

Validity Argument for
NAEP Reporting on 12th Grade
Academic Preparedness for College



Ray Fields
National Assessment Governing Board
November 2013

National Assessment Governing Board

BOARD MEMBERSHIP (2013-2014)

Honorable David P. Driscoll
Chair
Former Commissioner of Education
Melrose, Massachusetts

Susan Pimentel
Vice Chair
Educational Consultant
Hanover, New Hampshire

Andrés Alonso
Professor of Practice
Harvard Graduate School
of Education
Harvard University
Cambridge, Massachusetts

Lucille E. Davy
President and CEO
Transformative Education
Solutions, LLC
Pennington, New Jersey

Louis M. Fabrizio
*Data, Research and
Federal Policy Director*
North Carolina Department
of Public Instruction
Raleigh, North Carolina

Honorable Anitere Flores
Senator
Florida State Senate
Miami, Florida

Rebecca Gagnon
School Board Member
Minneapolis Public Schools
Minneapolis, Minnesota

Shannon Garrison
Fourth-Grade Teacher
Solano Avenue Elementary School
Los Angeles, California

Doris R. Hicks
*Principal and
Chief Executive Officer*
Dr. Martin Luther King, Jr. Charter
School for Science and Technology
New Orleans, Louisiana

Andrew Dean Ho
Assistant Professor
Harvard Graduate School
of Education
Harvard University
Cambridge, Massachusetts

Honorable Terry Holliday
Commissioner of Education
Kentucky Department of Education
Lexington, Kentucky

Richard Brent Houston
Assistant Superintendent
Shawnee Public Schools
Shawnee, Oklahoma

Hector Ibarra
Eighth-Grade Teacher
Belin Blank International Center
and Talent Development
Iowa City, Iowa

Honorable Tom Luna
*Idaho Superintendent
of Public Instruction*
Boise, Idaho

Terry Mazany
President and CEO
The Chicago Community Trust
Chicago, Illinois

Tonya Miles
General Public Representative
Mitchellville, Maryland

Dale Nowlin
Twelfth-Grade Teacher
Columbus North High School
Columbus, Indiana

Joseph M. O'Keefe, S.J.
Professor
Lynch School of Education
Boston College
Chestnut Hill, Massachusetts

W. James Popham
Professor Emeritus
University of California, Los Angeles
Wilsonville, Oregon

B. Fielding Rolston
Chairman
Tennessee State Board of Education
Kingsport, Tennessee

Cary Sneider
Associate Research Professor
Portland State University
Portland, Oregon

Honorable Leticia Van de Putte
Senator
Texas State Senate
San Antonio, Texas

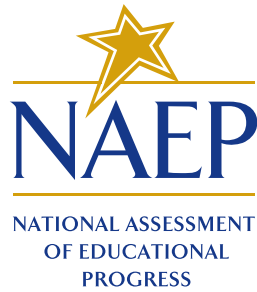
Ex-officio Member
John Q. Easton
Director
Institute of Education Sciences
U.S. Department of Education
Washington, D.C.

www.nagb.org | www.nagb.gov

Suggested Citation:

Fields, R. (2013). *Validity Argument for NAEP Reporting on 12th Grade Academic Preparedness for College*. Washington, DC: National Assessment Governing Board.

Validity Argument for NAEP Reporting on 12th Grade Academic Preparedness for College



Ray Fields
National Assessment Governing Board
November 2013

Acknowledgements

The author would like to thank the following individuals and organizations, whose many contributions were instrumental in the preparation of the validity argument contained herein.

Michael Kane, whose theoretical framework was used as the basis for the validity argument, served as technical advisor throughout its preparation.

Andrew Ho conceived the general form of the inferences that are at the heart of the validity argument.

The members of the Governing Board Committee on Standards, Design and Methodology—Chair Louis M. Fabrizio, Vice Chair Fielding Rolston, Andrew Ho, Terry Holliday, Tonya Miles, W. James Popham, Leticia Van De Putte and Ex Officio Member John Q. Easton—provided rigorous oversight and critical review of work in progress.

Gregory Cizek and Mark Reckase provided the independent technical reviews of the validity argument that appear in Appendix A and Appendix B, respectively.

ACT, Inc., the College Board, and the Florida Department of Education were essential partners, whose collaboration and participation were vital to the viability and quality of the NAEP 12th grade academic preparedness research.

Finally, many thanks go to Jason Smith, Jacqui Lipson, Neby Ejigu, Peter Hahn, and Dan Kaufman of Widmeyer Communications, Inc. for proofreading and report design services.

Contents

Acknowledgements	ii
List of Figures	iv
Executive Summary	v
I. Introduction	1
Rationale for NAEP Reporting on 12 th Grade Academic Preparedness	1
Defining Academic Preparedness for College	3
Domain Definition for Academic Preparedness for College in Reading and Mathematics	4
The Central Issue: Validity	4
II. Validity Argument	6
Overview	6
Proposed Inferences	8
Limitations on Interpretation and Other Caveats	8
Discussion of the Claims and Evidence	10
Indicators: College Board and ACT College Readiness Benchmarks	12
Indicators: First Year GPA of B- or Better and Remedial/non-Remedial Placement	12
NAEP Assessment Frameworks Were Revised to Measure Academic Preparedness	18
Content Alignment Studies Found Significant Overlap between NAEP and the ACT, SAT and ACCUPLACER	21
The Content Alignment Studies: Overview	22
The Content Alignment Studies: Reading Results	23
The Content Alignment Studies: Summary Discussion for Reading	31
The Content Alignment Studies: Mathematics Results	32
The Content Alignment Studies: Summary Discussion for Mathematics	38
Discussion of Test Uses and Consequences in Relation to the Proposed Inferences	40
Summary and Conclusion	40
III. Epilogue and Concluding Considerations	43
IV. References	45
Appendices	
Appendix A: Technical Review of Validity Argument: Gregory Cizek	A-1
Appendix B: Technical Review of Validity Argument: Mark Reckase	B-1
Appendix C: Motion on Reporting 12 th Grade Academic Preparedness for College	C-1
Appendix D: Prototype Chapter for NAEP Reporting	D-1
Appendix E: Background Information for Figure 1 and Figure 2	E-1

List of Figures

Figure 1 NAEP 12 th Grade Preparedness Research: Mathematics	15
Figure 2 NAEP 12 th Grade Preparedness Research: Reading	17
Figure 3 Examples of Objectives Added to the 2009 Grade 12 Mathematics Framework.	20

Executive Summary

The National Assessment Governing Board is conducting a program of research to determine the feasibility of the National Assessment of Educational Progress (NAEP) reporting on the academic preparedness of U.S. 12th grade students, in reading and mathematics, for college and job training.

Twelfth grade is the end of mandatory schooling for most students and represents the transition point to adult postsecondary pursuits. Currently, there are no nationally representative indicators to report to the public on the academic preparedness of 12th grade students for college and job training. As the only source of national and state-representative student achievement data at the 12th grade, NAEP is uniquely positioned to serve as an indicator of 12th grade academic preparedness. The research and conclusions contained herein are limited to academic preparedness for college.

Defining Academic Preparedness for College

In the United States in 2013, there is no single definition of “academic preparedness for college” used by colleges for admission and placement. Postsecondary education in the U.S. is a complex mix of institutions, public and private, that have different admission requirements and different procedures and criteria for placing individual students into education programs.

In order to design the NAEP 12th grade preparedness research, a working definition of preparedness was needed. As a working definition, it would change and evolve based on the research results. The working definition used for the NAEP preparedness research is intended to apply to the “typical” student in the “typical” college, not to highly selective institutions, and thus, to the vast majority of prospective students:

Academic preparedness for college refers to the reading and mathematics knowledge and skills needed to qualify for placement into entry-level, credit-bearing, non-remedial courses that meet general education degree requirements in broad access 4-year institutions and, for 2-year institutions, for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

The Central Issue: Validity

Having made the decision to determine the feasibility of NAEP reporting on 12th grade academic preparedness, the Governing Board recognized that the central concern would be establishing the validity of inferences about 12th grade academic preparedness that are to be made from NAEP scores and used in NAEP reports. The Governing Board would need to ensure that the content of NAEP 12th grade reading and mathematics assessments was appropriate for measuring academic preparedness and that research was conducted to collect evidence by which the validity of proposed inferences could

be evaluated. Finally, a formal validity argument would need to be developed, specifying the proposed inference(s) for NAEP reporting, the underlying assumptions or propositions, and the evidence related to the assumptions or propositions.

Accordingly, the Governing Board

- revised the NAEP frameworks for the 2009 12th grade reading and mathematics assessments with the explicit purpose of measuring academic preparedness for college and job training,
- appointed a special panel of technical experts to recommend a program of research on 12th grade academic preparedness (National Assessment Governing Board, 2009a),
- approved and conducted a comprehensive set of preparedness research studies (<http://www.nagb.org/what-we-do/preparedness-research.html>), and
- adopted the model for a validity argument described by Michael Kane (Kane, 2013).

The first phase of the Governing Board's program of preparedness research is completed. The studies were conducted in connection with the 2009 NAEP 12th grade assessments in reading and mathematics. More than 30 studies of five distinct types were conducted.

The study results are used in this validity argument as evidence in support of statements related to academic preparedness for college that have the following general form:

The percentage of students in the NAEP distribution at or above a particular score level in reading or mathematics on 12th grade NAEP is a plausible, or reasonable, estimate of the percentage of 12th grade students who are academically prepared for college.

This NAEP score interpretation depends on four prior claims (or assumptions):

- The 12th grade NAEP results in reading and mathematics provide unbiased, accurate estimates of the percentages of students at or above specified score levels on the NAEP scales in reading and mathematics for 12th grade students in the United States.
- Performance on 12th grade NAEP assessments in mathematics and reading is positively related to other measures associated with outcomes reflecting academic preparedness for college.

- There is a point on the NAEP scale that corresponds to other measures, indicators, and outcomes associated with academic preparedness for college (i.e., possession of a specific level of academic proficiency, attainment of a first-year overall college GPA of B- or better, and placement into entry-level, credit-bearing, non-remedial college courses).
- The positive relationship between NAEP and the other indicators and outcomes is meaningful in terms of academic preparedness for college, not merely a statistical artifact, because the 12th grade reading and mathematics domains measured by NAEP were specifically designed to measure academic preparedness for college.

The first claim is supported by the combination of the content of the NAEP assessment frameworks and the NAEP test items, the NAEP sampling designs, and the statistical models used to generate estimates of score distributions at each grade level and in each content area. These elements in support of the claim are well established, documented, and evaluated; therefore, the attention of the validity argument is directed primarily to the second, third, and fourth claims.

The second claim is supported by a statistical relationship study that examined student performance on the NAEP 12th grade reading and mathematics assessments to performance on the SAT reading and mathematics tests, as well as the respective college readiness benchmarks established by the College Board for these tests, which, in turn, are related to outcomes associated with academic preparedness for college.

The third claim was evaluated with multiple sources of evidence in addition to the SAT/NAEP statistical relationship study. These include a longitudinal study of Florida 12th grade students and analyses of the 2005 and 2009 NAEP High School Transcript Studies.

The fourth claim is supported by the fact that the Governing Board reviewed the NAEP 12th grade reading and mathematics frameworks for the purpose of making NAEP a measure of academic preparedness for college; made changes to the frameworks accordingly; and conducted a comprehensive set of content alignment studies that found a high degree of match between NAEP and tests that are used for college admission and placement.

The results of these analyses were consistent across studies and across years. They support the conclusions that students in the NAEP 12th grade distribution at or above the Proficient achievement level in reading and at or above 163 on the NAEP score scale for mathematics are

- likely to be academically prepared for entry-level, credit-bearing, non-remedial courses in broad access 4-year institutions and, for 2-year institutions, for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions, and
- not likely to need remedial/developmental courses in reading or mathematics in college.

Thus, the NAEP 12th grade preparedness research results support the inferences that:

For reading:

Given the design, content, and characteristics of the NAEP 12th grade reading assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness,

the percentage of students scoring at or above Proficient on Grade 12 NAEP in reading is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in reading that would make them academically prepared for college.

For mathematics:

Given the design, content, and characteristics of the NAEP 12th grade mathematics assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness,

the percentage of students scoring at or above a score of 163 on the Grade 12 NAEP scale in mathematics is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics that would make them academically prepared for college.

Including these inferences in NAEP 12th grade reports will add meaning to the interpretation of the NAEP 12th grade results. In the past, NAEP has reported on what students know and can do in important subjects. Now, NAEP can serve as an indicator that conveys the percentage of the nation's 12th grade students that are academically prepared for entry-level, credit-bearing college courses without remediation.

However, steps must be taken to avoid potential misinterpretation. NAEP reports using these inferences must describe the limitations on interpretation and provide explicit caveats for consideration by the reader. In addition, the reports should explain the rationale for NAEP reporting on academic preparedness and describe appropriate and inappropriate uses of the results.

I. Introduction

Rationale for NAEP Reporting on 12th Grade Academic Preparedness

The National Assessment Governing Board is conducting a program of research to determine the feasibility of the National Assessment of Educational Progress (NAEP) reporting on the academic preparedness of U.S. 12th grade students, in reading and mathematics, for college and job training.

Since 1969, NAEP has reported to the public on the status and progress of student achievement in a wide range of key subjects at grades 4, 8, and 12. NAEP provides national and state-representative results, results for 21 urban districts, and results by subgroups of students (e.g., by race/ethnicity, gender, and for students with disabilities and English language learners). NAEP, by law, does not provide individual student results.

The Governing Board's initiative on 12th grade academic preparedness began in March 2004, with the report of a blue-ribbon panel.¹ The panel was composed of K-12 education leaders—the “producers” of high school graduates—and leaders in business, postsecondary education, and the military—the “consumers” of high school graduates.

The panel members recognized the importance of 12th grade as the gateway to postsecondary education and training, and viewed NAEP as a “truth teller” about student achievement. These distinguished state and national leaders recommended unanimously that “NAEP should report 12th grade students’ readiness for college-

credit coursework, training for employment, and entrance into the military” (National Commission on NAEP 12th Grade Assessment and Reporting, 2004, p. 6). They stated that “America needs to know how well prepared its high school seniors are... [only NAEP] can provide this information... and it is necessary for our nation’s continued well-being that it be provided” (National Commission on NAEP 12th Grade Assessment and Reporting, 2004, p. 2).

The Governing Board approved this recommendation, with a minor modification. The term “readiness” was changed to “academic preparedness” and “entrance into the military” was subsumed by “job training.”

“Readiness” was changed to “academic preparedness” because “readiness” is broadly understood to include both academic preparedness and other characteristics needed for success in postsecondary education and training, such as habits of mind, time management, and persistence (Conley, 2007). NAEP does not purport to measure such characteristics. Rather, NAEP is designed to measure academic knowledge and skills.

“Entrance into the military” was subsumed by “job training” with the intention of identifying occupations with civilian and military counterparts and utilizing the military’s experience as the world’s largest occupational training organization and its extensive research on the relationship between performance on the Armed Service Vocational Aptitude Battery (ASVAB) and job training outcomes.

¹ The blue-ribbon panel was known officially as the National Commission on NAEP 12th Grade Assessment and Reporting.

The Governing Board approved the 12th grade academic preparedness initiative because it believes that the academic preparation of high school students for postsecondary education and training is important to the nation's economic well-being, national security, and democratic foundations (see Governing Board resolution of May 21, 2005 at <http://www.nagb.org/content/nagb/assets/documents/policies/resolution-on-preparedness.pdf>).

Indicators of many kinds are used to monitor critical aspects of national life and inform public policy. These include economic indicators (e.g., gross domestic product), health indicators (e.g., cancer rates), and demographic indicators (e.g., population trends by race/ethnicity and gender). The Governing Board believes that NAEP reporting on the academic preparedness of 12th grade students would serve as a valuable indicator of the human capital potential of rising generations of citizens, a nation's greatest resource.

The Governing Board is not alone in recognizing the importance of 12th grade academic preparedness for the nation. A wide array of state and national leaders has embraced the goal that 12th grade students graduate "college and career ready." These include the leadership and members of the National Governors Association (NGA), the Council of Chief State School Officers (CCSSO), the Business Roundtable (BRT), the U.S. Chamber of Commerce (the Chamber), the Council on Foreign Relations, and the Obama administration. The reason for this attention to 12th grade academic preparedness is well summarized by a statement of the Business Coalition for Student Achievement,

an organization coordinated by BRT and the Chamber:

Ensuring that all students graduate academically prepared for college, citizenship and the 21st century workplace...is necessary to provide a strong foundation for both U.S. competitiveness and for individuals to succeed in our rapidly changing world. (See <http://www.biz4achievement.org/>)

The NGA and CCSSO have collaborated to develop Common Core State Standards (CCSS) for mathematics and English language arts. These standards are aimed at fostering college and career readiness by the end of high school. The CCSS have been adopted formally by 45 states, several territories and the Department of Defense Education Activity. Viewing the need for rigor in education standards and outcomes through the lens of national security, a similar conclusion was made in the report of the Independent Task Force on U.S. Education Reform and National Security of the Council on Foreign Relations. The Task Force was co-chaired by former New York City School Chancellor Joel Klein and former Secretary of State Condoleezza Rice. The Obama administration has stated that "educating every American student to graduate from high school prepared for college and for a career is a national imperative" (Fields & Parsad, 2012, pp. 3-4).

Twelfth grade is the end of mandatory schooling for most students and represents the transition point to adult postsecondary pursuits. If it is essential for students to graduate from high school academically prepared for college and job training, it is essential for the public and policymakers to know the degree to which this is occurring.

A trusted indicator is needed for reporting to the public and policymakers on the status of 12th grade academic preparedness in the U.S., but no such indicator exists. State tests at the high school level are typically administered at 10th and 11th grade. College admission tests, like the SAT and ACT, are administered before the 12th grade, generally to self-selected samples of students.

State tests and college admission tests do not provide a measure of what students know and can do at the very end of K-12 education. Even if these state tests and college admission tests were administered at the 12th grade, they could not be combined to produce nationally representative results.

NAEP is the only source of national and state-representative student achievement data at the 12th grade. As such, NAEP is uniquely positioned to serve as an indicator of 12th grade academic preparedness.

Defining Academic Preparedness for College

In the United States in 2013, there is no single, agreed-upon definition of “academic preparedness for college” used by colleges for admission and placement. Postsecondary education in the U.S. is a complex mix of institutions, public and private, that have different admission requirements and different procedures and criteria for placing individual students into education programs.

In this complex mix are 2-year institutions, 4-year public and private institutions with a wide range of selectivity, and proprietary schools. Institutions range from highly selective (i.e., with admission

criteria including very high grade point averages, successful completion of rigorous high school coursework, and very high SAT and/or ACT scores) to open admission (i.e., all applicants are admitted).

Even within institutions, requirements may vary across majors or programs of study. For example, the mathematics and science high school coursework and academic achievement needed for acceptance into an engineering program in a postsecondary institution may be more rigorous than the general requirements for admission to the institution or for a degree in elementary education in the institution.

In order to design the NAEP 12th grade preparedness research, a working definition of preparedness was needed. As a working definition, it would change and evolve based on the research results. The Governing Board’s Technical Panel on 12th Grade Preparedness Research recommended use of the following working definition, which defines academic preparedness for college as

... the academic knowledge and skill levels in reading and mathematics necessary to be qualified for placement...into a credit-bearing entry-level general education course that fulfills requirements toward a two-year transfer degree or four-year undergraduate degree at a postsecondary institution [without the need for remedial coursework in those subjects] (National Assessment Governing Board, 2009a, p. 3).

This definition was intended to apply to the “typical” college, not to highly selective institutions, and thus, to the vast majority of

prospective students, or about 80% of the college freshmen who enrolled in 2-year and 4-year institutions within 2 years following high school graduation (Ross et al., 2012, p. 175). To make this clear, the definition is further elaborated as follows:

Academic preparedness for college refers to the reading and mathematics knowledge and skills needed to qualify for placement into entry-level, credit-bearing, non-remedial courses that meet general education degree requirements (ECNRG) in broad access 4-year institutions and, for 2-year institutions, for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

This is consistent with the approach used by the College Board and ACT, Inc. in developing their respective college readiness benchmarks, which are used as external referents in the NAEP 12th grade preparedness research. The ACT benchmarks “represent predictive indicators of success for *typical* students at *typical* colleges” (Allen & Sconing, 2005, p. 3). The SAT benchmarks are “an indication of college readiness at a typical college” (College Board, 2011, p. 4).

Domain Definition for Academic Preparedness for College in Reading and Mathematics

The working definition described above set the stage for designing the preparedness research studies, but begged a basic question—What are the reading and mathematics knowledge and skills needed to qualify for placement into ECNRG

and are they measured by NAEP? This question would be addressed by examining the degree of content match between NAEP and multiple widely accepted external sources that had developed domain definitions for academic preparedness for college in mathematics and reading.

A perfect match between two different sources could not be expected, but a sufficient content match between NAEP and each of a multiple of relevant widely accepted external sources would, collectively, support the inference that the needed knowledge and skills are measured by NAEP. Consequently, the Governing Board identified the following external sources for content comparison with NAEP: The American Diploma Project (ADP) benchmarks for mathematics and English, the ACT College Readiness Standards for Mathematics and Reading, and the ACT, SAT, and ACCUPLACER assessments for reading and mathematics. The results of the content comparison studies between NAEP and these other sources are described in the validity argument below.

