MAKING NEW LINKS
12th Grade and Beyond

Technical Panel on 12th Grade Preparedness Research
Final Report

National Assessment Governing Board
U.S. Department of Education

WHAT IS NAEP?

The National Assessment of Educational Progress (NAEP) is a continuing and nationally representative measure of trends in academic achievement of U.S. elementary and secondary students in various subjects. For four decades, NAEP assessments have been conducted periodically in reading, mathematics, science, writing, U.S. history, civics, geography, and other subjects. By collecting and reporting information on student performance at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education.

National Assessment Governing Board

The National Assessment Governing Board was created by Congress to formulate policy for NAEP. Among the Governing Board's responsibilities are developing objectives and test specifications and designing the assessment methodology for NAEP.

Members

Darvin M. Winick, Chair
President
Winick & Associates
Austin, Texas

Amanda P. Avallone, Vice Chair
Assistant Principal & 8th Grade Teacher
Summit Middle School
Boulder, Colorado

David J. Alukonis
Chairman
Hudson School Board
Hudson, New Hampshire

Gregory Cizek
Professor of Educational Measurement
University of North Carolina
Chapel Hill, North Carolina

Carol D'Amico
President and CEO
Conexus Indiana
Indianapolis, Indiana

Honorable David P. Driscoll
Former Commissioner of Education
Commonwealth of Massachusetts
Malden, Massachusetts

Louis Fabrizio
Director
Division of Accountability Services
N.C. Department of Public Instruction
Raleigh, North Carolina

Anitere Flores
Legislator
Florida House of Representatives
Miami, Florida

Alan J. Friedman
Consultant
Museum Development and Science Communication
New York, New York

David W. Gordon
Sacramento County Superintendent of Schools
Sacramento County Office of Education
Sacramento, California

Robin C. Hall
Principal
Beecher Hills Elementary School
Atlanta, Georgia

Kathi M. King
12th Grade Teacher
Messalonskee High School
Oakland, Maine

Kim Kozbial-Hess
4th Grade Teacher
Hawkins Elementary School
Toledo, Ohio

Henry Kranendonk
Mathematics Curriculum Specialist
Milwaukee Public Schools
Milwaukee, Wisconsin

James S. Lanich
President
California Business for Educational Excellence
Sacramento, California

Honorable Cynthia Nava
Chair, Education Committee
New Mexico State Senate
Las Cruces, New Mexico

Honorable Steven L. Paine
State Superintendent of Schools
W.V. Department of Education
Charleston, West Virginia

Honorable Sonny Perdue
Governor
Atlanta, Georgia

Susan Pimentel
Educational Consultant
Hanover, New Hampshire

Andrew C. Porter
Dean, University of Pennsylvania
Graduate School of Education
Philadelphia, Pennsylvania

Warren T. Smith, Sr.
Vice President
Washington State Board of Education
Olympia, Washington

Mary Frances Taymans, SND
Executive Director
Secondary Schools Department
National Catholic Educational Association
Washington, DC

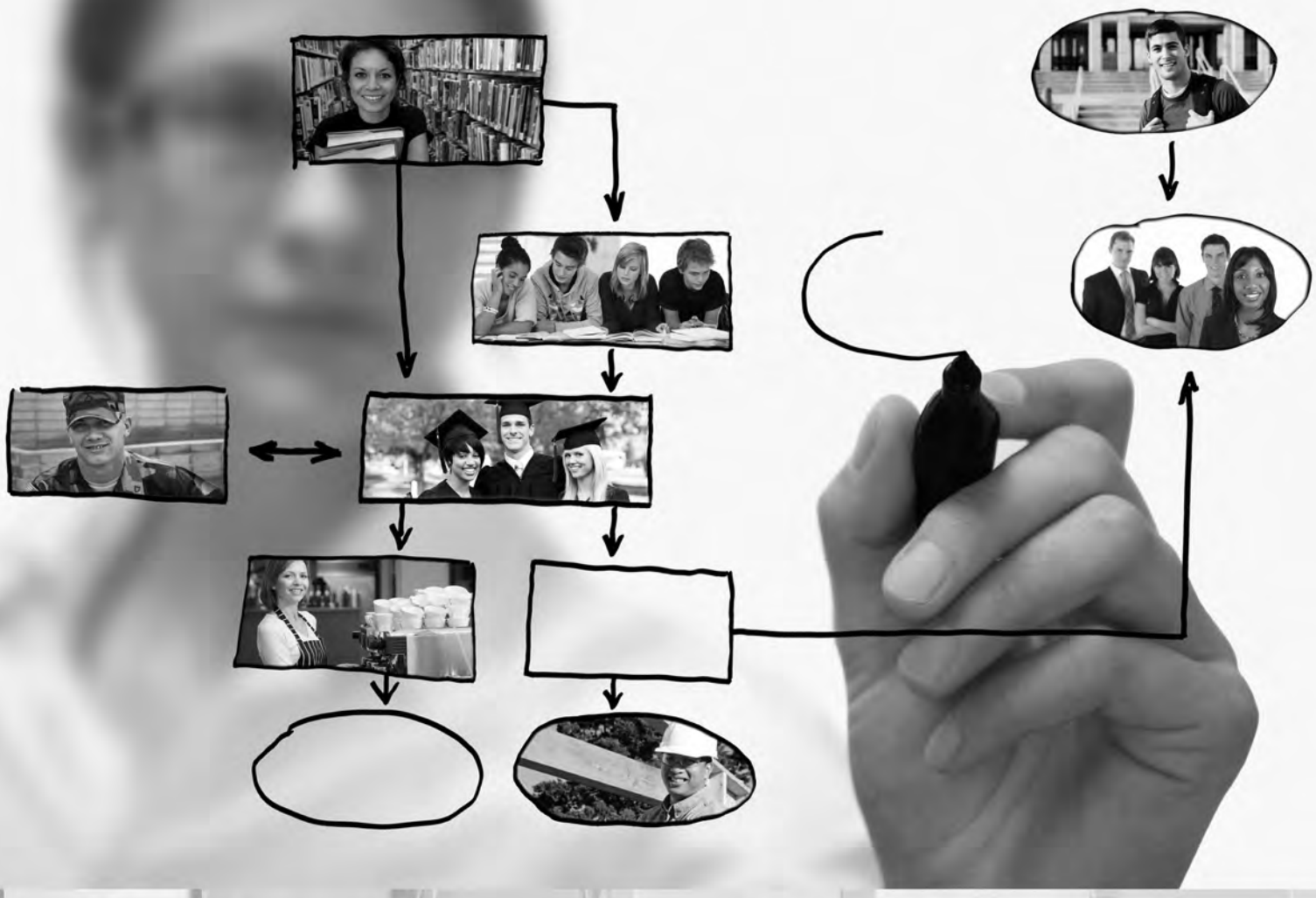
Oscar A. Troncoso
Principal
Anthony High School
Anthony Independent School District
Anthony, Texas

Eileen Weiser
Civic Leader
Ann Arbor, Michigan

Ex-officio Member

John Q. Easton
Director
Institute of Education Sciences
U.S. Department of Education
Washington, DC

MAKING NEW LINKS
12th Grade and Beyond



Technical Panel on 12th Grade Preparedness Research
Final Report

National Assessment Governing Board
U.S. Department of Education

Technical Panel on 12th Grade Preparedness Research

Michael Kirst (Chair)
Professor Emeritus of Education and Business Administration
Stanford University

John Campbell (Member)
Professor of Psychology
Professor of Human Resources and Industrial Relations
University of Minnesota

David T. Conley (Member)
Professor of Educational Policy and Leadership
Director of the Center for Educational Policy Research
University of Oregon

Michael Kane (Member)
Director of Research
National Conference of Bar Examiners

Mark David Milliron (Member)
President and CEO
Catalyze Learning International

Robert Mislevy (Adjunct Member)
Professor of Measurement, Statistics, and Evaluation
University of Maryland

George C. Thornton, III (Member)
Professor of Industrial Organizational Psychology
Colorado State University

Susan Loomis (Project Director)
Assistant Director for Psychometrics
National Assessment Governing Board

Acknowledgments

The Technical Panel on 12th Grade Preparedness Research is grateful for the participation and contributions of Governing Board Chair Darvin Winick and Board staff Interim Executive Director Mary Crovo, Assistant Directors Susan Loomis and Ray Fields, and Research Associate Michelle Blair, as well as former Executive Director Charles Smith. The Panel also expresses appreciation to Susan Loomis and Michelle Blair for their work in reviewing the Panel's deliberations and compiling its recommendations for development of the final report.

