To identify a set of 20 occupations from which 5-7 occupations can be selected for judgmental standard-setting studies

Project Period
September 2008 to August 2010

Contractor and Principal Investigator
ACT, Oliver Cummings and Helen Palmer

Based on the Program of Preparedness Research adopted by the National Assessment Governing Board, research is being conducted to study the relationship between grade 12 NAEP and other indicators of preparedness for college and job training Five categories of research are currently included in this research program: content alignment studies, statistical relationship studies, judgmental standard-setting studies, surveys, and benchmarking studies.

This study brief focuses on one of the studies prerequisite to judgmental standard-setting in relation to preparedness for job training.

Methodology
The study had three major components: (1) identify 20 civilian occupations from a set of criteria provided by the Governing Board, (2) determine which of these 20 occupations have a match in the military from publicly available information, and (3) attempt to identify up to 3 training programs for each job and the reading and mathematics tests and cut scores needed to qualify for the training.

Identify 20 Civilian Occupations
ACT used data from the U. S. Department of Labor, specifically O*NET[21] and the Bureau of Labor Statistics (BLS), to identify civilian occupations that meet the criteria provided by the National Assessment Governing Board (NAGB). These criteria were:

- Occupations in 0*NET Zones 2 and 3 that require substantial postsecondary training, but do not require a bachelor's degree.
- Coverage of a wide span of occupational families and industry sectors to the extent possible.
- Recognizability of occupations, i.e., they should be familiar to the public.
- High current employment level and high levels projected into the future, to cover a large proportion of jobs and a pace of job openings not projected to decline in the future.
- High growth projected into the future, to consider occupations in which high growth in important areas is expected.
- Reading and mathematics requirements to qualify for job training should represent a range of skill levels along the NAEP reporting scale.

Determine Comparability with Military Occupations

[21] The O*NET program is the nation's primary source of occupational information. The O*NET database contains information on hundreds of standardized and occupation-specific descriptors and is available on-line. http://www.onetcenter.org/overview.html

This activity involved the matching of the exemplar civilian jobs to military occupations based on job title using readily available resources on the internet: O*NET's Code Connector Search Engine, and the Department of Defense's and the National Crosswalk Service Center's (NCSC) Military Crosswalk Database. For those civilian occupations which had military counterparts, the next step was to compare tasks for the civilian and military occupations in order to determine the similarity of the occupations. ACT gathered task lists for civilian occupations from O*NET. Military task lists were from the U.S. Army's official website (goarmy.com) or the Army Credentialing Opportunities On-line (COOL) website.

Identify Training Programs and Reading and Mathematics Qualifications

ACT's database of WorkKeys profiles was searched for profiles related to training programs. Training program-related profiles were found for three of the exemplar occupations: Plumber, Electrician and Practical Nurse. WorkKeys skill levels needed to enter a training/apprenticeship program for each of the three occupations was identified. The majority of the search effort focused on searching on-line resources for postsecondary training programs, particularly those at colleges or community colleges.

Key Results

The study identified the following 20 occupations for consideration. An asterisk indicates that a match with a military occupation could not be confirmed. Training programs or courses were identified in every case. There was great variability in the availability of information about whether reading and mathematics proficiency were tested and, if tested, the tests used and the cut-scores needed to qualify.

OCCUPATION TITLE	TRAINING PROGRAMS FOUND? (Y/N)
Automotive master mechanics	Y
Bookkeeping, auditing, accounting clerks	Y
Civil engineering technicians	Y
Computer support specialists	Y
Customer service representatives*	Y
Dental hygienists	Y
Electrical engineering technician*	Y
Electrician	Y
Hairdressers, hairstylists, and cosmetologists*	Y
Licensed practical and licensed vocational nurses	Y
Medical records and health information technician	Y
Nursing aides, orderlies, and attendants*	Y
Paralegals and legal assistants	Y
Pharmacy technicians	Y
Plumber	Y
Police patrol officers	Y

Preschool teachers, except special education*	Y
Radiologic technologists	Y
Real estate sales agents*	Y
Registered nurses	Y