The Central Issue: Validity

Having made the decision to determine the feasibility of NAEP reporting on 12th grade academic preparedness, the Governing Board recognized that the central concern would be establishing the validity of inferences about 12th grade academic preparedness that are to be made from NAEP scores and used in NAEP reports. The Governing Board would need to ensure that the content of NAEP 12th grade reading and mathematics assessments was appropriate

for measuring academic preparedness and that research was conducted to collect evidence by which the validity of proposed inferences could be evaluated. Finally, a formal validity argument would need to be developed, specifying the proposed inference(s) for NAEP reporting, the underlying assumptions or propositions, and the evidence related to the assumptions or propositions.

Accordingly, the Governing Board

- revised the NAEP frameworks for the 2009 12th grade reading and mathematics assessments with the explicit purpose of measuring academic preparedness for college and job training,
- appointed a special panel of technical experts to recommend a program of research on 12th grade academic preparedness (National Assessment Governing Board, 2009a),
- approved and conducted a comprehensive set of preparedness research studies, and
- adopted the model for a validity argument described by Michael Kane (Kane, 2013).

The first phase of the Governing Board's program of preparedness research is completed. The studies were conducted in connection with the 2009 NAEP 12th grade assessments in reading and

mathematics. More than 30 studies of five distinct types have been conducted. Study results are available, and the complete studies are posted at <http://www.nagb.org/what-we-do/preparedness-research.html>. The National Center for Education Statistics (NCES) has provided additional data drawn from analyses of the 2005 and 2009 High School Transcript Studies conducted in connection with the NAEP 12th grade assessments in those years (see Appendix E).

From this research, Governing Board staff developed a proposed interpretation of NAEP performance in reading and mathematics related to 12th grade academic preparedness for college. Following below is the validity evidence for the proposed interpretation, presented in the form of a validity argument. The validity argument provides a statement of the proposed interpretation and the main assumptions inherent in the proposed interpretation in terms of academic preparedness for college. These assumptions are then evaluated using several lines of evidence, which were found to converge both for reading and for mathematics.

II. Validity Argument

Overview

The National Assessment of Educational Progress (NAEP) program is designed to provide information about student achievement in reading, mathematics and other content areas at the 4th, 8th, and 12th grades. The items for the assessments are developed according to content frameworks and test specifications developed by the National Assessment Governing Board. Scientific sampling procedures are used to produce estimates of score distributions representative of the national population of students at each grade level, as well as estimates representative of public school students in individual states and in 21 urban school districts.

The NAEP results do not produce scores for individual students, but rather, group estimates. The NAEP results are reported based on the estimated score distributions, by average score, percentiles, and in terms of the percentages of students at or above three performance standards used for NAEP reporting, called achievement levels, that are designated Basic, Proficient, and Advanced.

The purpose of the research reported here was to examine whether the interpretation of 12th grade NAEP results in reading and mathematics could be extended to include statements about the percentage of U.S. 12th graders who are academically prepared for college and, if such an interpretation were found to be defensible, to determine the specific statements about academic preparedness that were supportable by the research evidence. The specific statements would

be based on the following general definition for academic preparedness, used in relation to the NAEP preparedness research:

the reading and mathematics knowledge and skills needed to qualify for placement into entry-level, credit-bearing, non-remedial courses that meet general education degree requirements in broad access 4-year institutions and, for 2-year institutions, for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

The NAEP assessment program is well established and regularly evaluated, with ample technical documentation of the interpretation of the results at all three grade levels. Therefore, the technical quality, accuracy, and representativeness of the NAEP results in terms of the estimated distributions of U.S. 12th graders on the NAEP scales in reading and mathematics will be taken as a given and as a starting point for additional inferences about the academic preparedness of U.S. 12th graders for college.

In particular, the intent of this validity argument is to examine the evidence in support of statements related to academic preparedness for college for use in reporting NAEP 12th grade results that would have the following general form:

The percentage of students in the NAEP distribution at or above a particular score level in reading or mathematics on 12th grade NAEP is a plausible, or reasonable, estimate of the percentage of 12th grade students who are academically prepared for college.

This interpretation would depend on four prior claims (or assumptions):

1. The 12th grade NAEP results in reading and mathematics provide unbiased, accurate estimates of the percentages of students at or above specified score levels on the NAEP scales in reading and mathematics for 12th grade students in the United States.
2. Performance on 12th grade NAEP assessments in mathematics and reading is positively related to other measures associated with outcomes reflecting academic preparedness for college.
3. There is a point on the NAEP scale that corresponds to other measures, indicators, and outcomes associated with academic preparedness for college (i.e., possession of a specific level of academic proficiency, attainment of a first-year overall college GPA of B- or better, and placement into entry-level, credit-bearing, non-remedial college courses).
4. The positive relationship between NAEP and the other indicators and outcomes is meaningful in terms of academic preparedness for college, not merely a statistical artifact, because the 12th grade reading and mathematics domains measured by NAEP were specifically designed to measure academic preparedness for college.

The first claim is supported by the combination of the content of the NAEP assessment frameworks and the NAEP test items, the NAEP sampling designs, and the statistical models used to generate estimates of score distributions at each grade level and in each content area. These elements in support of the claim are well established, documented, and evaluated; therefore, the attention of the validity argument will be directed primarily to the second, third, and fourth claims.

The second claim is supported by a statistical relationship study that examined student performance on the NAEP 12th grade reading and mathematics assessments to performance on the SAT reading and mathematics tests, as well as the respective college readiness benchmarks established by the College Board for these tests, which, in turn, are related to outcomes associated with academic preparedness for college.

The third claim was evaluated with multiple sources of evidence that were highly convergent. These include the SAT/NAEP statistical relationship study, a longitudinal study of Florida 12th grade students, and analyses of the 2005 and 2009 NAEP High School Transcript Studies.

The fourth claim is supported by the fact that the Governing Board reviewed the NAEP 12th grade reading and mathematics frameworks for the purpose of making NAEP a measure of academic preparedness for college, made changes to the frameworks accordingly, and conducted a comprehensive set of content alignment studies to determine the degree of match between NAEP and tests that are used for college admission and placement.

Further, the results from the examination of the NAEP content provide a counterargument to a possible falsifying claim about the positive relationships discussed in the second and third claims. The falsifying claim would be that the positive relationships between NAEP and the other indicators were merely statistical artifacts, due to factors extraneous to academic preparedness for college, akin to finding a high correlation between height and passing rates on a state driving test. The counterargument is that the relationships are meaningful because the NAEP 12th grade reading and mathematics assessments were intentionally designed to measure academic

preparedness for college and that the evidence supports the conclusion that the NAEP 12th grade assessments do measure academic preparedness for college.

Proposed Inferences

The research results were consistent across studies and across years. On the basis of the research, the following inferences were formulated.

For reading:

Given the design, content, and characteristics of the NAEP 12th grade reading assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness:

the percentage of students scoring at or above Proficient on Grade 12 NAEP in reading is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in reading that would make them academically prepared for college.

For mathematics:

Given the design, content, and characteristics of the NAEP 12th grade mathematics assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness:

the percentage of students scoring at or above a score of 163 on the Grade 12 NAEP scale in mathematics is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics that would make them academically prepared for college.

In contrast to the inference for reading, which is set at the Proficient level, the inference for mathematics is set at a score of 163 on the

NAEP mathematics scale. This score is strongly supported by the consistent research results across years and data sources, but is below and significantly different from the cut score for the Proficient level for NAEP 12th grade mathematics, which is 176.

The research results for mathematics do support a related inference—that students in the distribution at or above the NAEP Proficient level in mathematics are likely to be academically prepared for college. However, the percentage of such students would be substantially less than the percentage in the distribution at or above 163, and thus, would underestimate the percentage of 12th grade students in the U.S. who are academically prepared for college.

For these reasons, and to have the proposed inferences for reading and mathematics as parallel as possible, the proposed inference for reading is formulated in relation to the Proficient achievement level, and the proposed inference for mathematics is formulated in relation to the NAEP mathematics scale score of 163.

Limitations on Interpretation and Other Caveats

False negatives and false positives. Some proportion of 12th grade students scoring below Proficient on the 12th grade NAEP Reading or below a score of 163 on the Mathematics Assessment are

- likely to be academically prepared for ECNRG college courses in broad access 4-year institutions and, for 2-year institutions, for entry-level placement into degree-bearing programs designed to transfer to 4-year institutions, and

- not likely to need remedial/developmental courses in reading or mathematics in college,

but with a lower probability than those at or above Proficient in reading or 163 in mathematics.

In addition, some proportion of 12th grade students scoring at or above Proficient on the 12th grade NAEP Reading or 163 on the Mathematics Assessment may

- not be academically prepared for ECNRG college courses in broad access 4-year institutions and, for 2-year institutions, for entry-level placement into degree-bearing programs designed to transfer to 4-year institutions, and
- need remedial/developmental courses in reading or mathematics in college.

Not a preparedness standard. The proposed inferences are not intended to represent or be used as standards for minimal academic preparedness for college. The proposed inferences are intended solely to add meaning to interpretations of the 12th grade NAEP reading and mathematics results in NAEP reports.

Academically prepared for college. The proposed inferences are intended to apply to the typical degree-seeking entry-level college student at the typical college. Thus, “academically prepared for college” refers to the reading and mathematics knowledge and skills needed for placement into ECNRG courses in broad access 4-year institutions and, for 2-year institutions, the general policies for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

It is important to note the focus on “placement” rather than “admission.” This distinction is made because students who need remedial courses in reading, mathematics or writing may be admitted to college, but not placed into regular, credit-bearing courses. The criterion of importance is qualifying for regular credit-bearing courses, not admission.

The proposed inferences are not intended to reflect academic requirements for highly selective postsecondary institutions; additional academic requirements for specific majors or pre-professional programs, such as mathematics, engineering, or medicine; or academic requirements applicable to entry into certificate or diploma programs for job training or professional development in postsecondary institutions.

The proposed inferences are focused on the first year of college; they do not support conclusions about college persistence beyond the first year or completion of a degree. The inferences will necessarily apply in general across a broad range of programs and majors, but should not be applied specifically to any particular program or major.

GPA of B- or better. The selection of “first-year GPA of B- or better” as a referent was made because of its use as a research-based criterion in defining college readiness benchmarks developed by an acknowledged leader in college testing programs—the College Board. The College Board had agreed to partner with the Governing Board in a study linking performance on 12th grade NAEP with the SAT. Another leader in college testing programs, ACT, Inc., has developed similar benchmarks for its college admission assessments using a similar criterion and similar

methodology. Because they are based on credible research related to college outcomes, and because performance on the respective tests could be linked to performance on NAEP, the college readiness benchmarks used by these testing programs were embraced as relevant, useful points of reference for the NAEP preparedness research.

The College Board has set a score of 500 on the SAT Mathematics and Critical Reading tests as its college readiness benchmarks in those areas. Based on its research, the College Board has determined that the score of 500 predicts, with a probability of .65, attainment of a first-year overall GPA of B- or higher (Wyatt, Kobrin, Wiley, Camara & Proestler, 2011, p. 5). Similarly, the ACT college readiness benchmarks are based on research indicating a .50 probability of attaining first-year grades in relevant courses (e.g., college algebra and courses requiring college-level reading) of B or better and .75 probability of C or better (Allen & Sconing, 2005, pp. 2-3, see also <http://www.act.org/solutions/college-career-readiness/college-readiness-benchmarks/>).

The proposed inferences are not intended to convey that a B- or any particular grade should be deemed a standard or goal for postsecondary student outcomes. This criterion was selected to foster comparability across the preparedness research studies, where applicable. However, it does seem self-evident that achieving a first-year GPA of B- or better, without enrollment in remedial/developmental courses, lends support to the likelihood of having possessed academic preparedness for first-year college courses upon entry to college.

Data limitations. Although the preparedness research studies are comprehensive and the results consistent and mutually confirming, for reading they are limited to one year for data at the national level and to one state-based longitudinal study. For mathematics, there are two separate years of data at the national level and one state-based longitudinal study. Therefore, more evidence exists to support the plausibility of inferences related to mathematics than to reading.

Preparedness for job training. The completed research with respect to academic preparedness for job training does not support conclusions relative to the NAEP scale and will not be addressed at this time.

Discussion of the Claims and Evidence

1. The 12th grade NAEP results in reading and mathematics provide unbiased, accurate estimates of the percentages of students at or above specified score levels on the NAEP scales in reading and mathematics for 12th grade students in the United States.

The proposed inferences are premised in part on the capability of NAEP to report percentages of students scoring at or above a certain score on the NAEP 12th grade reading and mathematics scales. The technical qualities of the NAEP scales make them well suited to this purpose.

The NAEP sampling, scaling, IRT modeling, and statistical procedures are widely accepted and well documented (for example, see National Center for Education Statistics, 2010, pp. 70-71), and have been periodically evaluated over two decades (for example, see complete list of research conducted by the NAEP Validity Studies Panel at http://www.air.org/reports-products/index.cfm?fa=viewContent&content_id=890 and “Evaluation of the National Assessment of Educational Progress: Study Reports” at <http://www2.ed.gov/rschstat/eval/other/naep/naep-complete.pdf>).

Other than issues relating to the comparability among the state-level NAEP samples of inclusion rates of students with disabilities and students who are English language learners (about which the Governing Board and NAEP have taken and continue to take significant action), there is little dispute about the appropriateness of the NAEP sampling, scaling and statistical procedures for estimating the percentage of students scoring at or above a selected NAEP scale score.

This is relevant because the proposed inferences that are the subject of this validity argument are interpretations to add meaning to the reporting of NAEP 12th grade reading and mathematics results at particular score levels. The percentages of students at or above particular score levels (e.g., the NAEP achievement levels) have been estimated with accuracy and reported regularly, beginning with assessments in 1992. The proposed inference for reading would use the cut score for 12th grade Proficient as the basis for reporting. The proposed inference for mathematics would use the score of 163 on the NAEP 12th grade scale as the basis

for reporting, which is between the Basic and Proficient achievement levels. Clearly, reporting NAEP results using the proposed inferences will not impair the accuracy of the estimates of the percentages of students scoring at or above the identified points on the NAEP score scales.

2. Performance on 12th grade NAEP assessments in mathematics and reading is positively related to other measures associated with outcomes reflecting academic preparedness for college.

In designing the NAEP preparedness research program, the Governing Board determined that it would be essential to examine how performance on NAEP relates to performance on other measures and outcomes associated with academic preparedness for college. The research program studied the relationship between performance on NAEP and performance on the SAT and ACT college admission tests, including the respective college readiness benchmarks that had been established by these testing programs. The data sources for these analyses were: the NAEP/SAT linking studies (see report at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/SAT-NAEP_Linking_Study.pdf), the Florida longitudinal study (see report at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/Florida_Statistical_Study.pdf), and the 2005 and 2009 NAEP High School Transcript Studies (see <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007467> and <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011462>). Further analyses of these data are found in

Appendix E. An additional data source was the Governing Board's survey of postsecondary education institutions' use of tests and the cut scores on those tests for determining whether incoming students need remedial instruction in reading and mathematics (Fields & Parsad, 2012).

In addition, the research program examined directly the relationship between performance on NAEP and postsecondary outcomes (e.g., first-year college GPA and placement into remedial/developmental courses versus regular credit-bearing courses), analyzing data from the Florida longitudinal study.

The results of these studies will be discussed both in this section and the next section of the validity argument. In this section, background is provided on the indicators that were examined and the results of the NAEP/SAT linking study. The NAEP/SAT linking study is discussed in this section because, as the most recent large-scale national study, it serves as a focal point for discussing the results of the other studies. Thus, in section 3, the results of the other statistical linking studies are discussed in relation to the NAEP/SAT linking study.

Indicators: College Board and ACT College Readiness Benchmarks

The College Board and ACT, Inc. have established college readiness benchmarks for the SAT and the ACT in a number of subjects tested, including reading and mathematics. The SAT College Readiness Benchmark for critical reading and mathematics is a score of 500 on the respective tests. According to the College Board's research, a score of 500 predicts, with a .65 probability, a

first-year GPA of B- or better. The ACT College Readiness Benchmark for reading is a score of 21.² According to ACT's research, the benchmark score predicts, with a .50 probability, a grade of B or better (or .75 probability of a C or better) in first-year courses requiring college reading, such as history and the social sciences. A score of 22 on the ACT mathematics tests predicts a .50 probability of a grade of B or better in a first-year mathematics course, or a .75 probability of a grade of C or better. The College Board research and the ACT research are based on the first-year outcomes of their respective test takers.

Indicators: First-Year GPA of B- or Better and Remedial/non-Remedial Placement

The Governing Board has a partnership with the state of Florida as a part of the Board's program of preparedness research. Florida was one of 11 states that volunteered to provide state-representative samples of 12th grade students for the 2009 NAEP reading and mathematics assessments. Under the partnership, the Florida 12th grade sample is being followed through the postsecondary years via the highly developed Florida longitudinal education data system. For comparability with the SAT College Readiness Benchmarks, the Governing Board analyzed the Florida data to determine the average score and interquartile range for the NAEP test takers with a first-year GPA of B- or better. In addition, the Governing Board analyzed the Florida data to determine the average score and interquartile range for the NAEP test takers who were and who were not placed into remedial reading or remedial mathematics in their first year of college.

² In 2013, after the NAEP research described herein was completed, the ACT College Readiness Benchmark for Reading was changed from 21 to 22.

Analysis of Results for Mathematics

The statistical linking study examining performance on the NAEP 12th grade mathematics assessment and performance on the SAT mathematics test yielded a correlation of .91. This high correlation clearly supports inferences about NAEP performance in relation to SAT performance. The study also examined how performance on NAEP relates to the SAT College Readiness Benchmark for mathematics (i.e., a score on the SAT mathematics test of 500). The SAT benchmark provides “an indication of college readiness at a typical college” (College Board, 2011, p. 4). This is consistent with the Governing Board’s working definition of academic preparedness cited previously:

Academic preparedness for college refers to the reading and mathematics knowledge and skills needed to qualify for placement into entry-level, credit-bearing, non-remedial courses that meet general education degree requirements in broad access 4-year institutions and, for 2-year institutions, for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

The SAT College Readiness Benchmark for mathematics is relevant to student outcomes in college, for it is “the SAT score associated with a 65 percent probability of earning a first-year GPA of 2.67 (B-) or higher” (Wyatt, Kobrin, Wiley, Camara & Proestler, 2011, p. 5). The average NAEP score of students scoring at the College Readiness Benchmark for mathematics was 163 (see Figure 1). As will be demonstrated in the discussion of the third claim, there are additional data corroborating

this level of performance on the 12th grade NAEP mathematics assessment to outcomes in college.

Analysis of Results for Reading

The statistical linking study examining performance on the NAEP 12th grade reading assessment and the SAT critical reading test resulted in a correlation of .74. Although it may not be high enough to predict the performance of individual students from one test to another (which is not required to support the proposed inference for reading), it is sufficient to support the group-level inferences reported by NAEP.

Performance on NAEP was also examined in relation to the SAT College Readiness Benchmark for critical reading (i.e., a score on the SAT critical reading test of 500). The SAT benchmark provides “an indication of college readiness at a typical college” (College Board, 2011, p. 4). This is consistent with the Governing Board’s definition of academic preparedness discussed in the results for mathematics above.

The SAT College Readiness Benchmark for critical reading is relevant to student outcomes in college, for it is “the SAT score associated with a 65 percent probability of earning a first-year GPA of 2.67 (B-) or higher” (Wyatt et al., 2011, p. 5). The average NAEP score of students scoring at the College Readiness Benchmark for reading was 301 (see Figure 2). As will be demonstrated in the discussion of the third claim, there are additional data corroborating this level of performance on the 12th grade NAEP reading assessment to outcomes in college.

3. There is a point on the NAEP scale that corresponds to other measures, indicators, and outcomes associated with academic preparedness for college (i.e., possession of a specific level of academic proficiency, attainment of a first-year overall college GPA of B- or better, and placement into entry-level, credit-bearing, non-remedial college courses).

In addition to the NAEP/SAT Linking Studies (NSLS) described above, analyses were conducted using data from several other studies. A high degree of convergence was found across the studies. The results are described below, first for mathematics and then for reading.

Analysis of Results for Mathematics

Companion statistical relationship studies to the NSLS for mathematics examined data from the 2005 and 2009 national NAEP High School Transcript Studies (HSTS) and from a longitudinal study under a partnership with the Florida Department of Education (FLS). In 2009, Florida was one of 11 states that volunteered to participate in 12th grade state NAEP in reading and mathematics. Using the highly developed Florida longitudinal database, the students in the 12th grade NAEP samples were followed into postsecondary public institutions.

Analyzing data from the transcripts of NAEP test takers, the HSTS examined performance on 12th grade NAEP mathematics in relation to performance in mathematics on the SAT and ACT college admissions tests in 2005 and 2009. The FLS study examined performance on the NAEP 12th grade mathematics assessment in relation to the SAT and ACT college readiness benchmarks, first-

year overall college GPA, and whether students were placed into non-remedial college courses. The study results are displayed in Figure 1.

The focal point for the discussion of these results is the 2009 NSLS because it is the most recent of the national studies. The average NAEP score is 163 for students with an SAT score at the College Readiness Benchmark for mathematics of 500.