National Assessment Governing Board

Darvin M. Winick, *Chair*
Amanda P. Avallone, *Vice Chair*
Mary Crovo, *Interim Executive Director*
Susan Loomis, *Project Director and Assistant Director for Pyschometrics*
Michelle Blair, *Research Associate*

Suggested Citation: National Assessment Governing Board (2009).
Making New Links, 12th Grade and Beyond: Technical Panel on 12th Grade Preparedness Research Final Report.

For further information, contact:
National Assessment Governing Board, 800 North Capitol Street, N.W.,
Suite 825, Washington, DC 20002–4233
www.nagb.org

June 2009

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-0328

Table of Contents

Executive Summary . . . iii

Section 1 **TECHNICAL PANEL ON 12TH GRADE PREPAREDNESS RESEARCH** . . . 1

CHARGE TO THE PANEL . . . 1

Composition of the Technical Panel . . . 1

RESPONSE TO THE CHARGE . . . 2

Initial Steps of the Panel . . . 2

Meetings of the Technical Panel . . . 2

Defining Preparedness for NAEP Research . . . 2

Panel Perspectives on the Mechanics of Reporting NAEP Preparedness . . . 4

Section 2 **GOAL OF NAEP PREPAREDNESS STUDIES AND RESPONSIVE STUDY DESIGNS** . . . 5

STRATEGY: A MULTIMETHOD APPROACH . . . 5

OVERVIEW OF THE RECOMMENDED METHODOLOGIES . . . 5

Content Alignment . . . 6

Statistical Relationships . . . 7

Judgmental Standard Setting . . . 8

National Survey . . . 9

Model of the Interrelationships for the Group of Studies . . . 10

COMPILATION OF STUDY FINDINGS FOR NAEP REPORTING . . . 11

Maximum Information from All Studies . . . 11

The Array of Study Results: Potential Outcomes . . . 12

CONSIDERING THE CHARGE TO THE TECHNICAL PANEL . . . 14

Statistical Linking Studies . . . 14

Benchmarking Studies . . . 14

Studies to Yield Information About Examinees Who Score Below the Basic Achievement Level . . . 15

Section 3 **POSTSECONDARY ACTIVITIES OF INTEREST** . . . 17

COLLEGE PREPAREDNESS . . . 17

Issues . . . 17

Responsive Strategy: Use Preparedness Indicators to Locate Points on the NAEP Scale . . . 18

WORKPLACE PREPAREDNESS . . . 18

Issues . . . 18

Responsive Strategy: Use Exemplar Occupations . . . 19

SPECIAL CHALLENGES TO THE RECOMMENDED METHODOLOGIES . . . 22

Collaboration Opportunities for Accessing Data . . . 22

Technical Challenges in Study Design . . . 22

Section 4 **SUMMARY OF PROPOSED RESEARCH STUDIES AND PRIORITIZATION** . . . 25

OVERVIEW OF THE RECOMMENDED STUDIES . . . 25

Content Alignment . . . 26

Statistical Relationships . . . 27

Judgmental Standard Setting 28
National Survey 30
ADVISORY GROUPS FOR IMPLEMENTATION AND ANALYSIS PHASES OF STUDIES 31
CRITERIA USED TO ASSIGN PRIORITY FOR THE STUDIES 31
Section 5 **ADDITIONAL RECOMMENDATIONS FOR PREPAREDNESS RESEARCH AND REPORTING** 33
Encourage Parallel Studies by External Organizations 33
Be Mindful of the Evolving Context of Preparedness 33
Address Caveats to NAEP's Reporting on 12th Graders' Preparedness 34
Proceed with Reporting Workplace Training Preparedness on NAEP if Supported by the Research 35
Consider a Research Approach to Enable Affirmative Statements About Examinees Scoring Below the Basic Achievement Level 35
Evaluate the Feasibility of Using the Newly Identified NAEP Preparedness Reference Points to Report Preparedness 35
Section 6 **FUTURE STUDIES RECOMMENDED BY THE TECHNICAL PANEL** 37
Studies to Incorporate Additional Exemplar Occupations 37
Benchmarking Studies 37
Studies with Other State Longitudinal Databases 37
Reference Course Studies 37

ENDNOTES 39

REFERENCES 40

APPENDICES 41
Appendix A: List of Panel Meetings and Objectives 43
Appendix B: Guiding Principles for Preparedness 45
Appendix C: Considerations for Terms Related to Preparedness 47
Appendix D: Timelines For Completing the Studies 49
Appendix E: Recommended Studies for 2009 NAEP 12th Grade Preparedness Reporting 51
Appendix F: Summary Report on Recommendations of Expert Content Alignment Group 53

Executive Summary

The Technical Panel on 12th Grade Preparedness Research, convened by the Governing Board, consists of seven members with expertise in a variety of measurement and policy areas related to preparedness. The purpose of the Panel was to assist the National Assessment Governing Board in planning research and validity studies that will enable the National Assessment of Educational Progress (NAEP) to report on the preparedness of 12th graders for postsecondary education and job training after they graduate from high school. The first round of studies will be conducted before and during the 2009 NAEP 12th grade assessments of reading and mathematics, and the Board plans to begin this new type of reporting with these NAEP 2009 results, scheduled for release in 2010.

The Panel's deliberative process engaged each Panel member's expertise to refine ideas; gather supplementary materials; convene representatives of testing companies and partner organizations; and review the advantages and disadvantages of various sources of data. At each step in the process, the Technical Panel considered a range of alternatives and feasibility issues and then made choices to advance to the next point in deliberations.

KEY RECOMMENDATIONS

Use a variety of methodologies for NAEP preparedness studies in order to determine if mutually confirmatory evidence exists.
A multimethod approach is a sound and reasonable way to gain understanding of this complex set of issues and interrelationships. There is not a particular study that would comprehensively address the feasibility and validity issues for prospective NAEP preparedness reporting. The four recommended types of studies are:

- content alignment;
- statistical relationships with other assessments and postsecondary outcomes data;
- criterion-based judgmental standard setting; and
- national surveys.

Highlight the focus on reading and mathematics academic skills and avoid representing NAEP's preparedness reporting as the single, authoritative definition of preparedness.
Several national conversations include capabilities beyond academics when addressing preparedness and readiness. NAEP measures academic aspects of student achievement, and it is important to clearly communicate this focus of NAEP preparedness research to avoid misrepresentation and overstatement.

Maximize the information produced from all studies.
In comparing NAEP with other assessment instruments used as indicators of preparedness, Panel members have noted there may be overlap and there may be non-overlap. The Panel sees equal importance in describing the characteristics of overlap and the characteristics of non-overlap. These sources of information should be used to provide context and rigor for NAEP preparedness research and reporting.

Be mindful of the evolving context of preparedness.
There has been a substantial increase in the development of policies and standards to promote preparedness of students transitioning from high school to postsecondary endeavors. The Panel recommends careful positioning with respect to this dynamic

context. A contextual statement should be added to Report Cards to explain what NAEP can do and what NAEP cannot do in its reporting of 12th grade student preparedness. The statement should explain the definition that NAEP is using for preparedness. NAEP's capabilities and the NAEP definition of preparedness should be presented in the larger policy context.

Conduct preparedness validity research as an iterative process with additional studies for NAEP 2009 and beyond.
The Panel recommends that additional studies be conducted to enable continued preparedness reporting beyond NAEP 2009. To build on the foundation set by the NAEP preparedness studies for NAEP 2009 reading and mathematics and to address the evolving national context of preparedness, the Panel has proposed specific additional studies for NAEP preparedness research. These studies represent an incremental approach, including study designs such as benchmarking studies to administer NAEP to groups of interest, studies with additional state databases, studies to examine additional occupations, and studies to develop composite college or job training courses to represent the knowledge and skills needed to be prepared for entry.

RECOMMENDED STUDY DESIGNS

Content Alignment
The Technical Panel recommends that, as an essential step, content alignment studies be conducted for each assessment used as an indicator for reporting preparedness on the 12th grade NAEP scale. In order to use other assessments as indicators of preparedness and to capitalize on their preparedness research for interpretations of NAEP results related to preparedness, NAEP and the other assessments should measure similar content in a similar way. Content alignment studies will provide evidence of the extent to which the two assessments are aligned and provide a basis for interpreting the relationships of scores on the two assessments.