The other study results are consistently convergent with the NSLS results. The average NAEP mathematics scores for 12th grade students scoring at the SAT College Readiness Benchmark of 500 for mathematics are compared first for the 2005 HSTS and the 2009 NSLS. The average scores are 161 and 163, respectively.

These results are confirmed by the FLS. The average NAEP mathematics score for the 12th grade Florida NAEP test takers who scored at the SAT College Readiness Benchmark of 500 was 160, much like the 2009 NSLS results and the 2005 HSTS results.

As discussed elsewhere in this validity argument, the ACT College Readiness Benchmark for mathematics is defined somewhat differently than the SAT College Readiness Benchmark for mathematics. However, it is noteworthy that even with this different definition, the results from the 2005 HSTS, 2009 HSTS, and 2009 FLS analyses for the ACT (169, 166, and 164, respectively) are consistent and very similar to the results for the 2009 NSLS.

To answer the question—“What is the relationship between performance on NAEP and actual student outcomes?”—we look to the

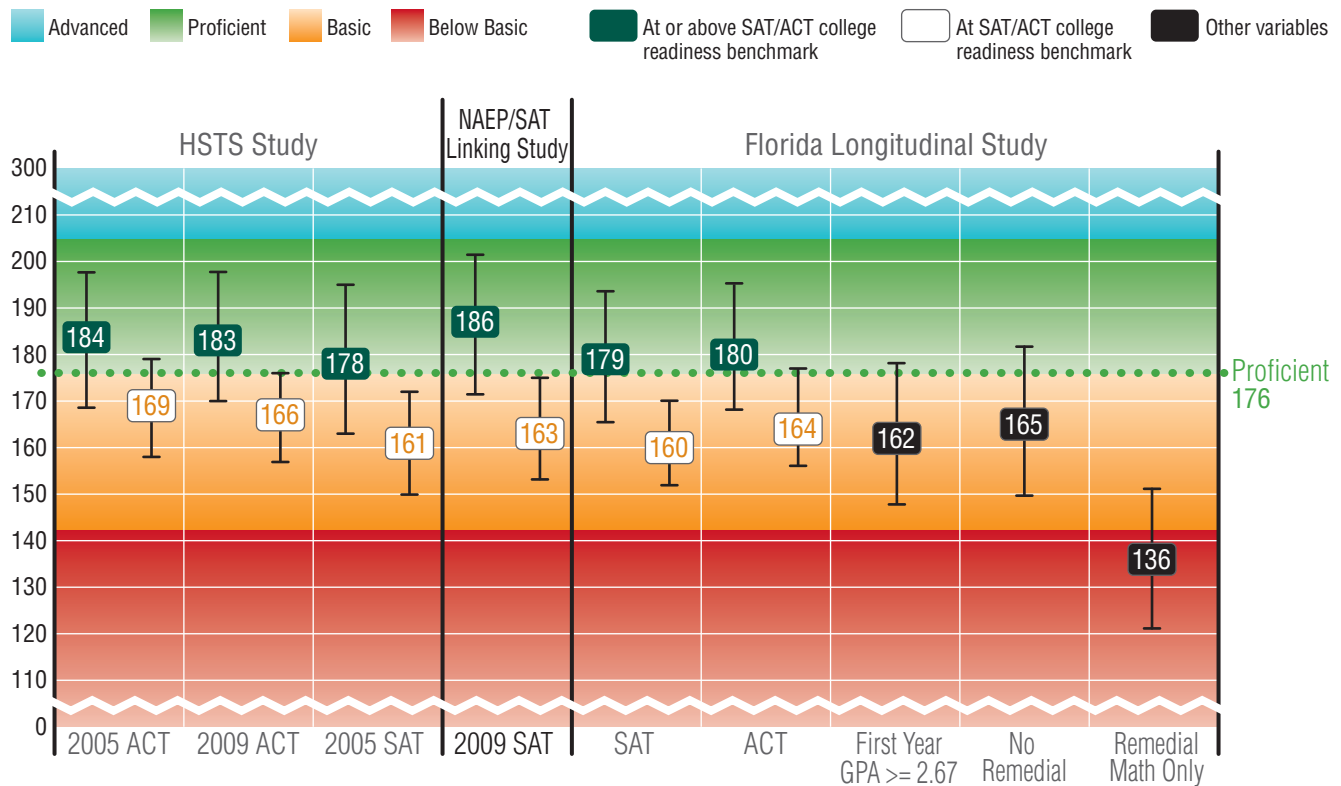
FLS results. First we examine the average NAEP mathematics score for the 12th grade Florida NAEP test takers who attained a first-year GPA of B- or better. The average NAEP score for these students was 162. This is consistent with the SAT College Readiness Benchmark analyses and further

supports the inference that students at or above 163 on the 12th grade NAEP mathematics scale are likely to be academically prepared and attain a first-year GPA of B- or better. It follows, of course, that students who are academically prepared will not require remedial courses.

Figure 1

NAEP 12th Grade Preparedness Research: Mathematics

Average Scores and Interquartile Ranges on NAEP for Selected Variables and SAT/ACT College Readiness Benchmarks, from the 2009 NAEP/SAT Linking Study, 2005 High School Transcript Study, 2009 High School Transcript Study, and 2009 Florida Longitudinal Study



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, *NAEP High School Transcript Study (2005, 2009)*, previously unpublished analyses, March 2013; National Assessment Governing Board, *Statistical Linking of National Results from NAEP and SAT (2009)*, previously unpublished analyses, March 2013; National Assessment Governing Board, *Longitudinal Statistical Relationships for Florida NAEP Examinees: First-Year College Performance Outcomes (2009–2010)*, previously unpublished analyses, March 2013.

Thus, another outcome of interest is placement of entry-level students into remedial college courses versus non-remedial credit-bearing courses. Here again, we look to the FLS as a data source. The average NAEP mathematics score was 165 for the Florida NAEP test-takers not placed into remedial courses, which is consistent with the NSLS score of 163 on the NAEP 12th grade mathematics scale. Furthermore, the average NAEP score of students who were placed into remedial mathematics courses in college was 136, much lower and significantly different from the NSLS score of 163.

The FLS results, together with the SAT and ACT analyses, lend support to the conclusions that students scoring at or above 163 on the 12th grade mathematics scale are likely to be academically prepared for ECRNG college courses and not likely to need remedial courses in mathematics.

These convergent, consistent results across years and across studies support the proposed inference that the percentage of students scoring at or above a score of 163 on the Grade 12 NAEP scale in mathematics is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics that would make them academically prepared for college.

Analysis of Results for Reading

The companion statistical relationship study to the NSLS for reading examined data from a longitudinal study under a partnership with the Florida Department of Education (FLS). In 2009, Florida was one of 11 states that volunteered to participate in 12th grade state NAEP in reading and mathematics. Using the highly developed Florida longitudinal database, the students in the 12th grade NAEP samples were followed into postsecondary public institutions.

The FLS study examined performance on the NAEP 12th grade reading assessment in relation to the SAT and ACT college readiness benchmarks for reading, first-year overall college GPA, and whether students were placed into non-remedial college courses. The study results are displayed in Figure 2.

The focal point for the discussion of these results is the 2009 NAEP/SAT Linking Study (NSLS) for reading, because it is the most recent of the national studies. The average NAEP score is 301 for students with an SAT score at the College Readiness Benchmark for critical reading of 500. A NAEP score of 301 in 12th grade reading is not significantly different from the cut score for the 12th grade Proficient achievement level (302).

The FLS results are consistently convergent with the NSLS results. The average NAEP reading score was 299 for the 12th grade Florida NAEP test takers who were not placed into remedial courses in their first year. The average score was 298 for those who had a first-year overall GPA of a B- or better. These data, which show the relationship between performance on NAEP and actual student outcomes, provide strong confirmation that students scoring at or above Proficient on the NAEP 12th grade reading assessment are likely to be academically prepared for ECNRG college courses.

As discussed elsewhere in this validity argument, the ACT College Readiness Benchmark for reading is defined differently than the SAT College Readiness Benchmark for reading. However, it is noteworthy that even with this different definition, the ACT results from the 2009 FLS analysis are similar to the NSLS analysis and the FLS outcome data.

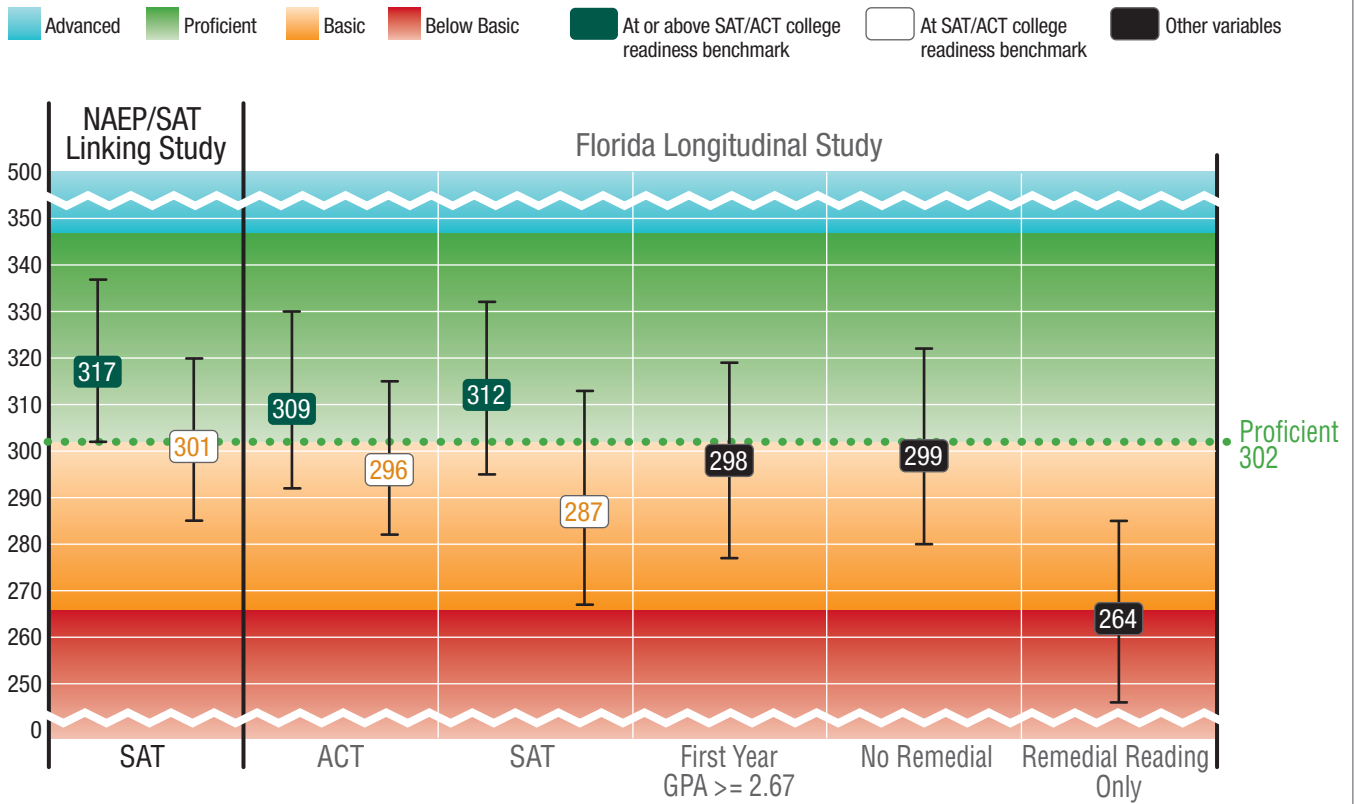
Taken together, these results support the inference that students scoring at or above Proficient on the NAEP 12th grade reading scale are likely to be academically prepared for ECNRG college courses.

In conclusion, these results suggest that the percentage of students at or above the Proficient level in reading on 12th grade NAEP would provide a plausible (or reasonable) estimate of the percentage of 12th grade students in the U.S. who are academically prepared for college.

Figure 2

NAEP 12th Grade Preparedness Research: Reading

Average Scores and Interquartile Ranges on NAEP for Selected Variables and SAT/ACT College Readiness Benchmarks, from the 2009 NAEP SAT Linking Study and 2009 Florida Longitudinal Study



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, previously unpublished analyses, March 2013; National Assessment Governing Board, *Statistical Linking of National Results from NAEP and SAT (2009)*, previously unpublished analyses, March 2013; National Assessment Governing Board, *Longitudinal Statistical Relationships for Florida NAEP Examinees: First-Year College Performance Outcomes (2009–2010)*, previously unpublished analyses, March 2013.

4. The positive relationship between NAEP and the other indicators and outcomes is meaningful in terms of academic preparedness for college, not merely a statistical artifact, because the 12th grade reading and mathematics domains measured by NAEP were specifically designed to measure academic preparedness for college.

NAEP Assessment Frameworks Were Revised to Measure Academic Preparedness

The National Assessment Governing Board intentionally revised the NAEP 12th grade reading and mathematics assessment frameworks with the purpose of measuring academic preparedness for college.

On March 5, 2004, the Governing Board accepted the report of the National Commission on NAEP 12th Grade Assessment and Reporting. The Commission recommended that “NAEP should report 12th grade students’ [academic preparedness] for college-credit coursework, training for employment, and entrance into the military” (National Commission on NAEP 12th Grade Assessment and Reporting, 2004, p. 6).

For NAEP to report on 12th grade academic preparedness for college, it must measure relevant content at the 12th grade. The content of each assessment is determined by the NAEP assessment frameworks, which the Governing Board is responsible for developing and approving. Accordingly, the Governing Board decided that the extant NAEP frameworks

intended for the NAEP 2009 reading and mathematics assessments at the 12th grade would be reviewed. The review would identify changes needed to measure 12th grade academic preparedness for college.³ Examples of the changes made are described in the next two subsections.

Assessments at the 12th grade in reading and mathematics are conducted at least once every four years. In 2004, when the Board decided to proceed with the 12th grade academic preparedness initiative, 2009 was the next assessment year in which the 12th grade reading and mathematics assessments could be affected by framework changes.

In September 2004, the Governing Board contracted with Achieve, Inc. (Achieve) to review the NAEP 12th grade reading and mathematics assessment frameworks and identify where changes, if any, would be needed. Achieve had established the American Diploma Project (ADP) “...to improve postsecondary preparation by aligning high school standards, graduation requirements and assessment and accountability systems with the demands of college and careers” (see www.achieve.org/adp-network). The ADP had conducted research to identify key competencies in English and mathematics needed for high school graduates who aspire to higher education. They refer to these as the “ADP benchmarks.” The type of colleges that were the target for the ADP research was similar to the “typical colleges” in the Governing Board’s research. These were the “two- and four-year colleges and universities in each of the ADP partner states...[that] enroll the vast majority

³ The review also addressed academic preparedness for job training, but that part of the NAEP preparedness initiative is not being addressed in this validity argument.

of high school graduates going on to college: community colleges, as well as four-year state institutions, but generally not the more highly selective ‘flagship’ campuses” (American Diploma Project, 2004, p. 107).

The research and expertise of the American Diploma Project was widely accepted and was brought to bear in reviewing the NAEP frameworks for 12th grade reading and mathematics. Achieve convened a panel of nationally recognized experts in reading and a panel of nationally recognized experts in mathematics. The panels comprised individuals from the K-12, postsecondary, research, and policy spheres, who were knowledgeable about academic preparedness for college reading and college mathematics. The panels compared the 12th grade NAEP reading and mathematics frameworks and the ADP benchmarks.

Reading

The Achieve reading panel found considerable similarity between NAEP and the ADP benchmarks for English, although not perfect agreement. This is displayed in the side-by-side chart on pages 30-40 of the Achieve Reading Report (Achieve, 2005). The English benchmarks have eight major components and objectives under each component. Three of these major components were deemed “Not Applicable” to the reading domain: writing, research, and media.

For almost all of the applicable objectives under the five major components that were applicable to the reading domain, the Achieve reading panel found matches in the NAEP 2009 reading framework. Overall, the panel concluded that “... the 2009 NAEP Reading Framework...was aligned to the ambitious [ADP] benchmarks” (Achieve, 2005, p. 2).

The reading panel also listed items in the NAEP framework that are not found in the ADP English benchmarks. For example, under Argumentation and Persuasive Text, figurative language and rhetorical structure, including parallel structure and repetition, were present in the NAEP reading framework at grade 12, but not in the ADP benchmarks. Under Poetry, tone, complex symbolism, and extended metaphor and analogy were present in the NAEP reading framework but not in the ADP benchmarks. A complete listing of the items in the NAEP framework not present in the ADP benchmarks appears on page 41 of the Achieve Reading Report.

Although the Achieve reading panel concluded that the 12th grade NAEP reading framework for 2009 was aligned with the ADP benchmarks applicable to reading, the panel’s report does include six recommendations. The Governing Board approved these recommendations on February 14, 2005. For example, the Achieve reading panel recommended increasing the percentage of informational text passages from 60% to 70% and featuring additional items that ask students to compare texts. The changes were modest, sufficiently so to permit continuation of the 12th grade trend line from its initiation in 1992.

The NAEP reading framework used for the 2009, 2011, and 2013 assessments contains the following statement:

In May 2005, the Governing Board adopted a policy statement regarding NAEP and 12th grade preparedness. The policy states that NAEP will pursue assessment and reporting on 12th grade student achievement as it relates to preparedness for post-secondary education and training. This policy resulted from recommendations of the Board’s National Commission on NAEP 12th

Figure 3

Examples of Objectives Added to the 2009 Grade 12 Mathematics Framework

Number properties and operations	
b)	Analyze or interpret a proof by mathematical induction of a simple numerical relationship.
Measurement	
d)	Interpret and use the identity $\sin^2\theta + \cos^2\theta = 1$ for angles θ between 0° and 90° ; recognize this identity as a special representation of the Pythagorean theorem.
e)	Determine the radian measure of an angle and explain how radian measurement is related to a circle of radius 1.
f)	Use trigonometric formulas such as addition and double angle formulas.
g)	Use the law of cosines and the law of sines to find unknown sides and angles of a triangle.
Geometry	
e)	Use vectors to represent velocity and direction; multiply a vector by a scalar and add vectors both algebraically and graphically.
g)	Graph ellipses and hyperbolas whose axes are parallel to the coordinate axes and demonstrate understanding of the relationship between their standard algebraic form and their graphical characteristics.
h)	Represent situations and solve problems involving polar coordinates.
Data Analysis, Statistics, and Probability	
c)	Draw inferences from samples, such as estimates of proportions in a population, estimates of population means, or decisions about differences in means for two “treatments.”
e)	Recognize the differences in design and in conclusions between randomized experiments and observational studies.
k)	Use the binomial theorem to solve problems.
e)	Recognize and explain the potential errors caused by extrapolating from data.
Algebra	
e)	Identify or analyze distinguishing properties of linear, quadratic, rational, exponential, or trigonometric functions from tables, graphs, or equations.
j)	Given a function, determine its inverse if it exists and explain the contextual meaning of the inverse for a given situation.
h)	Analyze properties of exponential, logarithmic, and rational functions.
g)	Determine the sum of finite and infinite arithmetic and geometric series.

Grade Assessment and Reporting in March 2004. Subsequent studies and deliberations by the Board took place during 2004 and 2005.

In reading, the Board adopted minor modifications to the 2009 NAEP Reading Framework at grade 12 based on a comprehensive analysis of the framework conducted by Achieve, Inc. The current version of the reading framework incorporates these modifications at grade 12 to enable NAEP to measure and report on preparedness for postsecondary endeavors (National Assessment Governing Board, 2008b, p. v).

Mathematics

The mathematics review began with the 2007 NAEP mathematics framework, which was the most current and included the changes approved for the 2005 12th grade mathematics assessment. The Achieve panel examined the NAEP mathematics framework at the 12th grade in relation to the ADP benchmarks for mathematics. The Achieve panel developed proposed revisions to the assessment objectives for grade 12. While acknowledging differences in language and purpose, the Achieve mathematics panel concluded that the “overall mathematics frameworks of ADP and [12th grade] NAEP are remarkably similar” (Achieve, 2006, p. 9).

The Governing Board convened a panel of mathematicians and mathematics educators to review and revise the objectives in relation to the objectives for grades 4 and 8. The panel conducted focus groups with various NAEP constituents, using repeated rounds of reviews. The Governing Board approved the final set of grade 12 objectives on August 5, 2006 (see Figure 3). The changes to the framework were sufficiently modest to permit the continuation of the 12th grade trend line begun with the 2005 12th grade mathematics assessment under the previous 12th grade framework.

Like the reading framework, the 2009/2013 mathematics framework for grade 12 states the Board’s intention to measure 12th grade academic preparedness (National Assessment Governing Board, 2008a, pp. 2-3).

Conclusion

The Governing Board, by official action, revised the NAEP 12th grade reading and mathematics frameworks with the explicit purpose of measuring 12th grade academic preparedness for college, beginning with the 2009 assessments. Setting forth the measurement purpose and making relevant revisions to the NAEP assessment frameworks are necessary elements of the validity argument; however, they are not sufficient. Evidence must be considered with respect to the alignment of the framework and the test questions administered to the measurement purpose. This will be addressed in the next section.

Content Alignment Studies Found Significant Overlap between NAEP and the ACT, SAT and ACCUPLACER

The Governing Board conducted studies to determine the degree of content similarity between NAEP 12th grade reading and mathematics assessments and relevant tests used for college admissions and placement.

The studies had two objectives. The first objective was to determine the degree to which the content of 12th grade NAEP assessments in reading and mathematics covers the reading and mathematics knowledge and skills needed for first-year college work. The SAT, ACT, and ACCUPLACER are well-established tests that assess individual students’ reading and mathematics proficiency in relation to college-level expectations.

The ACT is developed with the purpose of “...[measuring] as directly as possible the degree to which each student has developed the academic skills and knowledge that are important for success in college...” (ACT, 2007, p. 62).

The SAT is developed “to ensure that the topics measured on the SAT...reflect what is being taught in the nation’s high schools and what college professors consider to be required for college success” (Kim, Wiley & Packman, 2013, p. 3).

The ACCUPLACER has the purpose of “...determining if students are prepared for a college-level course or if they would benefit from a developmental course” (College Board, 2012, p. 2).

The SAT, ACT and ACCUPLACER in reading and mathematics are widely used for these purposes by admissions and placement professionals in postsecondary education institutions. These testing programs regularly conduct curriculum surveys, validity studies and other research to support their claims that the content measured is directly related to the reading and mathematics knowledge and skills needed to qualify for entry-level, credit-bearing courses (e.g., see the ACT curriculum studies for 2012, 2009, 2005, and 2002 at <http://www.act.org/research-policy/national-curriculum-survey/>, and the College Board *National Curriculum Survey on English and Mathematics* at <http://research.collegeboard.org/publications/content/2012/05/national-curriculum-survey-english-and-mathematics>).

Therefore, with the assumption that the SAT, ACT, and ACCUPLACER do measure the content needed for college-level work, significant content overlap between NAEP and these other assessments would support the conclusion that what NAEP measures covers the knowledge and

skills needed by college freshmen to be placed into entry-level, credit-bearing courses.

The second reason for conducting the content alignment studies was to provide information for interpreting the results of the statistical linking studies between NAEP and the other tests (discussed earlier), which measure academic preparedness for college. The linking studies were designed to examine how performance on NAEP compares with performance on the other tests, with the purpose of supporting inferences about academic preparedness for college. For NAEP to support inferences about academic preparedness for college based on the linking studies, a sufficient content match would be needed between NAEP and the other tests, not just a statistical relationship.

The Content Alignment Studies: Overview

The Governing Board conducted content alignment studies in reading and mathematics comparing the 2009 12th grade NAEP and the ACT, SAT, and ACCUPLACER reading and mathematics tests. Overall, considerable overlap was found between the ACT and NAEP and the SAT and NAEP, with some differences. NAEP was found to measure much of what is measured on the ACCUPLACER, but the reading and mathematics domains measured by NAEP were much broader than ACCUPLACER. More details are provided in the summaries of the individual studies below.

The general design for the content alignment studies was to compare the 12th grade NAEP frameworks in reading and mathematics with the analogous document for the other test, and then to compare the test items from one test to the framework/analogous document of the other test. The reviews were performed by subject-specific

(i.e., mathematics, reading) panels, composed of experts in mathematics or reading and English instruction at the high school and college levels.

Alignment studies that compare an assessment to the content standards on which it is based are relatively common and have well-established methodologies. However, this is not true for the types of alignment studies the Governing Board planned to conduct: content alignment studies comparing different assessment programs. Different assessment programs have different purposes, different approaches to describing the domain being measured, and, possibly, different “grain size” in the level of detail in describing the domain.

The Governing Board contracted with Norman Webb, a noted expert in content alignment studies, to prepare a design document for conducting the assessment to assessment alignment studies. The purpose was to put in place a methodology that considered the special challenges of assessment to assessment alignment studies and to foster comparability in the conduct of the studies and the reporting metrics across studies and contractors. (The link to the Webb design document is at <http://www.nagb.org/content/nagb/assets/documents/publications/design-document-final.pdf>). The Webb design was developed after the ACT alignment studies were completed. It was used in conducting the SAT and ACCUPLACER content alignment studies.

In the following sections are summaries of the content alignment study results, excerpted from the study reports. The results for the three content alignment studies in reading are presented first, followed by the three content alignment studies for mathematics, along with summary discussions for the reading and mathematics results.

The Content Alignment Studies: Reading Results

Reading: ACT. The Governing Board contracted with ACT, Inc. to conduct the content alignment study comparing the NAEP 12th grade reading assessment and the ACT reading test. The full report can be found at <http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACT-NAEP Math and Reading Content Comparison.pdf>.

The reading panel was composed of seven members with expertise in reading and/or English instruction at the high school and college levels. The panel was about evenly divided in terms of prior familiarity with either the ACT or NAEP reading domains.

The panel found considerable similarity in the content of the NAEP 12th grade reading assessment and the ACT. For example, the NAEP 12th grade reading framework was compared to the ACT reading domain and the ACT College Readiness Standards for reading. The ACT College Readiness Standards (CRS) are descriptions of the content (i.e., the knowledge and skills) measured by the ACT reading test in score bands along the ACT 1-36 point scale from 13-36 (see <http://www.act.org/standard/planact/reading/>).

The panel concluded that

All of the skills highlighted in the ACT [reading] domain and in the [ACT] College Readiness Standards [for reading] were identified within the NAEP Reading framework. In performing the comparison in the other direction—NAEP-to-ACT—it was the sense of the panel that the ACT measured primarily those skills that NAEP identifies as Locate/Recall and Integrate/Interpret skills, those that

pertain primarily to finding explicit information in text (what the ACT would call Referring skills) and to making inferences, drawing conclusions, and making generalizations from information within text (what the ACT would call Reasoning skills). The panel saw less evidence of the higher-level analytical and evaluative Critique/Evaluate skills in the ACT domain, and attributed that to the multiple-choice format of the ACT [whereas NAEP includes constructed response items as well as multiple-choice]. 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 is not a feature of the ACT. So, 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 (National Assessment Governing Board, 2009b, p. 17).

The reading panel also conducted an item classification study, in which the NAEP 12th grade reading items were classified in relation to the ACT College Readiness Standards for Reading:

A total of 152 Reading items (comprising 17 blocks) were classified in [the reading] study. Of these, 97 were multiple-choice (MC). Nine were dichotomously-scored (“incorrect” or “correct”) short constructed-response (DSCR) items. Thirty-three were polytomously-scored short constructed-response (PSCR) items, each scored using a three-point scoring rubric. Thirteen were extended constructed-response (ECR) items, each scored using a four-point rubric. Each DSCR had one creditable score category, each PSCR had two, and each ECR had three. Each Reading panelist,

therefore, assigned a total of 211 classifications to the NAEP Reading items [and rubric scoring categories] (p. 54).

An item or score category was deemed “classified” if there was majority agreement; that is, if at least four of the seven panel members agreed about the score band to which an item (or creditable score category under an item rubric) was assigned.

Of the 211 determinations to be made, there was only one for which there was no majority agreement (the assignment of a PSCR rubric to a CRS score band). Of the remaining 210 determinations, 181 were unanimous.

The reading panel was able to classify 137 items or rubric categories (about two-thirds of the determinations to be made) to the CRS score bands. Of the 97 multiple-choice items, 81 (or 84%) were classified. Of the 113 rubric score categories for items, 56 (or 50%) were classified. The reasons some multiple-choice items and rubric score categories could not be classified were related to the differences in the ACT and NAEP reading domains described above. These reasons include the presence of constructed response items in NAEP but not the ACT, the presence of items involving multiple texts in NAEP but not the ACT, and the greater presence of “Critique/Evaluate” type items in NAEP than the ACT.

Of the 137 classifications, 24 were in the score bands from 13-19; 113 of the classifications were in the score bands from 20-36. This is noted because the ACT College Readiness Benchmark for reading is 21. The ACT College Readiness Benchmark signifies the score at which a student has a 50% chance of attaining a grade of B or

better in a relevant subject and a 75% change of a C or better. In addition, the Governing Board conducted a survey of postsecondary institutions' use of tests in making entry-level decisions about placement into remedial or regular credit-bearing courses. With respect to the ACT, 18 was the mean reading score below which students were deemed to need remedial course work (Fields & Parsad, 2012, p. 19). Whereas this provides a context for the study results, it must be kept in mind that in making their judgments about item classifications, the panelists did not have data about NAEP item difficulty or data on how performance on NAEP compares with performance on the ACT.

Finally, although the study results support the conclusion that the 12th grade NAEP reading assessment measures content directly related to academic preparedness for college, it is noted that the study was conducted by ACT, Inc., not an independent third party. Further, because a different methodology was used, the study results are not directly comparable to the results for the

SAT and ACCUPLACER alignment studies in reading.

Reading: SAT. The Governing Board contracted with WestEd, an independent third party, to conduct the content alignment study comparing the NAEP 12th grade reading assessment and the SAT critical reading test. WestEd conducted the content alignment study using the design developed for the Governing Board by Norman Webb. The full report of the content alignment study can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/SAT-NAEP_Reading_Content_Comparison.pdf.

Overall, the study found similar content in the NAEP 12th grade reading assessment and the SAT critical reading test.

The following highlighted content is an excerpt from the Executive Summary of the report (pp. iv-vi):

Alignment of 2009 NAEP Grade 12 Reading and SAT Critical Reading, Executive Summary (Excerpt)

What is the correspondence between the reading content domain assessed by NAEP and that assessed by SAT?

The greatest commonality between the two tests is their shared emphasis on the broad skills of integrating and interpreting both informational and literary texts. This is evident in the majority of items from both tests aligned to NAEP Standard 2, "Integrate/Interpret," including many to Goal 2.1, "Make complex inferences within and across both literary and informational texts."

Despite the difference in the degree of specificity of the two frameworks (most NAEP objectives are much more finely grained than the SAT objectives), there is also considerable overlap at the level of more specific skills.

[continued]

Alignment of 2009 NAEP Grade 12 Reading and SAT Critical Reading, Executive Summary (Excerpt) *continued*

To what extent is the emphasis of reading content on NAEP proportionally equal to that on SAT?

Both tests had many of their item alignments to the same NAEP “Integrate/Interpret” objectives, often with similar percentages of alignments. Although there were some differences in emphasis, both tests also had notable percentages of alignments to SAT Objectives B.1.1–B.1.3 and B.1.5. Skills with overlap include inferring/analyzing the following:

- the “main idea” and “author’s purpose” (SAT Objective B.1.1 and NAEP Objectives 2.3.a and 2.1.f);
- the “tone and attitude” of an author or character (NAEP Objectives 2.2.a and 2.2.c and SAT Objective B.1.4);
- the use of “rhetorical strategies” (NAEP Objective 2.1.d and SAT Objective B.1.2); and
- connections between ideas, perspectives, or problems (NAEP Objective 2.1.b and SAT Objectives B.1.3 and B.1.5).

Additionally, in the area of greatest content overlap—items on both tests aligned to objectives for NAEP “Integrate/Interpret” and aligned to SAT “Passage-Based Reading” Objectives B.1.1– B.1.5—both tests met the typical threshold criteria for depth of knowledge consistency...

Despite these similarities, there are some notable differences in emphasis between the two assessments. Both tests assess vocabulary skills. However, NAEP addresses vocabulary exclusively in the context of passage comprehension, while the majority of SAT vocabulary items are in a sentence-completion format, in which context plays a more limited role. This difference reflects NAEP’s emphasis on the understanding of word meaning in context; the assessment is not intended to measure students’ prior knowledge of word definitions. The SAT sentence-completion items provide some context within the single sentence text, but in many cases, students’ success on the items almost certainly depends on their prior knowledge of word definitions.

In addition, panelists found considerably less emphasis in SAT than in NAEP on literal comprehension and critical evaluation, particularly the evaluation of the quality or effectiveness of an author’s writing, skills covered in the NAEP standards “Locate/Recall” (locating/recalling specific details and features of texts) and “Critique/Evaluate” (evaluating texts from a critical perspective), respectively. This difference suggests a greater emphasis on these skills in NAEP.

Even with the minimal coverage of NAEP “Locate/Recall” and “Critique/Evaluate” standards by SAT items, all NAEP items found a match in the SAT framework. However, the broad language of the SAT framework can

[continued]

Alignment of 2009 NAEP Grade 12 Reading and SAT Critical Reading, Executive Summary (Excerpt) *continued*

encompass the range of the NAEP items. For example, SAT Goal B.2, “Literal Comprehension,” refers to items that “ask what is being said” in a “small but significant portion of a reading passage,” a description that can easily accommodate most NAEP “Locate/Recall” items and objectives. In fact, nearly all items on the NAEP short version that were coded to “Locate/Recall” objectives in the NAEP framework were matched to SAT Goal B.2 in the SAT framework.

Similarly, SAT Objective B.1.3, to which approximately one-quarter of NAEP items aligned, includes “Evaluation,” the primary focus of NAEP “Critique/Evaluate.” The description in SAT Objective B.1.3 of items that “ask the test taker to evaluate ideas or assumptions in a passage” is compatible at a very general level with NAEP “Critique/Evaluate” objectives addressing the author’s point of view, logic, or use of evidence. SAT Objective B.1.2, “Rhetorical Strategies,” is also broad enough in its language to make it a reasonable match for some NAEP “Critique/Evaluate” items focused on “author’s craft” or use of “literary devices.” In the NAEP short version, all items that aligned to “Critique/Evaluate” objectives in the NAEP framework were aligned to either SAT Objectives B.1.2 or B.1.3, or both.

Are there systematic differences in content and complexity between NAEP and SAT assessments in their alignment to the NAEP framework and between NAEP and SAT assessments in their alignment to the SAT framework? Are these differences such that entire reading subdomains are missing or not aligned?

With regard to differences in content as described in the NAEP framework, SAT items had limited coverage of the knowledge and skills described by the NAEP standards “Locate/Recall” and “Critique/Evaluate.” This difference is also reflected in test format, with the use of longer reading passages and both constructed-response and multiple-choice items in NAEP. In comparison, all SAT items are multiple-choice. With regard to differences in content as described in the SAT framework, NAEP does not include sentence-completion items.

With regard to differences in complexity, NAEP items and objectives had a range of depth of knowledge including items at DOK Levels 1, 2, and 3, while SAT items and objectives were coded primarily at Levels 2 and 3.

Overall, the alignment results across the two sets of items and frameworks show a strong area of overlap in their coverage of SAT “Passage-Based Reading” objectives and NAEP “Integrate/Interpret” objectives, as well as some important differences.

Reading: ACCUPLACER. The Governing Board contracted with WestEd, an independent third party, to conduct the content alignment study comparing the NAEP 12th grade reading assessment and the ACCUPLACER reading test. The ACCUPLACER is used specifically to determine whether entry-level students have the reading skills necessary for college-level work or require remedial reading courses. WestEd conducted the content alignment study using the design developed for the Governing Board by Norman Webb. The full report of the content

alignment study can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACCUPLACER-NAEP_Reading_Content_Comparison.pdf.

Overall, the study found similar content in the NAEP 12th grade reading assessment and the ACCUPLACER reading test, although the content of NAEP is much broader and more complex. The following highlighted text is an excerpt from the Executive Summary of the report (pp. iv-vi):

Alignment of 2009 NAEP Grade 12 Reading and ACCUPLACER Reading Comprehension, Executive Summary (Excerpt)

What is the correspondence between the reading content domain assessed by NAEP and that assessed by ACCUPLACER?

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. This is evident in the majority of items on both tests (two-thirds to three-fourths) matched to the NAEP standard “Integrate/ Interpret: Make complex inferences within and across texts.” On both tests, the majority of alignments to “Integrate/Interpret” were to objectives that apply to informational text only or across both informational and literary texts.

The shared emphasis on the comprehension and interpretation of informational text can also be seen in the alignments on both tests to the ACCUPLACER framework. Although the ACCUPLACER standards do not explicitly refer to text type, they focus almost exclusively on elements typical of informational text. A majority of both NAEP and ACCUPLACER items were matched to the ACCUPLACER standard “Inferences,” and both tests had notable percentages of alignments to “Direct statements and secondary ideas” and “Applications.” A smaller percentage of items on both tests were aligned to “Identifying main ideas.”

To what extent is the emphasis of reading content on NAEP proportionally equal to that on ACCUPLACER?

As previously discussed, the alignments both within and across frameworks show that both tests emphasize the comprehension and interpretation of informational text, particularly through the use of inference. Within

[continued]

Alignment of 2009 NAEP Grade 12 Reading and ACCUPLACER Reading Comprehension, Executive Summary (Excerpt) *continued*

this broad area of convergence, however, there are differences in emphasis revealed in the alignments to specific objectives within both frameworks. In relation to the NAEP framework, the NAEP short-version items showed a far greater emphasis on the comprehension of vocabulary in context (Objective 4.a) and on the analysis of an author’s use of language (Objective 1.d). In relation to the ACCUPLACER framework, NAEP items showed more emphasis on the use of inference to interpret text (“Inferences”). The higher percentage of NAEP items aligned to “Applications” also reflects the greater emphasis in NAEP on understanding authors’ use of language.

In relation to the ACCUPLACER framework, the ACCUPLACER items showed a greater emphasis than the NAEP items on the identification of main ideas. In relation to the NAEP framework, the ACCUPLACER items showed more emphasis on the recall of specific details, facts, and information (NAEP 1.1.a).

In general, in the cross-framework alignments, the matches found in each test to the other’s framework (NAEP to ACCUPLACER and ACCUPLACER to NAEP) tended to be for the most general objectives within that framework. For example, the great majority of hits for ACCUPLACER items to NAEP objectives for “Integrate/Interpret” were to two of the most broadly stated NAEP objectives, “Draw conclusions” (2.3.b) and “Compare or connect ideas” (2.1.b). Many of the more specific NAEP objectives for “Integrate/Interpret,” such as “Find evidence in support of an argument” (2.2.c), received far fewer or no hits from ACCUPLACER items. Compared to ACCUPLACER, the NAEP items were more evenly distributed among NAEP objectives.

The majority of alignments for NAEP items to ACCUPLACER standards were also to the broadest of those standards—“Inferences” and “Applications,” both of which overlap in content with a number of NAEP objectives but at a higher level of generality. The more specific ACCUPLACER standard, “Identifying main ideas,” received far fewer alignments from NAEP items.

Are there systematic differences in content and complexity between the NAEP and ACCUPLACER assessments in their alignment to the NAEP framework and between the NAEP and ACCUPLACER assessments in their alignment to the ACCUPLACER framework? Are these differences such that entire reading subdomains are missing or not aligned?

In regard to differences in content, NAEP addresses reading skills related to both literary and informational text, while ACCUPLACER does not address reading skills specific to literary text. As expected, based on the framework-to-specifications [review]... ACCUPLACER items had minimal matches to NAEP objectives for literary text. The main area of alignment of ACCUPLACER items to the NAEP framework, NAEP objectives in “Locate/Recall” and “Integrate/Interpret,” applied to informational text only or to both informational and literary text.

[continued]

Alignment of 2009 NAEP Grade 12 Reading and ACCUPLACER Reading Comprehension, Executive Summary (Excerpt) *continued*

The ACCUPLACER items also had minimal to no coverage of the NAEP standard “Critique/Evaluate.” ... overall, the language of the ACCUPLACER objectives (“understand,” “comprehend,” “recognize”) places more emphasis on comprehension and interpretation of text (“distinguish the main idea from supporting ideas” or “perceive connections between ideas made—implicitly—in the passage”) than on critical analysis or evaluation (“Evaluate the strength and quality of evidence used by the author to support his or her position” in NAEP Objective 3.3.b, or “Judge the author’s craft and technique” in NAEP Objective 3.1.a).

In regard to complexity, both assessments were found to meet the criteria for depth of knowledge consistency in relation to their own framework. In relation to the NAEP framework, however, only the NAEP items met the criteria for DOK consistency for all NAEP standards. The ACCUPLACER items met the criteria for depth of knowledge consistency only for NAEP “Locate/Recall.”

Although the majority of the ACCUPLACER item alignments were to objectives for NAEP “Integrate/Interpret,” over half of these items were found to have a DOK level below that of the standard. In addition, the use of very short reading passages and exclusively multiple-choice items in ACCUPLACER may be less conducive to the more in-depth reasoning required by DOK Level 3. NAEP, by contrast, includes much longer reading passages and both multiple-choice and constructed-response items.

NAEP covers skills specific to the comprehension and analysis of literary text while ACCUPLACER does not. In addition, NAEP covers the skills of evaluating and critiquing text, skills not addressed by ACCUPLACER. Finally, NAEP has a wider range of cognitive complexity than ACCUPLACER, with a substantially higher percentage of items at DOK Level 3, requiring more in-depth analysis or evaluation. However, both tests show a similar emphasis on applying interpretive skills and inferential reasoning to the understanding of informational text.

Overall, the NAEP items covered a broader range of cognitive complexity than the ACCUPLACER items. This is also apparent in the frameworks. The three NAEP standards, defined in terms of three different “cognitive targets” (“Locate/Recall,” “Integrate/Interpret,” and “Critique/Evaluate”), cover a broader range of cognitive complexity supported by the use of longer reading passages and the inclusion of both short and extended constructed-response items. The language of the ACCUPLACER standards (“understand,” “comprehend,” “recognize”) places more emphasis on comprehension and interpretation of text (e.g., “distinguish the main idea from supporting ideas” in ACCUPLACER A, “Identifying main ideas,” or “perceive connections between ideas made—implicitly—in the passage” in ACCUPLACER C, “Inferences”) than on critical analysis or evaluation (e.g., “Evaluate the strength and quality of evidence” in NAEP 3.3.b, or “Judge the author’s craft” in NAEP 3.1.a). In addition, the use of very short reading passages and exclusively multiple-choice items in ACCUPLACER may be less conducive to the cognitive complexity typical of DOK Level 3 items. Although the NAEP items show a greater range of cognitive complexity and a greater emphasis on critical thinking, both tests show a similar emphasis on applying interpretive skills and inferential reasoning to the understanding of informational text.