Statistical Relationships
The Panel recommends a series of studies aimed at statistically relating NAEP and performance on other assessments that serve as indicators of preparedness for postsecondary education and for job training programs in the civilian and military sectors. The Technical Panel recommends that the strongest feasible form of linking should be used to establish statistical relationships between NAEP and the other assessments. It is important to note that the strongest form of linking, known as equating, will not be possible because equating involves relating scores of two tests built to the same specifications—same content, same difficulty, same reliability—which means that results are interchangeable. Because NAEP is a unique assessment with a different function and purpose, equating it with the other assessments of interest is not an option. Instead, statistical relationships, such as concordance or the use of equipercentile methods to establish a working relationship, will be most likely between NAEP and other assessments.

Judgmental Standard Setting
The Technical Panel recommends studies involving judgments by subject matter experts (SMEs) for each type of postsecondary activity, relative to pre-existing sets of academic performance standards (or knowledge, skills, and abilities statements). SMEs include content experts as well as other stakeholders, such as job training staff members and college academic advisors. The judgments of SMEs will be used to set preparedness cut scores based on the performance standards. For some studies, the training performance standards would need to be developed if a particular occupation, for example, did not have an appropriate set available for use in setting preparedness cut scores.

National Survey
The Technical Panel recommends a survey to collect data from a nationally representative sample of two- and four-year postsecondary education institutions. The survey would collect information about (1) the assessments used for course placement and (2) the cut scores on widely used standardized tests for placement into college credit coursework and vocational training programs, placement into remedial programs in reading and mathematics, and exemption from placement tests. The survey results will yield descriptive information related to results from other studies and provide a context for reporting NAEP preparedness research.

RECOMMENDED STUDIES FOR NAEP 12TH GRADE PREPAREDNESS REPORTING

HIGH PRIORITY

Content Alignment Studies for NAEP and Assessments of Postsecondary Preparedness

- Comparison with college admissions and placement examinations (ACT, ACCUPLACER, ASSET, COMPASS, SAT)
- Comparison with workplace eligibility and placement examinations (WorkKeys and ASVAB)

Statistical Relationship Studies for NAEP and Assessments of Postsecondary Preparedness

- Linking national NAEP scores with preparedness indicator scores from other assessments
- Linking 12th grade NAEP performance with longitudinal databases (score data for college admission and course placement; transcript data; and workplace data)

Judgmental Studies to Set NAEP Cut Scores for Workplace Preparedness (Military and Civilian)

- Identification of five to seven target occupations across various sectors
- Identification and development of eligibility criteria for each target occupation's job training program
- Setting NAEP reading and mathematics job training program cut scores

National Survey of College Course Placement Assessments and Cut Scores

MEDIUM PRIORITY

Judgmental Studies to Set NAEP Cut Scores for College Preparedness

- Setting NAEP reading and mathematics college preparedness cut scores using:
  - ACT College Readiness Standards
  - College Board Standards for College Success
  - Standards developed by subject matter experts specializing in college course placement ■

Section 1

Technical Panel on 12th Grade Preparedness Research

CHARGE TO THE PANEL

The purpose of the Technical Panel on 12th Grade Preparedness Research was to assist the National Assessment Governing Board in planning research and validity studies that will enable the National Assessment of Educational Progress (NAEP) to report on the preparedness of 12th graders for postsecondary education and job training after they graduate from high school.

At the first meeting of the Panel, Governing Board Chair Darvin Winick presented the Charge from the Board. In summary, the Technical Panel's Charge is to propose studies that will yield validity evidence for using grade 12 NAEP as a measure of preparedness or as a tool by which preparedness can be reported. As Mr. Winick has also remarked, one of the goals of this entire endeavor of developing NAEP preparedness research studies is to anchor the NAEP scale in more empirical data, and although the future success of this endeavor is unknown, the time to try is now.

Having established it as an independent Panel, Board members did not participate in Panel meetings. Reports and updates were provided to the Board at each quarterly meeting by Board staff members. Panel Chair Michael Kirst attended three Board meetings (November 2007, August 2008, and November 2008) to give updates to and to gather feedback from the full Board. Detailed technical updates were also shared with the Board's Committee on Standards, Design and Methodology. Further, the Panel accepted an invitation to meet with former U.S. Secretary of Education Margaret Spellings on September 17, 2008, and brief her on the work of the Technical Panel.

The studies recommended by the Panel are to be initiated in conjunction with the 2009 NAEP reading and mathematics assessments for 12th grade. The Board's goal is to begin this new type of reporting with preparedness indicators for these 2009 NAEP results, scheduled for release in 2010. The Panel, therefore, was charged to consider the desirability and feasibility of various research studies and to recommend research designs and priorities for implementation in 2008, 2009, and thereafter. The Panel was also asked to consider ways to provide information about the preparedness of 12th grade students who score below the Basic achievement level.

To support the Panel's focus on the technical aspects of preparedness research, the Panel was asked to consider the technical quality and feasibility of studies they recommended and the priority they assigned to each study. The Board will address issues regarding funding for recommended studies that are to be implemented.

Composition of the Technical Panel

The seven members of the Technical Panel represent breadth and depth of expertise in critical areas related to the transition and assessment of high school students as they enter the postsecondary worlds of college and work. Chaired by Michael Kirst, the group includes members with expertise in the civilian workplace, military training, two-year and four-year college requirements, and the overall transition to college, as well as NAEP and other assessments, validation methodology, and psychometrics. Some Panel members are NAEP experts, some are familiar with NAEP, and others are new to NAEP. The Panel has viewed this diversity as an asset to the deliberative process of developing an approach to measuring

and reporting preparedness for NAEP and reviewing study designs for achieving this goal.

RESPONSE TO THE CHARGE

Throughout its work, the Panel has recognized the dynamic and rapid expansion of initiatives, policies, and research related to readiness and preparedness. The Panel members have completed their work, and they have built a solid foundation for addressing vital issues that will continue to evolve for years to come.

Initial Steps of the Panel

The first meeting of the Technical Panel was designed to provide a general orientation to the Panel members regarding their Charge, the rationale for the goal of reporting preparedness for grade 12 NAEP, actions already undertaken by the Governing Board to pursue this goal, and characteristics of the NAEP program. Staff members of both the Governing Board and the National Center for Education Statistics presented information for discussion by the Panel. The NAEP briefings and related discussions enabled a common understanding of how the Panel's work would relate to the core goals of NAEP and the Governing Board.

As Panel members discussed their own relevant areas of expertise, the beginnings of a general approach started to emerge. They agreed that the general approach should focus on what was practical and what could be done, rather than on designing an ideal study. But they also agreed that looking at the attributes of an ideal study would determine what was possible and feasible for NAEP. The Panel set the broad strategy for approaching their Charge in practical terms: Add to the richness of NAEP by increasing understanding of how the performance of 12th graders on NAEP relates to preparedness for college and workplace training.

They resolved to initially focus their attention on delineating the feasibility of different approaches and study designs. Over the next few meetings, their deliberative process engaged each Panel member's expertise to refine ideas; gather supplementary materials; convene representatives of testing companies and partner organizations; and review the advantages and disadvantages of various sources of data. At each step in the process, the Technical Panel considered a range of alternatives and feasibility issues and then made choices to advance to the next point in deliberations.

Meetings of the Technical Panel

The Technical Panel convened for six in-person meetings and three teleconferences:

Meeting 1: June 5-6, 2007, Washington, DC
Meeting 2: August 13-14, 2007, Denver, CO
Teleconference 1: September 6, 2007
Meeting 3: October 2-3, 2007, Washington, DC
Teleconference 2: December 3, 2007
Meeting 4: February 6-7, 2008, Washington, DC
Meeting 5: April 16-17, 2008, San Francisco, CA
Teleconference 3: September 15, 2008
Meeting 6: September 17-18, 2008, Washington, DC

A brief summary of each meeting's objectives can be found in *Appendix A: List of Panel Meetings and Objectives*.

Defining Preparedness for NAEP Research

National Assessment Governing Board

Through the work of the National Commission on NAEP 12th Grade Assessment and Reporting in 2004, and the work in 2006 of the Board's Ad Hoc Committee on Planning for NAEP 12th Grade Assessments in 2009, the Governing Board had considered several ways to define preparedness. The Technical Panel used these earlier discussions to guide its work. The Board's deliberations have led to a focus on preparedness as a subset of readiness. Specifically, preparedness focuses on academic qualifications, which are measured by NAEP. Readiness includes behavioral aspects of student performance—time management, persistence, and interpersonal skills, for example—which are not measured by NAEP.

The recommendations to the Governing Board by the National Commission and the Ad Hoc Committee led to the following key principles. Please see a full listing of these principles in *Appendix B: Guiding Principles for Preparedness*. It is important to note that NAEP assesses reading and mathematics separately; therefore, preparedness in reading and mathematics for the same group of students cannot be examined simultaneously.

1. Preparedness for college refers to the reading and mathematics knowledge and skills necessary to qualify for placement into entry level college credit coursework without the need for remedial coursework in those subjects.
2. Preparedness for workplace training refers to the reading and mathematics knowledge and skills needed to qualify for job training; it does not mean that the student is ready to be hired for a job.
3. Preparedness does not mean success in postsecondary education and training. NAEP can provide valuable information by reporting eligibility to enter postsecondary activities. Measuring success directly requires individual student scores on NAEP and longitudinal studies with these data. These research design elements are not compatible with the NAEP design or the NAEP legislation.
4. Preparedness in the NAEP context must be limited to academic qualifications for postsecondary education and workplace training.
5. Preparedness for job training is based on the assumption that similar jobs in both the military and civilian sectors require approximately equal reading and mathematics knowledge and skills to qualify for entry.

Technical Panel's Deliberations

As noted earlier, the Panel viewed NAEP preparedness research as an effort to determine the feasibility of reporting statements about 12th graders' preparedness using NAEP results. The Panel has concluded that there is sufficient reason to proceed with these studies. In order to proceed, an initial operational definition of preparedness is needed. The Panel believes that the information about and understanding of preparedness in relation to NAEP will evolve and deepen as a result of preparedness research studies to be conducted in the short term and the long term. Concomitantly, the operational definition(s) of preparedness for NAEP will be developing and evolving. Results of the studies may indicate whether future changes are needed in the reading and mathematics frameworks and assessments to better measure preparedness. Based on the Governing Board's previous discussions of preparedness definitions and on the Panel's deliberations, the Technical Panel recommends that this research initiative go forward as an effort to develop NAEP as an indicator of academic preparedness for college and workplace training.

Defining Academic Preparedness for NAEP – Preparedness represents the academic knowledge and skill levels in reading and mathematics necessary to be qualified for placement into a job training program (for the workplace context) or 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 (for the college context). Academic preparedness is separate and different from college readiness because, in addition to academic skills, readiness encompasses behavioral aspects of individual performance related to success, and these additional attributes are not measured by NAEP. Examples of readiness characteristics include persistence, time management, interpersonal skills, and knowledge of the context of college. A comprehensive measure of all factors necessary to enter postsecondary education with a maximum likelihood of succeeding would address practices and programs geared toward both preparedness and readiness. However, the research initiative described in this report focuses on preparedness only—in part because this is what grade 12 NAEP is best equipped to measure, but also because academic skills in reading and mathematics constitute an important and foundational dimension of readiness.

Defining College Preparedness for NAEP – For the NAEP context, preparedness for college means a student has at minimum the reading and mathematics knowledge and skills to qualify for entry into a credit-bearing course en route to a four-year undergraduate degree. This includes many courses offered at two-year institutions, partly because two-year transfer degrees are often the full equivalent of a four-year institution's general education program. In addition, for the purposes of NAEP preparedness research, credit-bearing courses refer to the reading and mathematics knowledge and skills required in general education courses, which are typically "introductory" courses in core subject areas. In choosing this definition, the Panel recognizes there are ranges of institutions that will have different definitions of preparedness across majors and even across colleges within the same institution. Given the diversity of American institutions of postsecondary education, the Panel has identified key terms that may be used in a statement of 12th graders'

preparedness. Please see *Appendix C: Considerations for Terms Related to Preparedness.*

The Technical Panel has concluded that, with the appropriate validity studies, reporting on preparedness using NAEP seems feasible.

Defining Workplace Training Preparedness for NAEP – For the NAEP context, preparedness for workplace training requires that a student has the reading and mathematics knowledge and skills sufficient to qualify for placement into a job training program. Job training programs constitute a variety of pathways, including apprenticeship programs, community college technical certificates and job training programs, on-the-job training programs, and vocational institute or certification programs. While target occupations that require a bachelor's degree would be excluded from this definition of workplace preparedness, the Panel has noted that the academic requirements of the certification programs for some target occupations may exceed those of a two-year degree.

Panel Perspectives on the Mechanics of Reporting NAEP Preparedness

Connections Between Preparedness Research and Reporting

The Technical Panel has concluded that, with the appropriate validity studies, reporting on preparedness using NAEP seems feasible. The Panel identified early that there were two ways for NAEP to report on preparedness. NAEP could be established as a measure of preparedness. In this process, research would be conducted to identify whether the NAEP assessment has the properties needed to report on pre-specified statements of preparedness. A different strategy would involve identifying external indicators of preparedness and relating those indicators to the NAEP scale.

It is important to note that the Technical Panel has pursued the latter approach. With the findings of this research, the Governing Board will be equipped to determine valid statements and interpretations of NAEP performance in terms of preparedness. The broadness of these statements will be determined by the research findings. If a similar pattern of findings emerges from the research related to college and work, then it may be possible to have one general statement of preparedness that applies to both college and the workplace. The research, however, may support separate statements about what preparedness means for college versus what it means for the workplace. Further, research findings may only enable preparedness statements for particular contexts within postsecondary education or within the workplace. In all of these possible scenarios, the Panel has stressed that the Board will need to be especially careful in reporting "preparedness" to avoid misrepresentation and overstatement. The operational definition of preparedness is limited to academic skills in reading and mathematics as assessed by NAEP. It will be important to make this focus as explicit as possible in grade 12 NAEP preparedness reports.

In a related recommendation, the Technical Panel believes it is important to caution against interpreting NAEP's preparedness as the single authoritative definition or conception of preparedness. This understanding must be made clear to the public and the media.

Reporting Preparedness and Reporting Achievement

The Panel has also noted that achievement and preparedness are conceptually different. For example, the cut score indicating performance of a proficient learner may be higher than the cut score indicating performance that enables placement into a non-remedial credit-bearing general education course. The Panel has generally conceptualized NAEP reporting for grade 12 reading and mathematics to include achievement levels as well as preparedness reference points on the NAEP scale. Reference points are scores on the NAEP scale that represent indicators of preparedness for specific postsecondary contexts. These reference points will be identified through preparedness research. ■

Section 2

Goal of NAEP Preparedness Studies and Responsive Study Designs

The goal of NAEP preparedness studies is to establish potentially feasible and valid ways to report preparedness on the grade 12 NAEP score scale. Together, the recommended studies are meant to address this goal strategically.

STRATEGY: A MULTIMETHOD APPROACH

The Panel sees the need for multiple types of studies, which, taken together, would provide evidence to develop or support statements about preparedness based on NAEP performance. The Panel believes that such a multimethod approach is a sound and reasonable way to gain understanding of this complex set of issues and interrelationships. The Technical Panel supports using a variety of methodologies for NAEP preparedness studies in order to determine if mutually confirmatory evidence exists. No single study would comprehensively address the full range of feasibility and validity issues necessary to resolve all the technical and policy issues associated with NAEP preparedness reporting. The Panel has deliberated extensively to ensure an appropriate balance is achieved between qualitative and quantitative studies, given the potential impact of this work on public policy discussions and because many questions and concerns will be raised if preparedness is reported on the NAEP score scale.

The proposed studies and methodologies have been selected to support one another. The findings of the studies should provide information that will enable the Governing Board to make inferences and eventually reach conclusions about how best to report preparedness. Studies will help the Board determine the reference points or ranges on the NAEP scale associated with preparedness in different settings or contexts. The group of studies will be designed to provide information that allows each potential reference point or range to be evaluated relative to other identified points or ranges derived from other studies. Together, these reference points may provide enough information to report 12th grade NAEP reading and mathematics results in terms of preparedness.