The Content Alignment Studies: Summary Discussion for Reading

The NAEP 12th grade reading framework, test questions, and, for constructed-response items, the score category rubrics, were compared with the analogous domain descriptions and test questions for the ACT, SAT, and ACCUPLACER reading tests. These three tests are used for college admissions and placement. They are well established and have been used for these purposes for many years by professionals in postsecondary education. The test publishers regularly survey secondary and postsecondary educators about relevant content and have conducted research that supports the validity of the test content for the intended inferences and uses. The underlying assumption is that if the content of the 12th grade NAEP reading assessment is similar to the content of these reading tests, then the NAEP content is directly related to “academic preparedness for college.”

The ACT study found that “All of the skills highlighted in the ACT [reading] domain and in the [ACT] College Readiness Standards [for reading] were identified within the NAEP Reading framework.” At the same time, there was content measured by NAEP that was not present in the ACT reading test. In assigning 211 NAEP 12th grade reading items and rubric score categories to the ACT College Readiness Standards for reading, there were 137 positive classifications, or about 65% of the possible classifications. The multiple-choice items and rubric score categories that could not be classified were those that measured content not measured by the ACT reading test.

The SAT study found that “Overall, the alignment results across the two sets of items and frameworks show a strong area of overlap in their coverage of SAT ‘Passage-Based Reading’

objectives and NAEP ‘Integrate/Interpret’ objectives, as well as some important differences.” With respect to the differences, “...SAT items had limited coverage of the knowledge and skills described by the NAEP standards ‘Locate/Recall’ and ‘Critique/Evaluate.’ This difference is also reflected in test format, with the use of longer reading passages and both constructed-response and multiple-choice items in NAEP. In comparison, all SAT items are multiple-choice. With regard to differences in content as described in the SAT framework, NAEP does not include sentence-completion items.”

The ACCUPLACER study found that “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. This is evident in the majority of items on both tests (two-thirds to three-fourths) matched to the NAEP standard ‘Integrate/Interpret: Make complex inferences within and across texts.’ On both tests, the majority of alignments to ‘Integrate/Interpret’ were to objectives that apply to informational text only or across both informational and literary texts...Overall, the NAEP [frameworks and] items covered a broader range of cognitive complexity than the ACCUPLACER items...The three NAEP standards, defined in terms of three different ‘cognitive targets’ (‘Locate/Recall,’ ‘Integrate/Interpret,’ and ‘Critique/Evaluate’), cover a broader range of cognitive complexity supported by the use of longer reading passages and the inclusion of both short and extended constructed-response items.”

The results across the three studies are consistent. In general, the content of the ACT, SAT, and ACCUPLACER reading tests are present in NAEP, but NAEP is generally broader. Alignment

between NAEP and the other three respective assessments is substantial, but not perfect; perfect alignment is not expected. A component of the SAT critical reading assessment not present in NAEP is sentence completion, measuring vocabulary knowledge in a different way than NAEP does.

These results support the conclusion that

- The NAEP 12th grade reading assessment measures academic knowledge and skills that are also covered by other assessments designed and used to make judgments about the academic preparedness of college freshmen for placement into entry-level, credit-bearing, non-remedial college courses that meet general education degree requirements, and
- NAEP 12th grade reading test items and rubric scoring categories for items are appropriate for obtaining evidence of test takers' possession of knowledge and skills needed for college freshmen to be placed into ECNRG courses requiring college-level reading.

The Content Alignment Studies: Mathematics Results

Mathematics: ACT. The Governing Board contracted with ACT, Inc. to conduct the content alignment study comparing the NAEP 12th grade mathematics assessment and the ACT mathematics test. The full report can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACT-NAEP_Math_and_Reading_Content_Comparison.pdf.

The mathematics panel was composed of seven members with expertise in mathematics instruction at the high school and college levels. The panel was about evenly divided in terms of prior familiarity with either the ACT or NAEP mathematics domains.

The panel found considerable similarity in the content of the NAEP 12th grade mathematics assessment and the ACT. For example, the NAEP 12th grade mathematics framework was compared to the ACT mathematics domain and the ACT College Readiness Standards for mathematics. The ACT College Readiness Standards (CRS) are descriptions of the content (i.e., the knowledge and skills) measured by the ACT mathematics test in score bands along the ACT 1-36 point scale from 13-36 (see <http://www.act.org/standard/planact/math/index.html>). The panel concluded that

... the two assessments have much of their content domains in common. However, 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 addition, 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. With regard to the ACT-to-NAEP comparison, the Mathematics panel found nearly all of the ACT Mathematics domain and College Readiness Standards reflected in the NAEP Mathematics domain, but determined that a number

of the lower-level topics in the ACT Pre-Algebra subdomain were more consistent with Grade 8 NAEP topics. All of these points suggest that while there may be substantial overlap in what the two assessments measure and how they measure it, there are areas of difference, as well (pp. 18-19).

The mathematics panel also conducted an item classification study, in which the NAEP 12th grade mathematics items were classified in relation to the ACT College Readiness Standards for Mathematics.

An item or score category was deemed “classified” if there was majority agreement; that is, if at least four of the seven panel members agreed about the score band to which an item (or creditable score category under an item rubric) was assigned.

Of the 229 determinations to be made, panel members believed that every item or rubric category could be classified to some CRS score range. However, there were 39 for which there was no majority agreement (17 multiple-choice items and 22 rubric categories) on what the classification should be; therefore, those items were not considered assigned to a CRS score band. Of the remaining 190 determinations, 24 were unanimous, 142 involved classifications to adjacent score ranges, and 24 involved classifications to non-adjacent score ranges.

Of the 108 multiple-choice items, 91 (or 84%) were classified. Of the 121 rubric score categories for items, 99 (or 82%) were classified.

Of the 190 classifications, 10 were in the score bands from 13-19; 180 of the classifications were in the score bands from 20-36. This is noted

because the ACT College Readiness Benchmark for mathematics is 22. The ACT College Readiness Benchmark signifies the score at which a student has a 50% chance of attaining a grade of B or better in a relevant subject and a 75% chance of a C or better. In addition, the Governing Board conducted a survey of postsecondary institutions’ use of tests in making entry-level decisions about placement into remedial or regular credit-bearing courses. With respect to the ACT, 19 was the mean mathematics score below which students were deemed to need remedial course work in mathematics (Fields & Parsad, 2012, p. 13). Although this provides a context for the study results, it must be kept in mind that in making their judgments about content, the panelists did not have data about NAEP item difficulty or data on how performance on NAEP compares with performance on the ACT.

Finally, although the study results support the conclusion that the 12th grade NAEP mathematics assessment measures content that is also covered by another assessment that is designed and used to make judgments about academic preparedness for college, it is noted that the study was conducted by ACT, Inc., not an independent third party. Further, because a different methodology was used, the study results are not directly comparable to the results for the SAT and ACCUPLACER alignment studies in mathematics.

Mathematics: SAT. The Governing Board contracted with WestEd, an independent third party, to conduct the content alignment study comparing the NAEP 12th grade mathematics assessment and the SAT mathematics test. WestEd conducted the content alignment study using the design developed for the Governing Board

by Norman Webb. The full report of the content alignment study can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/SAT-NAEP_Math_Content_Comparison.pdf.

Overall, the study found similar content in the NAEP 12th grade mathematics assessment and the SAT mathematics test. The following highlighted content is an excerpt from the Executive Summary of the report (pp. iv-vi).

Alignment of 2009 NAEP Grade 12 Mathematics and SAT Mathematics, Executive Summary (Excerpt)

What is the correspondence between the mathematics content domain assessed by NAEP and that assessed by SAT?

At the standard level, the wording of the standards in the two frameworks is very similar. Both the NAEP and SAT frameworks include virtually the same five broad content categories, with SAT combining geometry and measurement into one standard. Each framework contains both general and specific objectives, although the SAT objectives, which are presented as content topics without indication of the cognitive level at which that content would be assessed, may be interpreted as more general than the NAEP objectives.

Although the structures of the two frameworks differ greatly beyond the standard level (including the NAEP framework having three levels while SAT has two), the mathematics areas typically expected of grade 12 students—number and operations, geometry and measurement, data analysis and probability, and algebra—are addressed in somewhat similar proportions.

To what extent is the emphasis of mathematics content on NAEP proportionally equal to that on SAT?

The greatest commonality between the two tests is their emphasis at the standard level. This is evident in the distribution of percentages of total hits from both assessments matched to each set of standards. Although there are some differences of emphasis, such as the full NAEP item pool's greater proportion of alignment to SAT "Data analysis, statistics, and probability," and the SAT short-version's greater proportion of alignment to SAT "Geometry and measurement," the proportions of alignments to "Algebra and functions" and "Number and operations" are comparable. There is also considerable overlap among some specific skills, with both assessments addressing many of the same NAEP "Number properties and operations" objectives and SAT objectives...

[continued]

Alignment of 2009 NAEP Grade 12 Mathematics and SAT Mathematics, Executive Summary (Excerpt) *continued*

Despite the difference in the degree of specificity of the two frameworks (most NAEP objectives are much more finely grained than the SAT objectives), it is clear that both assessments emphasize a number of the same or closely related skills. These include properties, equivalence, and operations on rational numbers (included in NAEP Goals 1.1 and 1.3 and included in SAT Objective N.2) and properties of two-dimensional shapes (included in NAEP Goals 3.1 and 3.3 and included in SAT Objective G.6).

Are there systematic differences in content and complexity between NAEP and SAT assessments in their alignment to the NAEP framework and between NAEP and SAT assessments in their alignment to the SAT framework? Are these differences such that entire mathematics subdomains are missing or not aligned?

While there is considerable overlap between the two assessments, primarily in the intersection of the NAEP “Algebra” and SAT “Algebra and functions” standards, there are notable differences as well. The SAT items had a somewhat limited range of coverage of the NAEP standards “Measurement,” “Geometry,” and “Data analysis, statistics, and probability,” with several goals receiving few item alignments. Even given the minimal coverage of some of the goals within each NAEP standard by SAT items, however, almost all NAEP items found a match in the SAT framework. The language of the objectives in the SAT framework is sufficiently broad to encompass the range of the NAEP items. For example, SAT Objective A.10, “Basic concepts of algebraic functions,” may accommodate most of the items aligning to the seven objectives within NAEP Goal 5.1, “Patterns, relations, and functions.” Finally, some NAEP items were found to be uncodable to the SAT objectives. These items assessed skills not present in the SAT framework.

The two tests are also similar in the average DOK [Depth of Knowledge] levels of items. However, while most items in both tests were found to be at DOK Level 2, NAEP items had a wider range of DOK than did SAT items, with more NAEP items coded to Levels 1 and 3. The Level 3 NAEP items often involved application of concepts through short or extended constructed-response items. Both tests also met depth-of-knowledge consistency overall (with each not meeting this criterion for only one standard as rated by one panel).

Overall, despite differences in alignment at the detailed specific objective level, differences in emphasis at the standard level, and a small difference in ranges of depth of knowledge, there is considerable overlap of content and complexity between [the NAEP 12th grade mathematics assessment and the SAT mathematics test].

Mathematics: ACCUPLACER. The Governing Board contracted with WestEd, an independent third party, to conduct the content alignment study comparing the NAEP 12th grade mathematics assessment and the ACCUPLACER mathematics test. The ACCUPLACER is used specifically to determine whether entry-level students have the mathematic knowledge and skills necessary for college-level work or require remedial mathematics courses.

WestEd conducted the content alignment study using the design developed for the Governing

Board by Norman Webb. The full report of the content alignment study can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACCUPLACER-NAEP_Math_Content_Comparison.pdf.

Overall, the study found similar content in the NAEP 12th grade mathematics assessment and the ACCUPLACER mathematics test, although the content of NAEP is much broader and complex. The following highlighted content is an excerpt from the Executive Summary of the report (pp. iv-vi).

Alignment of 2009 NAEP Grade 12 Mathematics and ACCUPLACER Mathematics Core Tests, Executive Summary (Excerpt)

What is the correspondence between the mathematics content domain assessed by NAEP and that assessed by ACCUPLACER?

The NAEP and ACCUPLACER assessments both cover certain content traditionally expected of grade 12 students, namely the two content subdomains of number or number operations and algebra (included in NAEP's "Number properties and operations" and "Algebra" standards and in ACCUPLACER's "Arithmetic," "Elementary algebra," and "College level math" standards), although their respective degrees of alignment and focus in these subdomains vary. Whereas the NAEP items focus primarily on number or number operations and algebra content at the grade 12 level, with an emphasis on problem solving and application of concepts at that grade level, the ACCUPLACER items span a wider developmental and grade-level range (from basic to more advanced).

This difference in focus is consistent with the purposes of the two assessments and their frameworks. The NAEP objectives are written to describe assessable content for grade 12 mathematics; thus, the 130 objectives tend to address the skills and concepts specific to that grade. The purpose of ACCUPLACER is to help determine appropriate placement for an individual student, and so the 87 ACCUPLACER objectives are spread more broadly across grade levels and are intended to be more general.

To what extent is the emphasis of mathematics content on NAEP proportionally equal to that on ACCUPLACER?

Regarding alignment to the NAEP framework, within the "Number properties and operations" and "Algebra" standards, NAEP items had broader overall coverage of the NAEP objectives than did ACCUPLACER. The 42

[continued]

Alignment of 2009 NAEP Grade 12 Mathematics and ACCUPLACER Mathematics Core Tests, Executive Summary (Excerpt) *continued*

NAEP items (the short version used for within-framework alignment) aligned to 72 NAEP objectives, whereas the 105 ACCUPLACER items (one complete form of each of the three ACCUPLACER Mathematics Core tests) aligned to only 56 NAEP objectives, with 44% of the ACCUPLACER item alignments aligning to only three NAEP objectives (all in “Number properties and operations” and “Algebra”). These differences in breadth and emphasis between the two assessments were evident across all NAEP standards. For example, in each assessment, items were aligned to four NAEP “Algebra” objectives for which the other assessment had no alignments, reflecting differences in emphasis within that standard.

Regarding alignment to the ACCUPLACER framework, ACCUPLACER items in the short version of 45 items covered all three standards—“Arithmetic,” “Elementary algebra,” and “College level math”—with a relatively even distribution, although “College level math” had the lowest percentage of item alignments. NAEP items in the full pool of 164 items also covered “Arithmetic,” “Elementary algebra,” and “College level math,” with a fairly even distribution of approximately one-third of NAEP codable items aligned to each standard, although “Elementary algebra” received somewhat fewer item alignments. Despite these differences in emphasis, however, considering only codable items, the percentages of alignments to each ACCUPLACER standard were relatively evenly distributed in both assessments and similar in distribution across assessments. At the objective level, the distribution of item alignments to objectives was relatively even on both tests, although each assessment was aligned to some objectives to which the other was not.

In summarizing cross-framework alignment, there was somewhat less even distribution of items than observed in within-framework alignment. The majority of items on each test were found to align to objectives on the other test. However, the 105 ACCUPLACER items aligned primarily (90%) to a total of seven out of 24 NAEP goals: three of the six goals from “Number properties and operations” in the NAEP framework, and four of the five goals in “Algebra.” Conversely, the NAEP items from the full pool of 164 items that aligned to the ACCUPLACER framework were distributed fairly evenly across the three ACCUPLACER standards and found to align to 75 ACCUPLACER objectives.

Are there systematic differences in content and complexity between NAEP and ACCUPLACER assessments in their alignment to the NAEP framework and between NAEP and ACCUPLACER assessments in their alignment to the ACCUPLACER framework? Are these differences such that entire mathematics subdomains are missing or not aligned?

Regarding differences in alignment of content, ACCUPLACER items had very limited coverage of measurement, geometry, and data analysis, content that is not included in the ACCUPLACER framework but that is included in the NAEP framework. Many NAEP items assessing these subdomains were found to be uncodable to the ACCUPLACER objectives (20 were rated uncodable by the majority of panelists in each panel). For other NAEP items that were aligned to an ACCUPLACER objective, there were often parts

[continued]

Alignment of 2009 NAEP Grade 12 Mathematics and ACCUPLACER Mathematics Core Tests, Executive Summary (Excerpt) *continued*

of those items not addressed by the objective. These items were coded as aligned, since they do assess an ACCUPLACER objective, but parts of the items also cover other skills not included in the ACCUPLACER framework.

Regarding differences in alignment of complexity, the items from both tests that aligned to the NAEP standards met the typical depth-of-knowledge (DOK) consistency threshold; that is, the items assessed the objectives at or above the DOK level of the objective. The items from both tests that aligned to the ACCUPLACER standards had somewhat different ranges of DOK. The ACCUPLACER short-version items were divided fairly evenly between Level 1 and Level 2. The NAEP items aligned to the ACCUPLACER framework had a wider range of DOK, with items at Level 1, 2, and 3, and a greater emphasis on Level 2 than was in the ACCUPLACER items.

The Content Alignment Studies: Summary Discussion for Mathematics

The NAEP 12th grade mathematics framework, test questions, and, for constructed response items, the score category rubrics, were compared with the analogous domain descriptions and test questions for the ACT, SAT, and ACCUPLACER mathematics tests. These three tests are used for college admissions and placement. They are well-established and have been used for these purposes for many years by professionals in postsecondary education. The test publishers regularly survey secondary and postsecondary educators about relevant content and have conducted research that supports the validity of the test content for the intended inferences and uses. The underlying assumption is that if the content of the 12th grade NAEP mathematics assessment is similar to the content of these mathematics tests, then the NAEP content is directly related to “academic preparedness for college.”

The ACT study found that “With regard to the ACT-to-NAEP comparison...nearly all of the ACT Mathematics domain and College Readiness Standards [are] reflected in the NAEP Mathematics domain, but...a number of the lower-level topics in the ACT Pre-Algebra subdomain were more consistent with Grade 8 NAEP topics.” In the NAEP-to-ACT comparison, there was uncertainty about “...whether a number of NAEP content topics—those pertaining to transformations, probability, statistics, and data analysis—are assessed by the ACT... and 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.”

The SAT study found similar content in the NAEP 12th grade mathematics assessment and the SAT mathematics test. “At the standard level, the wording of the standards in the two frameworks

is very similar. Both the NAEP and SAT frameworks include virtually the same five broad content categories, with SAT combining geometry and measurement into one standard...Although the structures of the two frameworks differ greatly beyond the standard level (including the NAEP framework having three levels while SAT has two), the mathematics areas typically expected of grade 12 students—number and operations, geometry and measurement, data analysis and probability, and algebra—are addressed in somewhat similar proportions...While there is considerable overlap between the two assessments, primarily in the intersection of the NAEP ‘Algebra’ and SAT ‘Algebra and functions’ standards, there are notable differences as well. The SAT items had a somewhat limited range of coverage of the NAEP standards ‘Measurement,’ ‘Geometry,’ and ‘Data analysis, statistics, and probability,’ with several goals receiving few item alignments. Even given the minimal coverage of some of the goals within each NAEP standard by SAT items, however, almost all NAEP items found a match in the SAT framework.”

The ACCUPLACER study found that “The NAEP and ACCUPLACER assessments both cover certain content traditionally expected of grade 12 students, namely the two content subdomains of number or number operations and algebra... although their respective degrees of alignment and focus in these subdomains vary... the 105 ACCUPLACER items aligned primarily (90%) to a total of seven out of 24 NAEP goals: three of the six goals from ‘Number properties and operations’ in the NAEP framework, and four of the five goals in ‘Algebra.’ Conversely, the NAEP items from the full pool of 164 items that aligned

to the ACCUPLACER framework were distributed fairly evenly across the three ACCUPLACER standards and found to align to 75 ACCUPLACER objectives...Regarding differences in alignment of content, ACCUPLACER items had very limited coverage of measurement, geometry, and data analysis, content that is not included in the ACCUPLACER framework but that is included in the NAEP framework. Many NAEP items assessing these subdomains were found to be uncodable to the ACCUPLACER objectives...”

The results across the three studies are consistent. In general, the content of the ACT, SAT, and ACCUPLACER mathematics tests are present in NAEP, but NAEP is generally broader. Alignment between NAEP and the other three respective assessments is substantial, but not perfect; perfect alignment is not expected.

These results support the conclusion that

- The NAEP 12th grade mathematics assessment measures academic knowledge and skills that are also covered by other assessments designed and used to make judgments about the academic preparedness of college freshmen for placement into entry-level, credit-bearing, non-remedial college courses that meet general education degree requirements for mathematics, and
- NAEP 12th grade mathematics test items and rubric scoring categories for items are appropriate for obtaining evidence of test takers’ possession of knowledge and skills needed for college freshmen to be placed into ECRNG college mathematics courses.