OVERVIEW OF THE RECOMMENDED METHODOLOGIES

Four study design methodologies are represented in the recommended NAEP preparedness studies:

- content alignment;
- statistical relationships with other assessments and postsecondary outcomes data;
- criterion-based judgmental standard setting; and
- national surveys.

As previously noted, this range of design types is meant to assess the extent to which research results are confirmatory and to provide context that better enables understanding of the findings from individual studies. For each of the four methodology descriptions below, the rationale for recommending the method is followed by the method's key research questions. For illustrative purposes only, the description of each methodology concludes with a list of possible outcomes or findings the studies may produce. The Panel has extensively discussed advantages and disadvantages of each proposed study design. The following sections summarize some of the considerations and reasoning that shaped the Panel's proposed study designs. For considerations specific to the features of the individual studies, please see *Section 4: Summary of Proposed Research Studies and Prioritization*.

Content Alignment

Rationale for Recommending Content Alignment Studies

The Technical Panel recommends that content alignment studies be conducted first for each assessment used as an indicator for reporting preparedness on the grade 12 NAEP scale. In order to use other assessments as indicators of preparedness and capitalize on their preparedness research for interpretations of NAEP results, it is important to know the extent to which NAEP and any other given assessment measure similar content in a similar way. To determine the level of similarity, a content alignment study should evaluate the two assessments in terms of the knowledge framework or domain of each, the technical specifications of each, and the specific assessment objectives against which each is designed. The evaluation should compare how assessments operationalize their technical specifications in their test forms and items measuring specified objectives. Content alignment studies help identify the degree to which the objectives or content descriptors are similar between tests and the degree to which different types of items on NAEP and the other test are substantially similar or different.

Full alignment between any two different assessment programs should not be expected.

Achieve, Inc., conducted an evaluation of the grade 12 NAEP Reading and Mathematics Frameworks for the Governing Board and provided extensive recommendations regarding the content and the distribution of content appropriate for assessing the preparedness of 12th graders for postsecondary activities including both college and workplace training. These recommendations were used to develop the grade 12 NAEP 2009 Reading Framework and the grade 12 NAEP 2009 Mathematics Framework and the item pools for each.

Although the frameworks and test specifications for the 2009 grade 12 NAEP assessments in reading and mathematics were developed with the goal of measuring preparedness of 12th graders for postsecondary activities in college and the workplace, the frameworks and specifications were not directly based on other assessments traditionally used as indicators of preparedness for postsecondary activities. The purposes of NAEP are unique relative to other assessments. Alignment studies, therefore, are needed to help them determine whether NAEP and the other assessments convey sufficiently similar meaning in terms of the knowledge and skills of examinees. These studies will investigate the extent to which the content domains of mathematics and reading assessed by NAEP align with those of other assessment instruments to be used as indicators of preparedness.

Full alignment between any two different assessment programs should not be expected. The content alignment studies, therefore, are meant to provide information about the degree and type of overlap between NAEP and other assessments of interest. With information at hand about the alignment between NAEP content and items and those of other assessments, the Governing Board will be able to interpret the statistical relationships between scores on NAEP and any other assessment within the all-important context of the underlying match in content assessed, level of cognitive complexity, and areas of emphasis between the measures. This awareness of actual content being tested helps to put score relationship measures into perspective.

The Technical Panel suggested that preliminary comparability studies be conducted to provide early signals about whether a full-scale content alignment study should be conducted. Full-scale content alignment projects can begin when item-level data are available from the 2008 NAEP reading and mathematics field trials. See *Appendix D: Timelines for Completing the Studies* for additional details.

Key Research Questions for Content Alignment Studies

1. What is the correspondence between the content domain assessed by NAEP and that of the specified assessment?
2. To what extent is the emphasis of NAEP content proportionally equal to that of the other assessment?

3. Are there systematic areas of difference between NAEP and the other assessment? Are these differences such that entire content sub-domains are missing or not aligned?

By providing answers to these questions, content alignment studies will provide evidence for interpreting the relationships between performances on the two assessments.

Potential Outcomes of Content Alignment Studies

It is useful to outline the Panel's vision for the types of findings that these studies may produce. Key prospective outcomes are noted below.

- NAEP is highly aligned with all parts of Assessment X; therefore, statistical relationships between NAEP and Assessment X may allow for the interpretation of NAEP results in terms of the how Assessment X is typically interpreted.
- NAEP is moderately aligned with all parts of Assessment Y or highly aligned with only some parts of Assessment Y; therefore, statistical relationships may allow for the interpretation of NAEP results in terms of how Assessment Y is typically interpreted for only some areas or features of Assessment Y.
- NAEP is not aligned or is not sufficiently aligned with any parts of Assessment Z; therefore, statistical relationship studies will not be pursued with Assessment Z.

Statistical Relationships

Rationale for Recommending Statistical Relationship Studies

NAEP in Relation to Other Assessments – The Technical Panel recommends a series of studies aimed at statistically relating NAEP and performance on other assessments that serve as measures of preparedness for college and for job training programs in the civilian and military sectors. The Panel has recommended that statistical relationships be established in each post-secondary area for which NAEP preparedness is to be reported. These empirical studies would develop some form of statistical linking and would enable interpretations of NAEP results in relation to the other assessments of interest. The assessments of interest would be well established with strong validity research to support their use as an indicator of student preparedness for college or training for the workplace. The Panel views content alignment studies as essential to interpreting findings from this type of research.

The Technical Panel recommends that the strongest feasible form of linking should be used to establish statistical relationships between NAEP and the other assessments. It is important to note that the strongest form of linking, known as equating, will not be possible because equating involves relating scores between two tests built to the same specifications—same content, same difficulty, same reliability—which means that results are interchangeable. Because NAEP is a unique assessment with a different function and purpose (most notably, providing group rather than individual results), equating with the other assessments of interest to preparedness research is not an option. Instead, statistical relationships, such as concordance or the use of equipercentile methods to establish a working relationship, will be most likely between NAEP and other assessments. As helpful background, this subset of statistical relationships is briefly described below.

> Concordance relates scores between two tests built to *different* specifications. Scores are comparable, but not interchangeable. Concordance results cannot be used in the same way as equated scores, and scores cannot be interpreted as if they are the same. The same statistical methods may be used for producing concordance scores as for producing equated scores, but the results cannot be validly interpreted in the same manner. (Examples: This technique has been used to relate ACT to SAT scores; and Armed Services Vocational Aptitude Battery (ASVAB) scores to ACT and SAT scores.)
>
> Equipercentile methods are another way to produce comparable or concorded scores that are based on the distribution of scores on the two assessments. Scores on one assessment are related to those on the other by aligning the distribution of scores such that the average score at the 90th percentile (50th, 25th, 10th, and so forth) is aligned to the score on NAEP at each of the target percentiles. (Example: This

technique has been used to relate NAEP and Trends in International Mathematics and Science Study (TIMSS) scores.)

NAEP in Relation to Postsecondary Outcome Indicators – Additional statistical studies should also examine NAEP scores in relation to postsecondary data regarding college transcripts and employment outcomes. These additional studies would not involve other assessments and would be descriptive in nature. Hence, it may be possible to see the relationship between NAEP score levels and grades in an entry-level college mathematics class; or how NAEP examinees who elected not to go to college fared in the job market immediately after high school. These studies are recommended as a way to build helpful context for NAEP preparedness reporting and for other potential NAEP preparedness research studies.

Key Research Questions for Statistical Relationship Studies

1. Can NAEP results be statistically linked to performance on the other assessment of interest?
2. Do statistical relationships hold across the full NAEP score scale?
3. What scores or score ranges on the NAEP score scale are related to preparedness cut scores on the other assessment?
4. How do the statistical analysis results relate to findings of other studies (both within the statistical relationship studies and across other study types in NAEP preparedness research)?

Potential Outcomes of Statistical Relationship Studies

Given the research design specified above, the list below represents possible outcomes of these studies and the Panel's ideas regarding the types of findings that these studies may produce.

NAEP in Relation to Other Assessments

- NAEP results and Assessment X results are highly correlated; these findings, in conjunction with findings of well-aligned content, suggest that performance on NAEP can be interpreted in ways similar to Assessment X.
- NAEP results and Assessment Y results are only moderately correlated. There may be information from the content alignment studies that can be used to examine statistical relationships for subsets of Assessment Y in relation to NAEP or subsets of NAEP in relation to Assessment Y. These sorts of analyses may enable more limited inferences than for Assessment X.
- There is little to no correlation between NAEP results and Assessment Z results, and it is not feasible to link performance on Assessment Z and NAEP.

NAEP in Relation to Postsecondary Outcome Indicators

- For example, there is a high probability that a 12th grader scoring between S_1 and S_2 on NAEP has a performance level associated with a grade of C or better in a general education course that can be used for credit toward a four-year degree.
- For example, there is a high probability that a 12th grader scoring between S_1 and S_2 on NAEP has an income level of $35,000 or higher in employment attained immediately after completing high school.

Judgmental Standard Setting

Rationale for Recommending Criterion-Based Judgmental Standard-Setting Studies

The Technical Panel recommends studies involving criterion-based judgments by subject matter experts (SMEs) for various postsecondary activities using pre-existing sets of reading and mathematics performance standards as the criteria. These studies would use a criterion-based judgmental standard-setting process to identify points on the NAEP scale that indicate preparedness for entry-level general education courses offering transfer credits or job training programs. These judgmental standard-setting studies make use of rigorously developed statements of required knowledge and skills to guide the judgments and increase consistency within and across the judges.

The criteria used for these studies will be taken from existing preparedness standards whenever possible. Subject matter experts for these studies include secondary and postsecondary content experts as well as individuals engaged in postsecondary activities of interest, such as job-training staff members, front-line supervisors, and academic advisors.

The purpose of these judgmental standard-setting studies is to recommend cut scores on NAEP that will serve as reference points on the NAEP scale to indicate preparedness for:

- College-level general education courses en route to a four-year degree; and
- Workplace training (military or civilian) in specified occupations or occupational clusters.

Key Research Question for Criterion-Based Judgmental Standard-Setting Studies

The following is the key research question that these judgmental standard-setting studies will address:

> Based on expert judgments through a rigorously designed standard-setting process, what scores on grade 12 NAEP represent the knowledge, skills, and abilities in reading or mathematics required to demonstrate academic preparedness in the subject area for a particular postsecondary activity?

Potential Outcomes of Judgmental Standard-Setting Studies

With this research design and its key question, the list below represents possible outcomes of the studies under this methodology to illustrate the Panel's thinking of how these studies will be useful.

- Grade 12 NAEP score ranges of R (for the reading test) and M (for the mathematics test) represent reading and mathematics skill levels that are comparable to the reading and mathematics skill levels in the "readiness" range of ACT's College Readiness Benchmarks.
- Grade 12 NAEP score ranges of R (for the reading test) and M (for the mathematics test) represent reading and mathematics skill levels that are comparable to the reading and mathematics skill levels in the "success" range of the College Board's Standards for College Success.
- Grade 12 NAEP score ranges of R (for the reading test) and M (for the mathematics test) represent the reading and mathematics skill levels required to qualify for training programs for medical technicians (or other selected occupations).

National Survey

Rationale for Recommending the Survey Study

The Technical Panel recommends a survey to collect data from a nationally representative sample of two- and four-year postsecondary education institutions. The survey would collect information about the assessments used for course placement and the cut scores on widely used standardized tests for placement into college credit coursework and vocational training programs. This survey would also collect the cut scores used for placement into remedial programs in reading and mathematics and the cut scores used to indicate exemption from placement tests. The survey results will yield descriptive information related to results from other studies and provide a context for reporting NAEP preparedness research. These data may be useful as feedback to panelists in a standard-setting study for preparedness research, for NAEP work on setting achievement levels, and for informing the Governing Board's decisions on these matters. The Panel has suggested that the Governing Board also array states' course placement indicators via either this survey or a broad document review of publicly available information.

Key Research Questions for the Survey Study

1. For each widely used "standardized test," what are the cut scores used to make placement decisions in postsecondary education institutions across the country?
2. For each widely used "standardized test," are there systematic patterns and central tendencies in cut scores for institutions or programs, according to level of selectivity, subject-areas, or disciplines of study, etc.?

Potential Outcomes of the Survey Study

- The average cut score on Assessment X used by postsecondary education institutions overall is R for the reading section of the assessment. The cut scores used for placement decisions cover a relatively narrow range, and there are clear patterns by institution type that are useful to NAEP reporting.
- The average cut score on Assessment Y used by postsecondary institutions of moderate selectivity is R for the reading section of the assessment. Across all levels of selectivity, the cut scores used for

placement decisions cover a wide range, but within selectivity levels, there are clear patterns that are useful to NAEP reporting.

- The average cut score on Assessment Z used by all postsecondary education institutions is R for the reading section of the assessment. The cut scores used cover a wide range, and there is not a clear pattern of results by selectivity or any of the other stratification variables.

Model of the Interrelationships for the Group of Studies

To build a foundation of information that can be used to report preparedness on NAEP, the Panel has deliberated about how the studies should work together to produce this evidence. In this context, it is not possible to develop studies that decisively support a particular interpretation; therefore, a set of interrelated studies involving different methods is likely to provide the best indication of how to report NAEP results in levels of preparedness. This general approach is sometimes referred to as "triangulation." This is partly represented in *Figure 2-1* below, though it is important to note that the arrows pointing to the NAEP scale are only meant to show studies with results that will be related directly to the NAEP scale and are not meant to imply the specific areas of the NAEP scale related to preparedness—these specific areas (or reference points) will not be identified until the studies are complete.

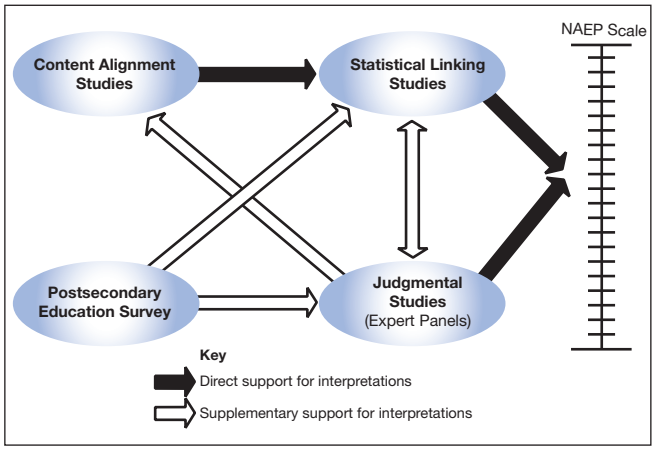


Figure 2-1. Interrelationships for the Group of Studies. *This demonstrates how the four principal study types relate to each other.*

Content alignment studies support interpretation of the statistical studies. To show this key supporting role, a solid arrow is shown above. If any of the performance standards used to set cut scores in the judgmental standard-setting studies are also directly related to an assessment program's scores, content alignment studies may provide useful evidence of content overlap that would not be apparent in a standard-setting study. For this reason, content alignment studies can be viewed as a supplement to the judgmental standard-setting studies relating NAEP to performance standards (or criteria) covered on other assessments. To represent this supplementary relationship, an outlined arrow is used. Content alignment studies do not produce results that can be interpreted directly in terms of the NAEP scale; thus, there are no arrows going directly from content alignment studies to the NAEP scale.

Statistical relationship studies and judgmental standard-setting studies both provide results that produce scores on the NAEP scale. The arrows pointing to the scale here are meant to show that the goal of these two study designs is to provide information that can be used to link various indicators of preparedness to the NAEP scale. The major arrows directed toward the NAEP scale are converging to signify the idea of mutually supportive relationships between the findings from these two types of studies. To the extent that the different studies yield similar NAEP scores, the interpretations in terms of preparedness will be strengthened.

The Technical Panel recommended that the results of the judgmental standard-setting studies and the statistical relationship studies be compared and evaluated relative to one another. These comparative evaluations show the extent to which the study results provide mutually confirmatory findings. For instance, if NAEP preparedness research were to use a set of reading and mathematics performance standards from the College Board in one of the judgmental studies, the Technical Panel would highly recommend a statistical relationship study to examine the relationship between the SAT and NAEP, and the results of this research would then be an external source of validation through another analytical lens.

The expert panels in the judgmental standard-setting studies could potentially be given results from the statistical studies to help them determine the cut scores they will recommend on the NAEP scale. Further, both the judgmental studies and the content alignment studies provide important insights into any statistical relationships identified through linking analyses. These supplementary mutually informative roles can help explain the relationships between the statistical relationship studies and the judgmental standard-setting studies. Timelines for the 2009 cycle may allow only one of these sets of studies to take place, but the goal is for both to occur eventually.