Discussion of Test Uses and Consequences in Relation to the Proposed Inferences

The National Assessment of Educational Progress is an independent monitor of student academic achievement in the United States. It reports on achievement at specific points in time and trends in achievement over time. NAEP reports to the public, national and state policymakers, and education leaders. It assesses student achievement at grades 4, 8, and 12 in important subjects. NAEP is used to compare performance across states and for 21 urban school districts. NAEP results are reported by gender, race/ethnicity, socioeconomic status, and for students with disabilities and students who are English language learners.

The audiences and the uses of NAEP are well established. They will not change as a result of the added meaning afforded by the inferences proposed in this validity argument. However, providing familiar external referents for performance on 12th grade NAEP will greatly enhance the understanding of NAEP results by its audiences.

Currently, there are either no or very low stakes consequences associated with the use of NAEP results. NAEP is not used as a basis for evaluating or diagnosing individual students, classroom or school performance, the effectiveness of individual teachers or administrators, or for any other accountability purpose. This will not change as a consequence of the inferences proposed in this validity argument.

Although the uses and consequences of NAEP will not change, employing the proposed inferences for NAEP reporting will bring a potential for misinterpretation. NAEP reports should include text explaining the limitations on interpretation and other caveats that were discussed in detail on pages 8-10.

Summary and Conclusion

The National Assessment Governing Board decided to determine the feasibility of transforming NAEP into a measure of academic preparedness for college. Consequently, the Governing Board made changes to the NAEP 12th grade reading and mathematics frameworks with the explicit purpose of measuring academic preparedness for college. The Governing Board conducted research that established a high degree of overlap between the content of the NAEP 12th grade reading and mathematics assessments and the content of widely used college admissions and placement tests.

Through a partnership with the College Board, performance on 12th grade NAEP was compared with performance on the SAT mathematics and critical reading assessments, with correlations of .91 and .74, respectively. Analyses of these data examined the average NAEP scores and inter-quartile ranges for students scoring “at” and “at or above” the College Board College Readiness Benchmarks for reading and mathematics. Similar analyses were conducted using data from the 2005

and 2009 NAEP High School Transcript Studies, using the college readiness benchmarks developed by ACT and by the College Board. A longitudinal study was conducted in partnership with the Florida Department of Education, following the 12th grade students in the 2009 state NAEP sample into Florida public postsecondary institutions, employing Florida's longitudinal database. The average NAEP scores and interquartile ranges were calculated for the Florida students in relation to the ACT or SAT college readiness benchmarks, whether they achieved a first-year GPA of B- or better, and whether they were placed into a remedial course in their first year of college.

The results of these analyses were consistent across studies and across years. They support the conclusions that students in the NAEP 12th grade distribution at or above the Proficient achievement level in reading and at or above 163 on the NAEP score scale for mathematics are

- likely to be academically prepared for entry-level, credit-bearing, non-remedial courses in broad access 4-year institutions and, for 2-year institutions, for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions, and
- not likely to need remedial/developmental courses in reading or mathematics in college.

That the NAEP sampling, scaling and statistical procedures yield accurate estimates of the percentage of students scoring at or above a

selected cut score (i.e., NAEP achievement level) is well established as a result of numerous validity studies and evaluations.

Thus, the NAEP 12th grade preparedness research results support the inferences that:

For reading:

Given the design, content, and characteristics of the NAEP 12th grade reading assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness:

the percentage of students scoring at or above Proficient on Grade 12 NAEP in reading is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in reading that would make them academically prepared for college.

For mathematics:

Given the design, content, and characteristics of the NAEP 12th grade mathematics assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness,

the percentage of students scoring at or above a score of 163 on the Grade 12 NAEP scale in mathematics is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics that would make them academically prepared for college.

Including these inferences in NAEP 12th grade reports will add meaning to the interpretation of the NAEP 12th grade results. In the past, NAEP has reported on what students know and can do in important subjects. Including these inferences will enable NAEP to serve as an indicator that conveys the percentage of the nation's 12th grade students that are academically prepared for entry-level, credit-bearing college courses without remediation.

However, steps must be taken to avoid potential misinterpretation. NAEP reports using these inferences must also include the limitations on interpretation and caveats described previously in this validity argument. In addition, the reports should explain the rationale for NAEP reporting on academic preparedness and describe appropriate and inappropriate uses of the results.

III. Epilogue and Concluding Considerations

Steps Taken Following Completion of the Validity Argument

The validity argument on the preceding pages is based on the model developed by Michael Kane. According to Kane, validity arguments “can be evaluated in terms of their clarity, coherence, and plausibility” (Kane, p. 14). Kane’s model does not prescribe a particular approach or explicit indicators for determining whether these criteria have been met. In the present case, a decision was made to subject the validity argument to external technical review and include the external review as a part of the validity argument document. Thus, after the validity argument was prepared, it was reviewed by two technical experts. The reviews conducted by the technical experts appear in Appendix A and Appendix B.

As a consequence of these technical reviews and associated guidance from the Governing Board’s Committee on Standards, Design and Methodology, the inferences presented in the validity argument were refined to make them more parallel in form. The refined inferences were approved by the Governing Board on August 3, 2013. The Governing Board Motion on Reporting 12th Grade Academic Preparedness for College that contains the approved inferences appears in Appendix C. To assist in the Board’s deliberations on the motion, a prototype chapter was prepared to illustrate what NAEP 12th grade reporting might include, which appears in Appendix D.

Considerations for the Future

The Governing Board began the preparedness initiative with a modest goal—to determine the feasibility of NAEP serving as an indicator of 12th grade academic preparedness for college and job training. The results of the first phase of research have confirmed that it is feasible for NAEP to serve as such an indicator with respect to academic preparedness for college. The Governing Board continues to study the feasibility of NAEP serving as an indicator of 12th grade academic preparedness for job training.

Likewise, the working definition of preparedness used in connection with the phase I research was modest—focusing on the academic knowledge and skills needed to qualify for placement into entry-level, credit-bearing courses without remediation, and specifically not on measures of postsecondary outcomes or success. This was because it was not known at the time whether longitudinal studies and studies linking performance on NAEP with performance on other relevant indicators would be technically possible to conduct. Even if the technical hurdles could be overcome, it was not certain whether the needed partnerships to carry out such studies would materialize. However, the technical hurdles were overcome, the partnerships did materialize, and the phase I research results demonstrated the technical feasibility of studying performance on NAEP in relation to a range of external indicators and postsecondary outcomes.

The internal logic and alignment among the definition of preparedness, the research projects, and the supported inferences is a validity consideration. Although the initial working definition for the preparedness research was appropriate to use as a starting point, it has been and continues to be an explicit principle that the working definition would change and evolve based on the research results. As described in the preceding validity argument, the NAEP-SAT linking study found substantial correlations between performance on NAEP and performance on the SAT. Through additional studies, average scores were estimated for NAEP 12th grade test-takers scoring at college readiness benchmarks set by the College Board and by ACT, Inc. on their respective college admissions tests. The Florida longitudinal study estimated average NAEP scores for students placed into entry-level, credit-bearing courses.

These indicators and outcomes are at a level that is somewhat beyond a “minimally acceptable” or “just qualified” level for placement into entry-level, credit-bearing courses. Thus, the average NAEP scores associated with these indicators and outcomes point more to solid academic achievement than to a “minimal” level of academic preparedness. This should be made clear in reporting NAEP 12th grade results in connection with the phase I research. Further, these results suggest that consideration be given to further refining the working definition of academic preparedness for college. As planning ensues for the phase II preparedness research, some attention should be given to updating the working definition in light of the phase I research results. For example, this could include addressing the level of performance and including postsecondary outcomes as elements of the working definition of 12th grade academic preparedness for college.

IV. References

- Achieve, Inc. (2005). *Recommendations to the National Assessment Governing Board on Aligning 12th Grade NAEP with College and Workplace Expectations: Reading*. Washington, DC.
- Achieve, Inc. (2006). *Recommendations to the National Assessment Governing Board on Aligning the 12th Grade NAEP with College, Workplace, and Military Expectations: Mathematics*. Washington, DC.
- ACT, Inc. (2007). *The ACT Technical Manual*. Iowa City, IA: Author.
- ACT, Inc. (2013). *College Readiness Standards: Reading*. Iowa City, IA: Author. Retrieved from <http://www.act.org/standard/planact/reading/index.html>.
- Allen, J., & Scoring, J. (2005). *Using ACT Assessment Scores to Set Benchmarks for College Readiness (ACT Research Series 2005-3)*. Iowa City, IA: ACT, Inc.
- American Diploma Project. (2004). *Ready or Not: Creating a High School Diploma that Counts*. Washington, DC: Achieve, Inc.
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Bozick, R., & Lauff, E. (2007). *Education Longitudinal Study of 2002 (ELS: 2002): A First Look at the Initial Postsecondary Experiences of the Sophomore Class of 2002*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- College Board. (2011). *The SAT[®] College and Career Readiness Benchmark User Guidelines*. Washington, DC: Author.
- College Board. (2012). *ACCUPLACER Program Manual*. Washington, DC: Author.
- Conley, D. T. (2007). *Toward a More Comprehensive Conception of College Readiness*. Eugene, OR: Educational Policy Improvement Center.
- Fields, R. & Parsad, B. (2012). *Tests and Cut Scores Used for Student Placement in Postsecondary Education: Fall 2011*. Washington, DC: National Assessment Governing Board.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50, 1-73.

- Kim, Y. K., Wiley, A., & Packman, S. (2013). *National Curriculum Survey on English and Mathematics, College Board Research Report 2011-13*. New York: The College Board.
- National Assessment Governing Board. (2008a). *Mathematics Framework for the 2009 National Assessment of Educational Progress*. Washington, DC: Author.
- National Assessment Governing Board. (2008b). *Reading Framework for the 2009 National Assessment of Educational Progress*. Washington, DC: Author.
- National Assessment Governing Board. (2009a). *Making New Links: 12th Grade and Beyond*. Washington, DC: Technical Panel on 12th Grade Preparedness Research.
- National Assessment Governing Board. (2009b). *NAEP in Relation to the ACT College Readiness Standards and Benchmarks in Reading and Math*. Washington, DC: ACT, Inc.
- National Assessment Governing Board. (2013). *Technical Report: NAEP 12th Grade Preparedness Research*. Washington, DC: Author. Retrieved from <http://www.nagb.org/what-we-do/preparedness-research.html>.
- National Center for Education Statistics. (2010). *The Nation's Report Card: Grade 12 Reading and Mathematics 2009 National and Pilot State Results (NCES 2011-455)*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.
- National Commission on NAEP 12th Grade Assessment and Reporting. (2004). *12th Grade Student Achievement in America: A New Vision for NAEP*. Washington, DC: Author.
- Ross, T., Kena, G., Rathbun, A., KewalRamani, A., Zhang, J., Kristapovich, P., & Manning, E. (2012). *Higher Education: Gaps in Access and Persistence Study (NCES 2012-046)*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: Government Printing Office.
- WestEd. (2010). *Comprehensive Report: Alignment of 2009 NAEP Grade 12 Reading and SAT Critical Reading*. San Francisco, CA: Author.
- WestEd. (2010). *Interim Report: Comparative Analysis of the Test Blueprints and Specifications for 2009 NAEP Grade 12 Mathematics and ACCUPLACER Mathematics*. San Francisco, CA: Author.
- WestEd. (2010). *Interim Report: Comparative Analysis of the Test Blueprints and Specifications for 2009 NAEP Grade 12 Mathematics and SAT Mathematics*. San Francisco, CA: Author.
- Wyatt, J., Kobrin, J., Wiley, A., Camara, W. J., & Proestler, N. (2011). *SAT Benchmarks: Development of a College Readiness Benchmark and its Relationship to Secondary and Postsecondary School Performance (Research Report 2011-5)*. New York, NY: College Board.

APPENDIX A

Review and Comment on

Validity Argument for NAEP Reporting on 12th Grade Academic Preparedness for College

Prepared for:

National Assessment Governing Board

Prepared by:

Gregory J. Cizek, PhD

Professor of Educational Measurement and Evaluation

University of North Carolina at Chapel Hill

cizek@unc.edu

June 30, 2013

Introduction

The National Assessment Governing Board (NAGB) sought input on the constellation of logical and empirical evidence it has amassed in support of certain claims centering on how scores on the 12th Grade National Assessment of Educational Progress (NAEP) might be interpreted with respect to college preparedness. The logic underlying those claims and the logical and empirical support for the claims can be referred to as the validity argument.

According to Kane (2013):

To validate an interpretation or use of test scores is to evaluate the plausibility of the claims based on the scores. An argument-based approach to validation suggests that the claims based on the test scores be outlined as an argument that specifies the inferences and supporting assumptions needed to get from test responses to score-based interpretations and uses. Validation then can be thought of as an evaluation of the coherence and completeness of this interpretation/use argument and of the plausibility of its inferences and assumptions. (p. 1)

The remainder of this paper presents the preparedness score interpretation claims proposed for the 12th grade NAEP scores and an overall evaluation of the plausibility of those claims.

To produce this evaluation, I relied primarily on two documents that presented the NAEP preparedness validity argument and evidence (Fields, 2013a, 2013b). A draft response to *Validity Argument for NAEP Reporting on 12th Grade Academic Preparedness for College* (Fields, 2013a) was submitted to the National Assessment Governing Board on May 29, 2013 (Cizek, 2013). This paper is a response to a revision of *Validity Argument for NAEP Reporting on 12th Grade Academic Preparedness for College* (Fields, 2013b).

The Proposed Interpretations and Claims

The proposed score interpretations related to college preparedness for NAEP Reading and Mathematics are the following:

READING – “The percentage of students scoring at or above Proficient on Grade 12 NAEP in reading is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in reading that would make them academically prepared for college.”

MATHEMATICS – “The percentage of students scoring at or above a score of 163 on the Grade 12 NAEP scale in mathematics is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics that would make them academically prepared for college.” (Fields, 2013b, p. 8)

The proposed interpretations are grounded in four claims (taken from Fields, 2013b):

- 1) The 12th grade NAEP results in reading and mathematics provide unbiased, accurate estimates of the percentages of students at or above specified score levels on the NAEP scales in reading and mathematics for 12th grade students in the United States.
- 2) Performance on 12th grade NAEP assessments in mathematics and reading is positively related to other measures associated with outcomes reflecting academic preparedness for college.
- 3) There is a point on the NAEP scale that corresponds to other measures, indicators, and outcomes associated with academic preparedness for college (i.e., possession of a specific level of academic proficiency, attainment of a first-year overall college GPA of B- or better, and placement into entry-level, credit-bearing, non-remedial college courses).
- 4) The positive relationship between NAEP and the other indicators and outcomes is meaningful in terms of academic preparedness for college, not merely a statistical artifact, because the 12th grade reading and mathematics domains measured by NAEP were specifically designed to measure academic preparedness for college.

Evaluation of Validity Evidence in Support of the Proposed Interpretations

Overall, my review and analysis leads me to conclude that the logical and empirical evidence amassed provides strong support for the proposed 12th Grade NAEP Reading and Mathematics score interpretations related to academic preparedness for college. The case for the validity of the interpretations is clear and coherent. The proposed interpretations are warranted in two ways: 1) by the accumulation of confirming evidence that is uniformly in the direction that would be hypothesized by the proposed interpretations; and 2) by the paucity of disconfirming evidence. On this point, it is noteworthy that the present validation effort appeared to be searching, objective, and contemplated the potential for disconfirming evidence.

It is my opinion, based on the evidence provided, that future NAEP reporting can provide reasonably confident and accurate indications of college preparedness in Reading and Mathematics.

It should be recognized, of course, that validation efforts typically should not be considered final or complete at any given juncture (see Cizek, 2012). Additional data can be gathered; additional experience with the test is gained; theory related to (in this case) college preparedness evolves; and new relationships

among variables can be explored. The following three recommendations suggest additional validation strategies or evidential sources that may have the potential to strengthen warrants for the intended preparedness score interpretations.

- 1) To enhance the clarity of the proposed interpretations, I offer the following recommendation: NAGB should consider making the score interpretations parallel by specifying the NAEP scale score associated with preparedness in Reading.

As currently worded, a defensible and specific scale score associated with preparedness is offered for NAEP Mathematics score interpretations; however, the interpretation for Reading is phrased as an achievement level: “The percentage of students in the 12th grade NAEP distribution at or above (Proficient for reading and a score of 163 for mathematics) is a plausible (or reasonable) estimate of the percentage of students who possess the knowledge, skills, and abilities in (reading or mathematics) that would make them academically prepared for college.”

The lack of parallelism in construction seems awkward, unnecessary, and potentially confusing to readers and users of this information. I recommend expressing both the Reading and Mathematics interpretations as NAEP scale scores, with elaboration as achievement levels if desired. An example of a slightly reworded interpretation along these lines would be:

“The percentage of students in the 12th grade NAEP distribution at or above a scaled score of XXX (Proficient) in Reading and a score of 163 in Mathematics is a plausible estimate of the percentage of students who possess the knowledge, skills, and abilities in those subjects that would make them academically prepared for college.”

- 2) To enhance the coherence of the proposed interpretations, I offer the following recommendation: NAGB should consider conducting additional research into the content coverage of the NAEP and the alignment of NAEP with traditional college admissions measures.

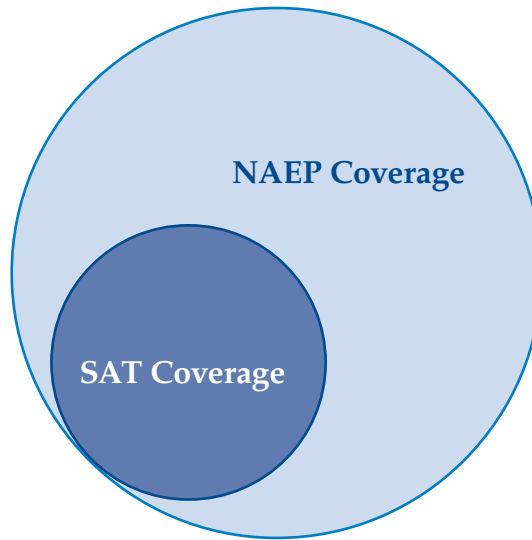
In its present form, it is argued that, in essence, the content of NAEP assessments in Reading and Mathematics covers everything that traditional college admissions measures (e.g., ACT, SAT, etc.) do, but also more. It is claimed that NAEP content coverage is “broader.” The Venn diagram below illustrates this claim:¹

¹ The Venn diagram and the reference to the SAT are presented only to illustrate relationships between content coverage on assessments. The diagram is not intended to represent the actual proportional content coverage between NAEP and college admissions assessments, nor that of the SAT in particular.

Figure 1 illustrates (ignoring the relative size of the circles) the claim that NAEP is somewhat of an umbrella assessment in terms of content coverage compared to the traditional college admissions measures on which alignment research has already been conducted. However, it is not clear that the fact that an umbrella relationship exists unequivocally supports the claim that NAEP assessments capture the same things about college preparedness as the college admissions tests or, importantly, that conclusions based on such alignment can unambiguously be made with respect to preparedness. For example, it would be theoretically possible for an examinee to score “Proficient” on NAEP Reading (and be deemed prepared for college) by getting very little of the “SAT- like” content correct on NAEP (that is, content deemed necessary for college success) and getting a lot of the “other” NAEP content correct (that is, the additional/broader content that may or may not necessarily be relevant to college preparedness).

Figure 1

Hypothetical content coverage between NAEP Assessment and College Admissions Assessment



- 3) To enhance the comprehensiveness of the proposed interpretations, I offer the following recommendation: NAGB should consider conducting additional research into the predictive validity of the NAEP with respect to college success.

Perhaps the most important variable assessed in the validation of traditional college admissions assessments is the ultimate criterion of college success—typically operationalized as first-year GPA, persistence, or some other variable. Although the validity evidence gathered so far links NAEP scores to scores on other measures that are, in turn, linked to college success, the present validity case for NAEP preparedness does not do so directly. For the future, independent evaluations of direct evidence regarding the extent to which NAEP preparedness scores are associated with college criterion outcomes would substantially bolster the evidence in support of the intended score interpretations.

Conclusion

The logical and empirical evidence gathered to date provides strong support for the proposed 12th Grade NAEP Reading and Mathematics score interpretations related to academic preparedness for college. The case for the validity of the interpretations is clear, coherent, and comprehensive. Recommendations were presented for future strategies to strengthen the validity case. Nonetheless, based on the empirical evidence and logical rationales to date, there appear to be strong warrants for the intended interpretations regarding NAEP reporting and indications of college preparedness in Reading and Mathematics.

References

- Cizek, G. J. (2013, May). Response to *Draft validity argument for NAEP reporting on 12th grade academic preparedness for college*. Report prepared for the National Assessment Governing Board, Washington, DC.
- Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods*, 17(1), 31-43.
- Fields, R. (2013, May 9). *Draft validity argument for NAEP reporting on 12th grade academic preparedness for college*. Washington, DC: National Assessment Governing Board.
- Fields, R. (2013, June 17). *Draft validity argument for NAEP reporting on 12th grade academic preparedness for college*. Washington, DC: National Assessment Governing Board.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1-73.

APPENDIX B

Comments on

Validity Argument for NAEP Reporting on 12th Grade Academic Preparedness for College

Prepared for:

National Assessment Governing Board

Prepared by:

Mark D. Reckase
Michigan State University

July 7, 2013

Beginning in March 2004, the National Assessment Governing Board (NAGB) began work to support the use of the 12th grade National Assessment of Educational Progress (NAEP) as a measure of “preparedness” of students for academic work at the college level. There are many challenges to this work, but one of the most important is to show that there is validity evidence to support the inference that students who are estimated to be above a specified level on the NAEP reporting score scale have the skills and knowledge to profit from credit-bearing, first-year, college-level coursework.