The postsecondary education survey provides contextual evidence that can be used to understand how the scores and score ranges identified in the statistical and judgmental studies relate to college course placement cut scores used across the country. The survey can also be used to evaluate the results of the statistical and judgmental studies in light of contextual evidence. This supplementary role is also displayed with an outlined arrow toward both the statistical and the judgmental studies. Because the survey data will be indirectly related to the NAEP scale through the statistical and judgmental studies, this study does not provide information directly interpretable on the NAEP scale. Hence, there is no arrow directly from the postsecondary education survey to the NAEP score scale.

Regarding all studies represented in *Figure 2-1* above, the Panel has stressed that although a plethora of data will be generated, an interpretive framework is needed to evaluate a priori whether the evidence to support preparedness statements will, in fact, be available. The following elements could be included in the interpretive framework:

(a) the strength of each type of evidence, e.g., extent of alignment, representativeness of a survey sample or respondents, qualifications of subject matter experts in a judgmental study, and level of agreement or consistency of judgments;
(b) the convergence of evidence across types of studies relevant to a targeted postsecondary activity;
(c) the number of study types contributing evidence to a targeted postsecondary activity;
(d) the consistency of evidence across studies for setting a composite cut score for a targeted postsecondary activity;
(e) the level of correspondence between a composite preparedness score and external validity evidence, e.g., course placement cut scores; and
(f) the relative strength of evidence to support inferences about preparedness for different postsecondary education and training endeavors.

A formal interpretive framework should be developed to be explicit about the reasoning for how the studies fit together and to promote transparency for the public. This interpretive framework is intended to help to make the conclusions clearer and more informative. This set of decision guides for the interpretive framework could be developed by a committee familiar with the issues of conducting preparedness research.

COMPILATION OF STUDY FINDINGS FOR NAEP REPORTING

Maximum Information from All Studies

The Panel advises that studies be designed to provide detailed information pursuant to the model of the studies' interrelationships as outlined below. For example, alignment methodologies should be sufficiently sophisticated to distinguish among direct matches, partial matches, and matches based on prerequisite skills required to address a particular reading or mathematical performance standard. A student who correctly applies the Pythagorean Theorem may actually be demonstrating three discrete skills that may be applicable in other mathematical problem solving (*Figure 2-2*). Listing these more discrete skills targeted by a performance descriptor involves substantial work, but it could provide invaluable information about which skills are covered by both assessments or which skills are covered by only one of the assessments.

Hence, if the Pythagorean Theorem, for example, were covered in only one assessment, the alignment study may reveal that the other assessment covered two of the three discrete skills applied in the Pythagorean Theorem. This would allow a "partial match" label, instead of finding no overlap for this particular content

area. Capturing these sorts of distinctions is imperative if all content alignment studies are to yield maximum information. These data can also be used to conduct sensitivity analyses to investigate the overall strength of one or more studies' results.

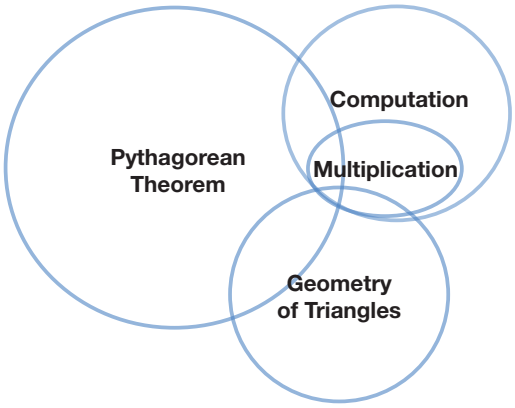


Figure 2-2. Venn Diagram for Skills Used in the Pythagorean Theorem. *This figure represents some of the discrete skills employed in mastering the Pythagorean Theorem. It also shows how these discrete skills may overlap. This demonstrates the implicit layering that may occur in academic performance standards.*

In a related recommendation, Panel members noted that information about non-overlap can be used to provide descriptive contexts for NAEP preparedness research and additional rigor in NAEP preparedness reporting. One proposal was to analyze items or content areas that appear most strongly aligned (between NAEP and the other assessment or performance standards of interest). It would also be useful to conduct analyses of the recorded reasons for SME ratings of non-alignment. Some of these efforts could be thought of as sensitivity analyses, possibly including an examination of the NAEP score when scaling NAEP with only the items that appear to be strongly aligned with the selected preparedness indicator(s) versus the NAEP score when scaling on the full set of NAEP items.

The Array of Study Results: Potential Outcomes

The proposed NAEP preparedness studies will likely yield a range of findings. Some studies may result in clear relationships whereas others may be less conclusive. The Technical Panel has discussed these possibilities and whether the full group of studies proposed for NAEP preparedness reporting is likely to produce results that provide firm guidance to the Board. The Panel has outlined some hypothetical scenarios to demonstrate the types of outcomes the Board may encounter.

Hypothetical Scenario #1: The Ideal Clustering of NAEP Preparedness Reference Points

In the ideal scenario, study results would demonstrate a convergent or logical pattern for each major postsecondary activity. The results would "cluster" around a few areas of the NAEP scale, enabling the Governing Board to formulate valid and comprehensive statements of preparedness. College preparedness reference points on the NAEP scale could cluster in one area, and workplace training preparedness reference points could cluster in the same area (*Figure 2-3*) or a different area (*Figure 2-4*). This would enable the Governing Board to formulate one to two comprehensive statements of preparedness that could be disaggregated as appropriate to fit specific contexts within college preparedness or within workplace training preparedness. A slight variant on this is a pattern with several clusters of reference points that fit logically together, where each cluster has a logical relationship to the other clusters.

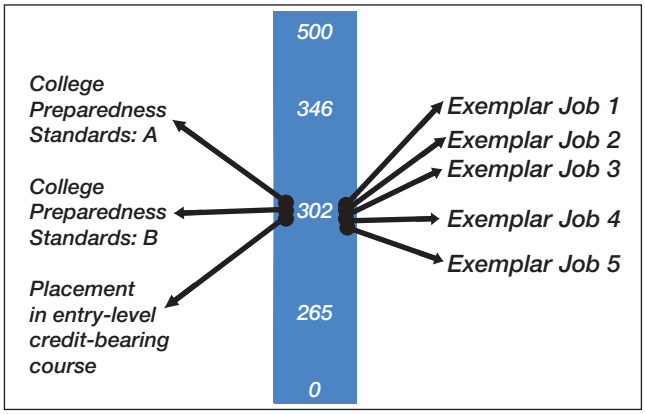


Figure 2-3. Depiction of All Reference Points Clustering in the Same Location.

Hypothetical Scenario #2: NAEP Preparedness Reference Points Clustering with Some Outliers

In this scenario, patterns emerge without the same degree of convergence as in Scenario #1 above. Here, outliers are evident. A general statement of preparedness could be used (with acknowledgement of the potential ambiguity), or it may be possible to report a larger number of statements specific to particular postsecondary contexts. One way to address outliers is to look within college preparedness, for example, at the range of reference points identified. For NAEP reporting purposes, it may be advisable to use only the reference points between the 10th and 90th percentiles of the range of identified reference points. This would focus preparedness statements exclusively on this set of reference points. Similarly, it may also be advisable to use the interquartile range, i.e., the range of reference points between the 25th and 75th percentiles of the range of identified reference points. As NAEP preparedness research evolves over the longer-term, it may be possible to conduct studies that would help to illuminate why the outliers exist. This could then lead to integrating some of the outliers into NAEP preparedness reporting over time.

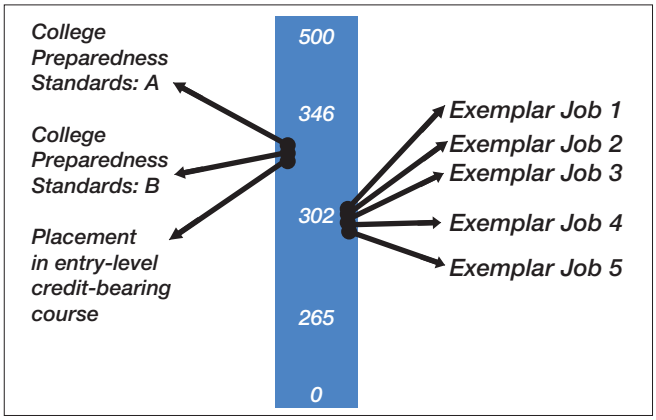


Figure 2-4. Depiction of Reference Points Clustering in a Few Locations.