During the nine year period of this effort, the thinking about the way that validity evidence is collected and reported has had some significant changes. Particularly over the last few years, the work of Michael Kane (e.g., Kane, 2013) has provided guidance about how to present validity evidence for the interpretation of the results of an academic test in the form of what is now called a “validity argument.” The document that I reviewed was one of the first that I have seen that takes this approach to heart and makes a highly credible effort to apply this perspective on validation. In one sense, this is not surprising because work on NAEP has tended to be at the forefront of innovative psychometrics, be it on the use of item response theory procedures or standard setting. In another sense, it is surprising that NAGB has adopted this approach because there are few practical models for the creation of a validity argument. Even though there may have been some risk in being among the first to report support for an inference using the validity argument, this document is quite successful at providing a well supported validity argument. It gives other testing programs a very nice model for future reports on the validation of inferences from test scores.

My general view is that this document presents solid support for the inference that the proportion of the examinee population that is estimated to be above the specified cut score on the NAEP reporting score scale meets the definition of “preparedness for credit-bearing, first-year college coursework.” The evidence that was collected to support the inference is quite extensive and the connection of the evidence to the argument is logical and compelling. There are also appropriate cautions about over interpretation of results. It is very nice to see the areas of weakness in the supporting documents as well as the strengths. This adds credibility to the conclusions from the validity argument. This is not to say that the argument could not be tightened and elaborated, but this is an impressive example of a validity argument for a complex inference from a complex assessment.

A More Detailed Analysis

Although I have a very positive reaction to the report, it is important to probe the specifics of the argument and the claims being made. This may be interpreted as a desire for even more detail than is given in the report, but there is always a need for balance between detail and clear communication. The report is already long and detailed. I am reluctant to suggest adding more to it. But I do want to highlight some specific issues about some of the assumptions and claims in the argument.

The following statement is the basic inference that is the focus of the argument.

The percentage of students in the NAEP distribution at or above a particular score level in reading or mathematics on 12th grade NAEP is a plausible, or reasonable, estimate of the percentage of 12th grade students who are academically prepared for college. (p. 6)

This statement is very rich in meaning. To fully understand it, some background information is assumed to be known by the reader. Some of this background is listed here, but the list may not be comprehensive.

1. NAEP produces an accurate representation of the distribution of achievement of students in the areas of reading and mathematics.
2. The estimate of the proportion of students above a cut score on the NAEP reporting score scale is fairly accurate.
3. Students who are estimated to be above the specified cut score are likely to have high school grades and college admissions test scores that will make them eligible for admission to college.
4. Those students who are eligible for admission attend college and enroll in entry-level, credit-bearing courses.
5. The skills and knowledge in reading and mathematics are prerequisite to learning the content presented in the entry-level, credit-bearing courses.

The first two entries in the list are well supported by the technical documentation for NAEP. There are many years of research studies and analyses that show the technical quality of the assessment program. The last three of the entries in the list are more difficult to support because NAEP does not provide accurate student level scores and the individual students who participate are usually not identified so their academic history following the NAEP administration cannot be recorded. It is here that the special studies and data collections that have been done by NAGB are important to fill in links of the validity argument.

A Slight Variation on the Validity Argument

During the process of reviewing the report on the validity argument, I took notes on component parts of the argument. In some cases, the purpose of the notes was to highlight assumptions that were not explicitly stated. In other cases, the purpose was to elaborate on a step in the validity argument. A summary of these notes in the form of a slightly different validity argument than the one given in the report is given below. This is not meant to imply a problem with the validity argument in the NAGB report, but rather to add some commentary on that argument.

1. There is a body of knowledge and skills that is taught at the secondary school level that is prerequisite to gaining admission into entry-level, credit-bearing courses at colleges and universities.
 - a. There seems to be strong evidence for this from the America Diploma Project and the analysis of the admission and placement tests.
 - b. It might be helpful to think of this in terms of a Venn diagram that shows the intersection and union of the content descriptions from all of these different sources. The argument should be made that NAEP is based on a reasonable sampling of content from the intersection or the union.
2. College admissions test scores and high school transcripts provide information about the prerequisite knowledge and skills, and these are used to make decisions about admissions to the entry-level courses.
 - a. This is easy to document, but it is not explicitly stated in the argument. Of course, different institutions use the information in different ways.
3. The knowledge and skills reflected in college admissions tests and high school transcripts that are prerequisite to the entry-level college courses can be described in some detail to allow the design of a test to assess the knowledge and skills.
 - a. This is clearly supported by the information from the studies.
 - b. It would be useful to have a summary description of the common components from all of the parts.

4. NAEP assessments provide information about student acquisition of the knowledge and skills described above.
 - a. This is the main thrust of all of the content analysis.
 - b. The argument is compelling, but it would be helpful to have a general content description that is the result of all of the content analysis.
5. There is a threshold value for the knowledge and skills defined above. If students do not meet this threshold, they will not be ready to take the entry-level courses.
 - a. The comparative data make a good argument for the existence of the cut score.
6. A cut score on NAEP is consistent with the threshold.
 - a. There is a good process for identifying a reasonable cut score on NAEP to correspond to #5.
 - b. The combination of information from different tests results in strong support for parts of the argument.
7. The proportion of students estimated to be above the cut score on NAEP gives a good estimate of the proportion who exceed the threshold for admission into entry-level courses.
 - a. This is well supported by the statistical analysis procedures if the argument for an appropriate cut score is supported. In this case, there is reasonable support for the cut score from the connection to placement and admissions tests.

From this argument, I believe that the following inference from NAEP reported results is supported: The proportion of students estimated to be above the specified cut score on the NAEP reporting score scale is a reasonable estimate of the proportion of students who have the prerequisite knowledge and skills in mathematics and reading to profit from entry-level, credit-bearing college courses.

Reference

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1 – 73.

APPENDIX C

National Assessment Governing Board Motion on Reporting 12th Grade Academic Preparedness for College

Motion on Reporting 12th Grade Academic Preparedness for College

The Committee on Standards, Design and Methodology (COSDAM) recommended adoption of the motion provided below.

The National Assessment Governing Board approved the motion on August 3, 2013.

Background

- COSDAM has responsibility for overseeing the Governing Board’s program of research on 12th grade academic preparedness for college and job training.
- COSDAM has reviewed:
 - the research results completed during 2010-2012,
 - the staff-prepared validity argument developed in support of inferences proposed,
 - the independent technical reviews of the validity argument, and
 - a staff-prepared prototype intended to exemplify how the research on 12th grade academic preparedness for college would be reported.
- The inferences proposed in the motion below are intended as preliminary statements for reporting purposes and specifically not as performance standards for academic preparedness for college.
- Governing Board staff have worked and will continue to work collaboratively with staff of the National Center for Education Statistics to prepare the full explanatory text about the NAEP 12th grade academic preparedness initiative.
- The prototype document in the Board materials is a starting point for what may be said about the preparedness initiative in the report of the NAEP 12th grade reading and mathematics administered in 2013, with adjustments to the text and format to be made as needed.
- Further research to be conducted during 2013-2015 will be considered by COSDAM and the National Assessment Governing Board to make adjustments, as appropriate, to the statements in the motion below.

Motion

The National Assessment Governing Board approves the following statements for use in the reporting of the NAEP 12th Grade Reading and Mathematics assessments administered in 2013.

Reading: Given the design, content, and characteristics of the NAEP 12th grade reading assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness:

the percentage of students scoring at or above a score of 302 on Grade 12 NAEP in reading is a plausible estimate of the percentage of students who possess the knowledge, skills, and abilities in reading that would make them academically prepared for college.

A score of 302 corresponds to the cut score for the Proficient achievement level in 12th grade reading.

In 2013, XX% of 12th graders nationally scored at or above 302 in reading.

Mathematics: Given the design, content, and characteristics of the NAEP 12th grade mathematics assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness,

the percentage of students scoring at or above a score of 163 on the Grade 12 NAEP scale in mathematics is a plausible estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics that would make them academically prepared for college.

A score of 163 in mathematics is between the cut scores for the Basic and Proficient achievement levels in 12th grade mathematics.

In 2013, XX% of 12th graders nationally scored at or above 163 in mathematics.

APPENDIX D

Prototype Chapter for NAEP Reporting

Towards NAEP as an Indicator of Academic Preparedness for College and Job Training

Ray Fields July 7, 2013

For over a decade, the National Assessment Governing Board has been conducting research to enable 12th grade NAEP to serve as an indicator of academic preparedness for college and job training. This chapter provides the rationale for pursuing this goal; the research results from studies conducted in connection with the 2009 administration of 12th grade NAEP; and the implications for NAEP 12th grade reporting.

Introduction

Indicators of many kinds are used to monitor critical aspects of national life and inform public policy. These include economic indicators (e.g., gross domestic product), health indicators (e.g., cancer rates), and demographic indicators (e.g., population trends by race/ethnicity and gender).

NAEP serves the public as a national and state indicator of education achievement at the elementary and secondary levels. NAEP monitors student achievement at key points in the elementary/secondary progression: grades 4, 8, and 12.

According to the National Assessment Governing Board, the 4th grade is the point at which the foundations for further learning are expected to be in place (e.g., when “learning to read” becomes “reading to learn”).

The 8th grade is the typical transition point to high school.

The 12th grade is the end of the K-12 education experience, the transition point for most students to postsecondary education, training, the military, and the work force (Draft Policy Statement on NAEP).

NAEP is the only source of nationally representative 12th grade student achievement results. State tests of academic achievement are usually administered before 12th grade and are quite different across the country. Likewise, college admission tests like the ACT and SAT are generally taken before 12th grade by a self-selected sample and, therefore, are not representative of all 12th graders.

Consequently, NAEP is uniquely positioned to serve as an indicator of academic preparedness for college and job training at grade 12—the point that represents the end of mandatory schooling for most students and the start of postsecondary education and training for adult pursuits.

A wide array of state and national leaders has embraced the goal that 12th grade students graduate “college and career ready.” These include the leadership and members of the National Governors Association (NGA), the Council of Chief State School Officers (CCSSO), the Business Roundtable (BRT), the U.S. Chamber of Commerce (the Chamber), a task force on education reform of the Council on Foreign Relations, and state and national political leaders (Fields and Parsad).

NAEP and Academic Preparedness

The Governing Board believes that NAEP reporting on the academic preparedness of 12th grade students would afford an invaluable public service: providing an indicator of the human capital potential of today's and future generations of the nation's population.

The Board began this initiative in 2004, after receiving recommendations from a distinguished blue-ribbon panel that had examined whether NAEP should continue assessing at the 12th grade.

The panel stated that "America needs to know how well prepared its high school seniors are... [only NAEP] can provide this information...and it is necessary for our nation's well-being that it be provided." The panel recommended that NAEP continue to assess at grade 12 and that the 12th grade assessment be transformed to measure preparedness for college, job training, and the military (National Commission on NAEP 12th Grade Assessment and Reporting, p. 2.)

To transform 12th grade NAEP into an indicator of academic preparedness, the Governing Board took several significant steps.

1. The Board determined that measuring academic preparedness for college and job training should be an intended purpose of 12th grade NAEP.
2. The Board contracted with Achieve, Inc. in 2005 to review the NAEP 12th grade reading and mathematics assessment frameworks and identify where changes, if any, would be needed. Modest changes were recommended.
3. Accordingly, the Board made changes to the frameworks to be used for the administrations of the 12th grade assessments, scheduled for 2009 and 2013.
4. In 2006, the Governing Board assembled a team of noted psychometricians, industrial/organizational psychologists, and K-12 and postsecondary researchers to serve as a technical panel, advising on validity research to conduct.
5. In 2008, the technical panel recommended a comprehensive program of research. The validity of statements about academic preparedness in NAEP reports would be affected by the degree to which the results were mutually confirming.

Figure 1 presents a model of the research program, with five types of research displayed, the interrelationships that would be examined, and the potential meaning of the research results in terms of the NAEP score scale.

6. The Governing Board began contracting for the research studies in 2008, in connection with the 2009 administration of the 12th grade reading and mathematics assessments. More than 30 research studies were completed during the period 2009-2012.

The Research Findings

The research findings were consistent across studies and across years. For example, the content of the 12th grade NAEP reading and mathematics assessments was found to be similar to widely recognized tests used for college admission and placement.

Performance by the same students on NAEP and the SAT mathematics and reading tests was correlated at 0.91 and 0.74, respectively.

Statistical linking studies examining performance on NAEP and the college admission tests found that the college readiness benchmarks set for the ACT and SAT reading and mathematics were in a range around the Proficient achievement levels on the 12th grade NAEP reading and mathematics assessments (see Figures 2 and 3).

A longitudinal study followed a representative sample of Florida 12th grade NAEP test-takers into the state’s public colleges (see Figures 2 and 3). The longitudinal study permitted an analysis of performance on NAEP and actual student outcomes. In the first year of this study, an analysis

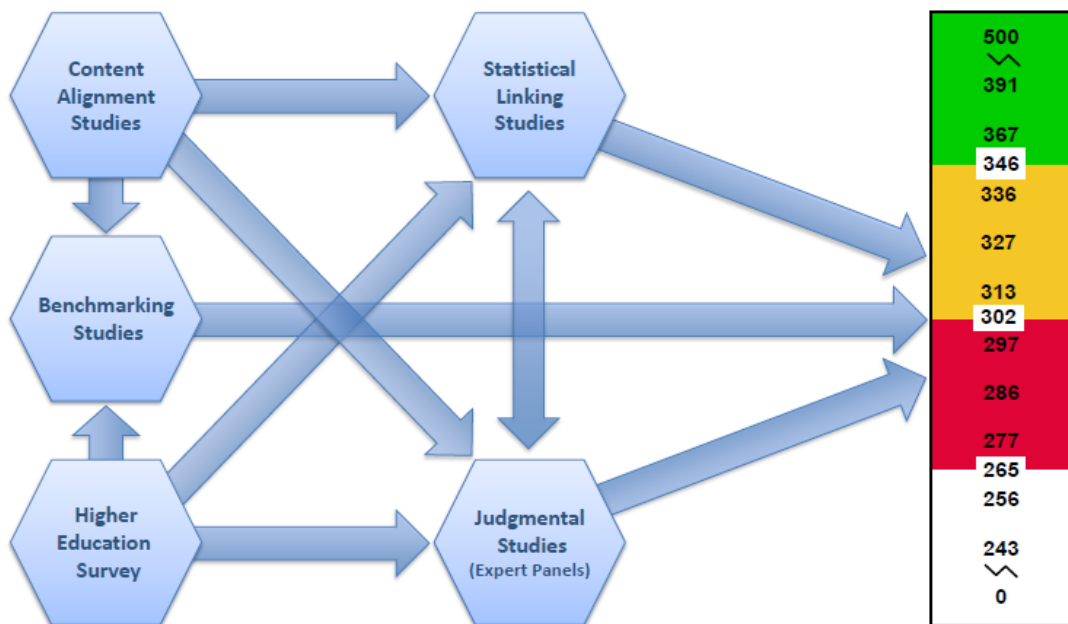
was conducted of performance on NAEP and (1) enrollment in regular versus remedial courses, and (2) first-year overall college grade point average (GPA). As with the other statistical studies, the average NAEP score of the students who were not placed into remedial courses or who had a first-year college GPA of B- or better was in a range around the 12th grade reading and mathematics Proficient achievement levels.

Results from the more than 30 studies were used to develop a validity argument to support proposed inferences (claims) about academic preparedness for college in relation to student performance on 12th grade NAEP. The validity argument was reviewed by two independent technical reviewers. The technical reviewers concluded that the validity argument supports the proposed inferences.

Figure 1

Model of the Preparedness Research Program

Study Interrelationships



The complete research reports and the validity argument, along with the two independent reviews, can be found at <http://www.nagb.org/what-we-do/preparedness-research.html>.

Although the research results support inferences about NAEP performance and academic preparedness for college, the research results do not support inferences about NAEP performance and academic preparedness for job training.

A second phase of NAEP preparedness research began in 2013 and is expected to be completed in time for reporting 12th grade results in 2015. The second phase of research results will be examined to determine the degree to which they confirm existing results.

A Transition to Reporting on Academic Preparedness

The reporting of the 12th grade results for 2013 represents a transition point for NAEP.

The interpretations of the 2013 NAEP 12th grade reading and mathematics results related to academic preparedness for college set forth in this report are considered foundational and subject to adjustment in the future.

These interpretations are included in this report because the technical reviewers found them to be technically defensible, but more importantly, to promote public discussion about their meaningfulness and utility.

The Context for Academic Preparedness for College

In the United States in 2013, there is no single, agreed upon definition of “academic preparedness for college” used by colleges for admission and placement. Postsecondary education in the U.S. is

a complex mix of institutions, public and private, that have different admission requirements and different procedures and criteria for placing individual students into education programs.

In this complex mix are 2-year institutions, 4-year public and private institutions with a wide range of selectivity, and proprietary schools. Institutions range from highly selective (i.e., with admission criteria including very high grade point averages, successful completion of rigorous high school coursework, and very high SAT and/or ACT scores) to open admission (i.e., all applicants are admitted).

Even within institutions, requirements may vary across majors or programs of study. For example, the mathematics and science high school coursework and academic achievement needed for acceptance into an engineering program in a postsecondary institution may be more rigorous than the general requirements for admission to the institution or for a degree in elementary education in that institution.

Defining Academic Preparedness for College

Given the diversity of postsecondary education institutions, it is essential to provide a reasonable definition of academic preparedness for NAEP reporting. The definition should be relevant to NAEP’s purpose of providing group estimates of achievement. (It is important to note that NAEP does not provide individual student results.) The definition should be meaningful to NAEP’s primary audiences: the general public and national and state policymakers.

The definition proposed in this report is intended to apply to the typical degree-seeking entry-level student at the typical college. For NAEP reporting, “academically prepared for college” refers to the reading and mathematics knowledge and skills

needed for placement into entry-level, credit-bearing, non-remedial courses in broad access 4-year institutions and, for 2-year institutions, the general policies for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

It is important to note the focus on “placement” rather than “admission.” This distinction is made because students who need remedial courses in reading, mathematics, or writing may be admitted to college, but not placed into regular, credit-bearing courses. The criterion of importance is qualifying for regular credit-bearing courses, not admission.

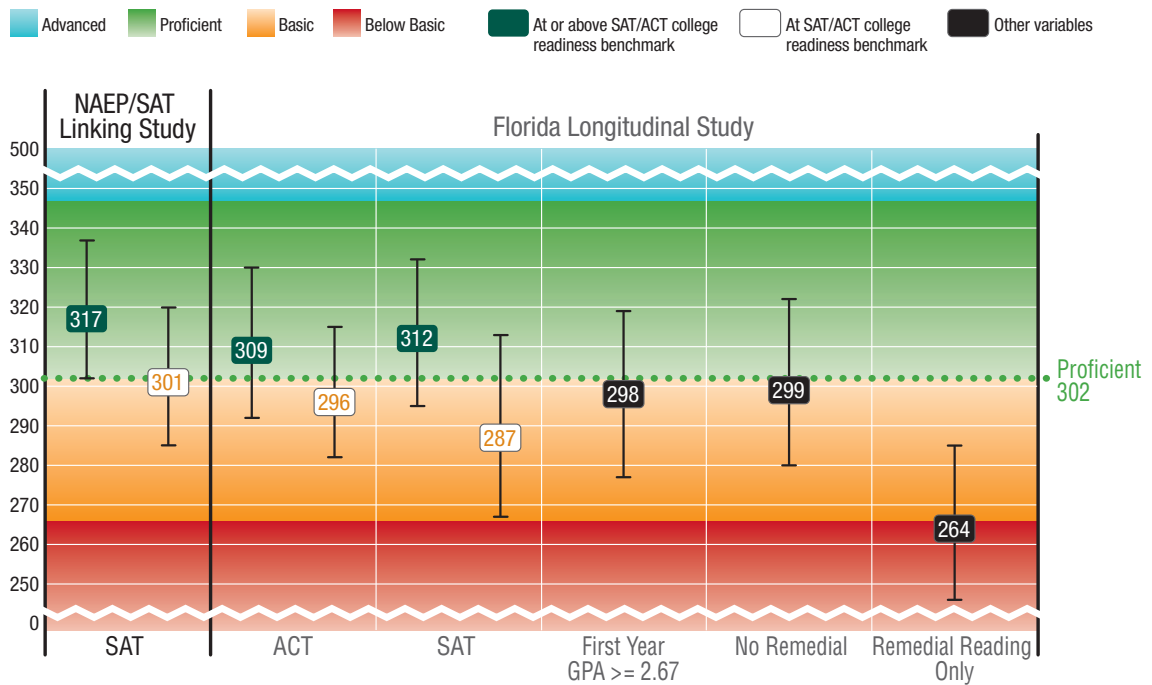
The definition is not intended to reflect

- academic requirements for highly selective postsecondary institutions;
- the additional academic requirements for specific majors or pre-professional programs, such as mathematics, engineering, or medicine; or
- academic requirements applicable to entry into certificate or diploma programs for job training or professional development in postsecondary institutions.

Figure 2

NAEP 12th Grade Preparedness Research: Reading

Average Scores and Interquartile Ranges on NAEP for Selected Variables and SAT/ACT College Readiness Benchmarks, from the 2009 NAEP SAT Linking Study and 2009 Florida Longitudinal Study



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, previously unpublished analyses, March 2013; National Assessment Governing Board, *Statistical Linking of National Results from NAEP and SAT (2009)*, previously unpublished analyses, March 2013; National Assessment Governing Board, *Longitudinal Statistical Relationships for Florida NAEP Examinees: First-Year College Performance Outcomes (2009–2010)*, previously unpublished analyses, March 2013.

The definition is focused on the first year of college; it does not address college persistence beyond the first year or completion of a degree. The definition will necessarily apply in general across a broad range of programs and majors, but should not be applied specifically to any particular program or major.

Proposed Inferences for NAEP Reporting

The NAEP preparedness research does not affect the NAEP results in any way. The distribution of student achievement is unchanged. That is, the average scores, the percentiles, and the

achievement level results are not impacted by the NAEP preparedness research.

The independent technical reviewers confirmed that the research findings support inferences about performance on NAEP 12th grade results in reading and mathematics in relation to academic preparedness for college.

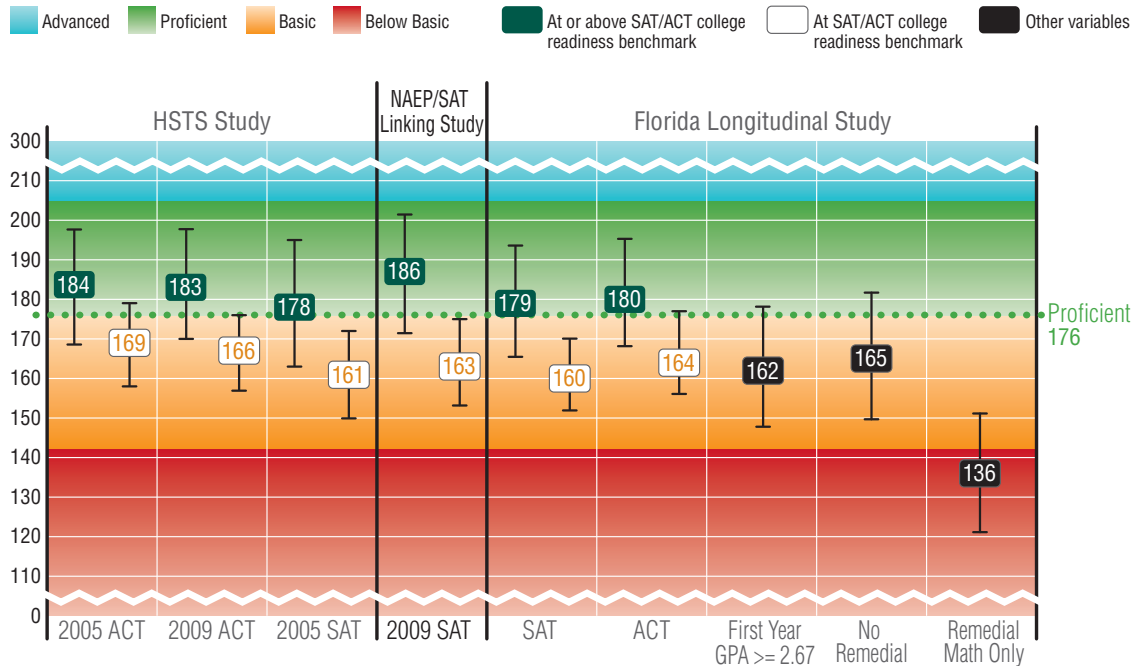
Proposed Inferences

In the NAEP/SAT linking study for reading (Figure 2), the average NAEP score for 12th grade students scoring at the SAT college readiness

Figure 3

NAEP 12th Grade Preparedness Research: Mathematics

Average Scores and Interquartile Ranges on NAEP for Selected Variables and SAT/ACT College Readiness Benchmarks, from the 2009 NAEP/SAT Linking Study, 2005 High School Transcript Study, 2009 High School Transcript Study, and 2009 Florida Longitudinal Study



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, *NAEP High School Transcript Study (2005, 2009)*, previously unpublished analyses, March 2013; National Assessment Governing Board, *Statistical Linking of National Results from NAEP and SAT (2009)*, previously unpublished analyses, March 2013; National Assessment Governing Board, *Longitudinal Statistical Relationships for Florida NAEP Examinees: First-Year College Performance Outcomes (2009–2010)*, previously unpublished analyses, March 2013.

benchmark for critical reading is 301, not significantly different from the Proficient cut score of 302. The results from the Florida longitudinal study are confirmatory.

These data, together with the content analyses that found NAEP reading content to be similar to college admission and placement tests, support the inference for reading that

Given the design, content, and characteristics of the NAEP 12th grade reading assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness:

the percentage of students scoring at or above Proficient on Grade 12 NAEP in reading is a plausible/reasonable estimate of the percentage of students who possess the knowledge, skills, and abilities in reading that would make them academically prepared for college.

In 2013, XX% of 12th graders (nationally) scored at or above Proficient in reading.

The study results support these inferences. However, there will be students scoring at or above Proficient who are not academically prepared and students scoring below Proficient who are academically prepared (i.e., there will be false positives and false negatives). This will be true for any assessment program that sets cut scores for a similar purpose.

In the NAEP/SAT linking study for mathematics (Figure 3), the average NAEP score for 12th grade students scoring at the SAT college readiness benchmark for mathematics is 163, lower than and significantly different from the Proficient

cut score of 176. The results from the High School Transcript Study and the Florida longitudinal study are confirmatory.

These data, together with the content analyses that found NAEP mathematics content to be similar to college admission and placement tests, support the inference for mathematics that

Given the design, content, and characteristics of the NAEP 12th grade mathematics assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness,

the percentage of students scoring at or above a score of 163 on the Grade 12 NAEP scale in mathematics is a plausible/reasonable estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics that would make them academically prepared for college.

In 2013, XX% of 12th graders (nationally) scored at or above 163 in mathematics.

To consider the plausibility of these estimates, comparisons can be made with the percentages of students who met the ACT or SAT college readiness benchmarks.

Information is available about students who were seniors in 2009 (ACT) and in 2010 (SAT). Thus, the ACT data are for the same student cohort as the NAEP data, but the SAT data are for a cohort that followed one year later.

It also must be noted that, unlike the NAEP results, neither the ACT nor the SAT results represent all 12th graders. Further, there is overlap among ACT

and SAT test-takers, with about 20% estimated to take both tests.

Assuming that a substantial portion of students who do not take either test are not academically prepared for college, it is not inconsistent that the NAEP percentages are lower than those for the respective college readiness benchmarks.

Percentages* Scoring at/above ACT and SAT College Readiness Benchmarks and at/above Proficient in Reading on NAEP and at/above 163 in Mathematics on NAEP		
	Reading	Mathematics
ACT (2009)	53	42
SAT (2010)	50	54
NAEP (2009)	38	40

* About 48% of 12th graders took the ACT or SAT. NAEP represents 100% of 12th graders.

Limitations on Interpretation and Other Caveats

False Negatives and False Positives

Some proportion of 12th grade students scoring below Proficient on the 12th grade NAEP Reading or below a score of 163 on the Mathematics Assessment are

- likely to be academically prepared for college
- not likely to need remedial/developmental courses in reading or mathematics in college,

but with a lower probability than those at or above Proficient in reading or 163 in mathematics.

In addition, some proportion of 12th grade students scoring at or above Proficient on the 12th grade NAEP Reading or 163 on the Mathematics Assessment may

- not be academically prepared for college
- need remedial/developmental courses in reading or mathematics in college.

Not a Preparedness Standard

The proposed inferences are not intended to represent or be used as standards for minimal academic preparedness for college. The proposed inferences are intended solely to add meaning to interpretations of the 12th grade NAEP reading and mathematics results in NAEP reports.

GPA of B- or Better

The variable “first-year GPA of B- or better” was selected because of its use as a research-based criterion in defining college readiness benchmarks developed by an acknowledged leader in college testing programs—the College Board. The College Board had agreed to partner with the Governing Board in a study linking performance on 12th grade NAEP with the SAT. Another leader in college testing programs, ACT, Inc., has developed similar benchmarks for its college admission assessments using a similar criterion and similar methodology. Because they are based on credible research related to college outcomes, and because performance on the respective tests could be linked to performance on NAEP, the college readiness benchmarks used by these testing programs were relevant, useful points of reference for the NAEP preparedness research.

The College Board has set a score of 500 on the SAT Mathematics and Critical Reading tests as its college readiness benchmarks in those areas.

Based on its research, the College Board has determined that the score of 500 predicts, with a probability of .65, attainment of a first-year overall GPA of B- or higher. Similarly, the ACT college readiness benchmarks are based on research indicating a .50 probability of attaining first-year grades in relevant courses (e.g., college algebra and courses requiring college-level reading) of B or better and .75 probability of C or better.

The proposed inferences are not intended to convey that a B- or any particular grade should be deemed a standard or goal for postsecondary student outcomes. This criterion was selected to foster comparability across the preparedness research studies, where applicable. However, it does seem self-evident that achieving a first-year GPA of B- or better, without enrollment in remedial/developmental courses, lends support to the likelihood of having possessed academic preparedness for first-year college courses upon entry to college.

Data Limitations

The NAEP preparedness research studies are comprehensive and the results consistent and mutually confirming. However, for reading, the statistical studies are limited to one year for data at the national level and to one state-based longitudinal study. For mathematics, there are two separate years of data at the national level and one state-based longitudinal study. Therefore, more evidence exists to support the plausibility of inferences related to mathematics than to reading.

Preparedness for Job Training

The completed research with respect to academic preparedness for job training does not support conclusions relative to the NAEP scale and will continue to be reviewed and discussed by the Governing Board.

Conclusion

The independent technical reviewers found the Governing Board's preparedness research to be methodical, rigorous, and comprehensive. They concluded that the research findings support the use of the proposed inferences in NAEP reports about 12th graders' academic preparedness for college.

The interpretations of NAEP results in relation to academic preparedness for college are being reported on a preliminary basis. They are provided to help foster public understanding and policy discussions about defining, measuring, validating and reporting on academic preparedness for college by NAEP and more broadly.

Including these inferences in NAEP 12th grade reports is intended to add meaning to the interpretation of the NAEP 12th grade results. However, the potential for misinterpretation exists. For these reasons, the section above on limitations on interpretation and other caveats is included in this chapter.

The Governing Board will monitor the use of these inferences as well as unintended consequences arising from their use as a part of the next phase of the preparedness research.

The next phase of the preparedness research is being conducted in connection with the NAEP reading and mathematics assessments administered in 2013. The research will be used to refine NAEP reporting on 12th grade academic preparedness.

[Note: References to be added.](#)

APPENDIX E

Background Information for Figure 1 and Figure 2

The tables on the following pages display the results of previously unpublished analyses. Data from these tables were used to prepare Figure 1 and Figure 2 that appear in the main body of the validity argument.

Table 1

**Average NAEP Mathematics Scores and Interquartile Ranges
In Relation to the SAT College Readiness Benchmark for Mathematics, 2009**

NAEP/SAT Analysis from Statistical Linking Study 2009	Average NAEP Mathematics Scores in Relation to SAT College Readiness Benchmark for Mathematics						
	N	Mean		25th		75th	
		Stat	SE	Stat	SE	Stat	SE
Average NAEP Scores of Students Scoring < SAT Benchmark Score of 500	6,470	144	0.5	130	0.8	160	0.7
Average NAEP Scores of Students Scoring = SAT Benchmark Score of 500	500	163	1.5	153	1.6	175	3.7
Average NAEP Scores of Students Scoring > SAT Benchmark Score of 500	8,290	187	1.1	173	0.9	203	1.5
Average NAEP Scores of Students Scoring > = SAT Benchmark Score of 500	8790	186	1.1	171	1.1	202	1.7
Total	15,260	168	1.2	147	1.1	190	1.7

SOURCE: National Assessment Governing Board, *Statistical Linking of National Results from NAEP and SAT (2009)*, previously unpublished analyses, March 2013.

Table 2

**Average NAEP Reading Scores and Interquartile Ranges
In Relation to the SAT College Readiness Benchmark for Critical Reading, 2009**

NAEP/SAT Analysis from Statistical Linking Study 2009	Average NAEP Reading Scores in Relation to SAT College Readiness Benchmark for Critical Reading						
	N	Mean		25th		75th	
		Stat	SE	Stat	SE	Stat	SE
Average NAEP Scores of Students Scoring < SAT Benchmark Score of 500	7,380	281	0.8	262	0.7	303	1.2
Average NAEP Scores of Students Scoring = SAT Benchmark Score of 500	620	301	1.9	285	2.0	320	3.6
Average NAEP Scores of Students Scoring > SAT Benchmark Score of 500	8,140	318	0.9	302	1.0	338	0.7
Average NAEP Scores of Students Scoring > = SAT Benchmark Score of 500	8,770	317	0.9	300	1.0	337	0.7
Total	16,150*	300	1.1	278	1.1	325	1.2

*Numbers may not add to totals due to rounding.

SOURCE: National Assessment Governing Board, *Statistical Linking of National Results from NAEP and SAT (2009)*, previously unpublished analyses, March 2013.

Table 3

**Average NAEP Mathematics Scores and Interquartile Ranges
In Relation to the SAT College Readiness Benchmark for Mathematics, 2005**

NAEP/SAT Analysis from High School Transcript Study Data, 2005	Average NAEP Mathematics Scores in Relation to SAT College Readiness Benchmark for Mathematics						
	N	Mean		25th		75th	
		Stat	SE	Stat	SE	Stat	SE
Average NAEP Scores of Students Scoring = SAT Benchmark Score of 500	70	160.6926	2.543	150.1689	2.0936	172.1845	2.3495
Average NAEP Scores of Students Scoring > = SAT Benchmark Score of 500	750	178.4017	1.2241	163.2184	1.1882	195.0469	2.0194

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, *NAEP High School Transcript Study* (2005), previously unpublished analyses, March 2013.

Table 4A

**Average NAEP Mathematics Scores and Interquartile Ranges
In Relation to the ACT College Readiness Benchmark for Mathematics, 2009**

NAEP/ACT Analysis from High School Transcript Study Data, 2009	Average NAEP Mathematics Scores in Relation to ACT College Readiness Benchmark for Mathematics						
	N	Mean		25th		75th	
		Stat	SE	Stat	SE	Stat	SE
Average NAEP Scores of Students Scoring = SAT Benchmark Score of 500	140	166.3915	1.2206	157.4843	2.2519	175.9394	1.39
Average NAEP Scores of Students Scoring > = SAT Benchmark Score of 500	1,200	183.0342	0.9603	169.2868	1.2031	197.0322	1.0457

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, *NAEP High School Transcript Study* (2009), previously unpublished analyses, March 2013.

Table 4B

**Average NAEP Mathematics Scores and Interquartile Ranges
In Relation to the ACT College Readiness Benchmark for Mathematics, 2005**

NAEP/ACT Analysis from High School Transcript Study Data, 2005	Average NAEP Mathematics Scores in Relation to ACT College Readiness Benchmark for Mathematics						
	N	Mean		25th		75th	
		Stat	SE	Stat	SE	Stat	SE
Average NAEP Scores of Students Scoring = ACT Benchmark Score of 22	50	169.3752	3.039	157.7917	4.1997	179.4143	3.232
Average NAEP Scores of Students Scoring > = ACT Benchmark Score of 22	450	183.8212	1.2741	169.4302	1.6421	198.2601	1.675

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, *NAEP High School Transcript Study* (2005), previously unpublished analyses, March 2013.

Table 5

Average NAEP Mathematics Scores and Percentiles In Relation to the SAT and ACT College Readiness Benchmarks for Mathematics for Students in the Florida 12th Grade State Sample, 2009

Variable	Average NAEP Mathematics Scores in Relation to SAT and ACT College Readiness Benchmarks for Mathematics												
	N	Mean		10th		25th		50th		75th		90th	
		Stat	SE	Stat	SE	Stat	SE	Stat	SE	Stat	SE	Stat	SE
SAT Test Taker	1,430	162	1.4	125	2.1	143	1.7	163	2.2	181	1.3	197	1.3
Without SAT Test Score	1,820	138	1.3	102	4.4	119	1.7	138	1.4	157	2.0	174	1.8
Average NAEP Scores of Students Scoring < SAT Benchmark Score of 500 for Mathematics	660	143	1.0	116	1.9	130	1.4	145	1.3	158	1.5	169	2.2
Average NAEP Scores of Students Scoring = SAT Benchmark Score of 500 for Mathematics	50	160	2.8	140	17.4	152	3.3	162	2.5	170	2.2	178	5.8
Average NAEP Scores of Students Scoring > SAT Benchmark Score of 500 for Mathematics	720	180	1.1	154	2.1	167	1.3	180	1.4	194	1.7	206	2.0
Average NAEP Scores of Students Scoring > = SAT Benchmark Score of 500 for Mathematics	770	179	1.1	152	2.9	166	1.7	179	2.0	193	1.1	206	4.2
Total	3,250	148	1.4	109	1.3	127	1.4	148	1.8	170	1.9	187	2.2
ACT Test Taker	1,530	152	1.9	113	1.3	130	2.2	152	3.3	174	2.9	192	2.1
Without ACT Test Score	1,720	144	1.2	106	1.9	124	2.7	145	1.3	165	1.3	183	2.5
Average NAEP Scores of Students Scoring < ACT Benchmark Score of 22 for Mathematics	1,000	138	1.2	107	2.4	122	1.2	139	1.5	156	1.4	169	1.9
Average NAEP Scores of Students Scoring = ACT Benchmark Score of 22 for Mathematics	80	164	2.1	140	12.4	156	4.6	168	2.1	177	6.5	183	2.2
Average NAEP Scores of Students Scoring > ACT Benchmark Score of 22 for Mathematics	450	183	1.3	158	5.5	173	0.8	185	1.9	197	2.1	208	1.3
Average NAEP Scores of Students Scoring > = ACT Benchmark Score of 22 for Mathematics	530	180	1.3	153	4.2	169	1.9	182	2.0	195	1.1	206	2.3
Total	3,250	148	1.4	109	1.3	127	1.4	148	1.8	170	1.9	187	2.2

*Numbers may not add to totals due to rounding.

SOURCE: National Assessment Governing Board, *Longitudinal Statistical Relationships for Florida NAEP Examinees: First-Year College Performance Outcomes (2009-2010)*, previously unpublished analyses, March 2013.

Table 6

Average NAEP Reading Scores and Percentiles In Relation to the SAT and ACT College Readiness Benchmarks for Reading for Students in the Florida 12th Grade State Sample, 2009

Variable	Average NAEP Reading Scores in Relation to SAT and ACT College Readiness Benchmarks for Reading												
	N	Mean		10th		25th		50th		75th		90th	
		Stat	SE	Stat	SE	Stat	SE	Stat	SE	Stat	SE	Stat	SE
SAT Test Taker	1,530	296	1.9	250	3.7	275	1.9	299	1.9	321	2.6	338	3.3
Without SAT Test Score	1,880	272	1.6	222	2.9	250	3.1	275	1.9	297	1.4	317	2.6
Average NAEP Scores of Students Scoring < SAT Benchmark Score of 500 for Reading	700	278	1.5	236	3.9	260	3.1	281	1.9	300	2.0	316	1.7
Average NAEP Scores of Students Scoring = SAT Benchmark Score of 500 for Reading	60	287	5.4	240	7.7	267	13.0	290	4.9	313	4.9	326	5.9
Average NAEP Scores of Students Scoring > SAT Benchmark Score of 500 for Reading	770	314	1.9	278	4.0	297	1.5	316	1.7	333	2.7	350	2.2
Average NAEP Scores of Students Scoring > = SAT Benchmark Score of 500 for Reading	830	312	1.9	274	4.1	295	2.0	315	1.9	332	2.4	349	3.0
Total	3,410	283	1.6	232	2.6	259	2.3	285	1.8	309	2.1	330	2.0
ACT Test Taker	1,590	288	2.2	237	3.5	264	2.5	290	3.1	314	2.8	334	2.1
Without ACT Test Score	1,820	278	1.7	228	2.8	256	1.8	281	2.0	305	2.1	325	2.8
Average NAEP Scores of Students Scoring < ACT Benchmark Score of 21 for Reading	860	270	2.0	225	2.8	250	2.4	273	1.9	292	3.0	310	2.8
Average NAEP Scores of Students Scoring = ACT Benchmark Score of 21 for Reading	100	296	4.5	259	11.7	282	8.2	300	7.3	315	5.8	330	2.1
Average NAEP Scores of Students Scoring > ACT Benchmark Score of 21 for Reading	630	311	2.2	272	4.2	294	2.5	314	1.3	332	2.5	349	2.9
Average NAEP Scores of Students Scoring > = ACT Benchmark Score of 21 for Reading	730	309	2.1	270	6.0	292	1.4	311	1.7	330	2.8	348	3.6
Total	3,410	283	1.6	232	2.6	259	2.3	285	1.8	309	2.1	330	2.0

*Numbers may not add to totals due to rounding.

SOURCE: National Assessment Governing Board, *Longitudinal Statistical Relationships for Florida NAEP Examinees: First-Year College Performance Outcomes (2009-2010)*, previously unpublished analyses, March 2013.