Hypothetical Scenario #3: NAEP Preparedness Reference Points with Minimal Discernible Patterns

The nature of the Panel's recommended multimethod research strategy is exploratory and is intended to establish the feasibility of reporting preparedness. Therefore, one potential outcome is that reference points for both college and workplace are dispersed across the NAEP scale with little or no discernible pattern (*Figure 2-5*). In Scenario #3, broad NAEP statements of preparedness would have less meaning than in Scenario #1, where patterns of convergence are evident. This may suggest a statement of preparedness that focuses on the median reference point or the mean reference point among the range of overall reference points on the NAEP scale. A prospective report may relate performance to many of these points to more fully communicate what NAEP can say about preparedness given that preparedness expectations do in fact vary substantially with context. Given that some reference points may be individually more meaningful, it may be advisable to report preparedness on NAEP using only a small number of reference points. Alternatively, it may be advisable for none of the reference points to be used for reporting. This may also lead to consideration of whether future changes in the NAEP frameworks and assessments are needed to further support reporting preparedness of 12th graders.

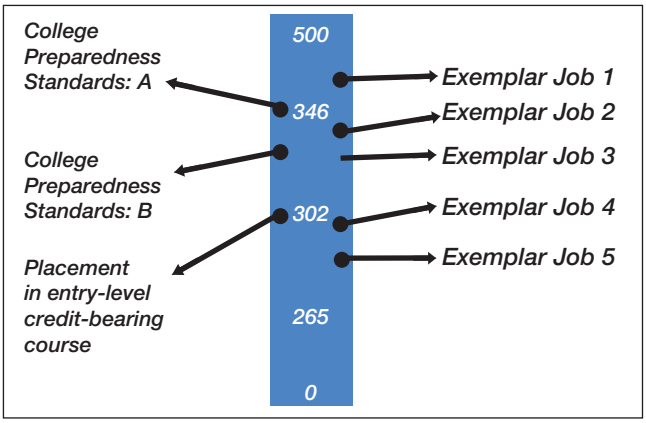


Figure 2-5. Depiction of Scattered Reference Points for Both College and Workplace Training Preparedness.

Possible Reporting Challenges for All Scenarios

Whatever pattern emerges from the study results, the Panel cautions that extreme care must be exercised to reduce the potential for misunderstanding and misinterpretation of the results. Several misconceptions may be likely, some of which have been mentioned already in this report. The following cautions are advised.

1. *Performance on NAEP relates to preparedness to be academically qualified for entry into college-level general education courses en route to a four-year*

degree or into a job training program; it does not predict admission to college or to a job training program. Nor does NAEP preparedness research strive to permit inferences about success in college or the workplace or for individual students. NAEP provides performance results for groups of students.

A comprehensive interpretive framework for using the research findings will help the Governing Board make decisions on how preparedness will be reported.

Performance outcomes in these settings are a function of a much broader array of abilities, skills, and personal characteristics, such as motivation, than are assessed by scores on NAEP reading and mathematics assessments. Admissions decisions can be highly idiosyncratic, particularly in selective settings, and these decisions often take into account factors other than academic preparedness in reading and mathematics. Complicating this further is the fact that the vast majority of postsecondary institutions admit all students who meet minimal qualification requirements, and these qualifications (based only on courses completed and grades achieved in high school) do not effectively gauge reading and mathematics skill levels. For this reason, prospective preparedness statements in NAEP reports should not be related to a student's probability of being admitted into a college or job training program. NAEP preparedness research is focused on 12th graders having the reading and mathematics knowledge and skills consistent with preparedness for college and occupational training environments as defined previously, not for admission to these environments. One way to think about a preparedness level is that the academic skills assessed in NAEP are presumed to be needed in a given postsecondary context, and the student who does not have them is severely challenged, i.e., is not prepared, in these particular respects.

2. *The data from preparedness research studies will inform the decisions on how preparedness will be reported on the NAEP scale; the data may not be the sole determinant of the decision.*

As noted earlier, this research initiative is about determining the feasibility of a relationship between NAEP performance and academic preparedness for college and occupational training. Because of the variety of data sources and types of data being proposed for NAEP preparedness research, a comprehensive interpretive framework for using the research findings will help the Governing Board make decisions on how preparedness will be reported. The research findings, on their own, will be only an intermediate step in the process.

CONSIDERING THE CHARGE TO THE TECHNICAL PANEL

In the Charge to the Technical Panel, several study designs were proposed for consideration and deliberation, and the Panel was asked to suggest additional studies. Considering each in turn, the studies mentioned in the Charge to the Panel were:

Statistical Linking Studies

The Panel has deliberated extensively on this type of study design, and it is included in the Panel's list of recommended studies for NAEP 2009 reading and mathematics.

Benchmarking Studies

The Charge to the Panel also asked Panel members to consider benchmarking studies in which NAEP might be administered to groups of interest, such as incoming college freshmen, military trainees, entry level employees of major employers or employer groups, and union apprenticeship trainees. The Panel has recommended that this study design be considered in the future, after the first set of studies for NAEP 2009 reading and mathematics is completed. See *Section 6* for a full list of the future studies recommended by the Panel.

The Panel noted factors that could complicate the identification of an appropriate sample for these studies. One complicating factor may be the existence of intervening variables, such as additional education or work experience between the time of graduation and the time that NAEP would be administered to the examinee. Another such factor would be the potential

differences in student motivation to perform well in these studies relative to their motivation in a standard NAEP administration. Differences in age and in the testing context (high school versus postsecondary setting) represent challenges as well. For the reasons mentioned above, good benchmarking studies would be difficult and expensive to conduct, and therefore not ideal for launching NAEP preparedness research. With the additional lead time for implementing this line of research and with the immense amount of information to be produced in the first set of NAEP preparedness research studies for NAEP 2009 reading and mathematics Report Cards, these studies hold potential to further inform prospective NAEP preparedness reporting. The benefit will be much higher for such studies if they are used to answer specific questions that are identified in the first round of studies.

Studies to Yield Information About Examinees Who Score Below the Basic Achievement Level

The Charge to the Panel indicated that information should be obtained about the preparedness of 12th grade students who score below the *Basic* achievement level. The multimethod research design will, by its nature, produce a series of reference points on the NAEP scale that will provide helpful information to better interpret the performance of students who score below the Basic achievement level. It is possible that a few preparedness reference points will be identified in this range of the NAEP score scale. However, to improve the likelihood that reference points will be identified in this range, the Panel urges obtaining as much information as possible; and this recommendation has been incorporated in the design of the studies related to workplace preparedness. The full range of cut scores should be evaluated, as well as the summary measure. These data may show a range of preparedness levels related to a particular job training program, which could be within the range of the NAEP score scale below the Basic achievement level. ■

Section 3

Postsecondary Activities of Interest

COLLEGE PREPAREDNESS

As stated earlier, the Panel recommends that the definition of college preparedness be focused on credit-bearing general education courses, including those eligible for transfer from two-year to four-year institutions. By definition, these courses are non-remedial, which is consistent with the Governing Board's earlier conceptualization of preparedness as being free from the need for remediation. It is important to note that the nature and degree of remediation a student may need are beyond the scope of the Panel's ideas for reporting preparedness on NAEP.

Altogether nearly half of all postsecondary education students are enrolled at two-year colleges, with an increasing number of students using the two-year college as a pipeline for four-year college transfer. Focusing on credit-bearing courses that offer transfer credits ensures the inclusion of two-year transfer programs in the determination of preparedness. The Panel's recommended study designs and definition of college preparedness address this and other key trends and issues listed below.

Issues

- *Articulation between two-year and four-year institutions* – Students seeking to transfer from two-year to four-year institutions face a range of state policies regarding transfer credits. In some states, completion of a two-year program is universally accepted by the state's four-year institutions as a signal of preparedness and eligibility. In other states, the four-year institutions may layer additional requirements that students must meet before being eligible to enter.
- *Institution-specific criteria for non-remedial course placement* – Tremendous variation exists across postsecondary institutions in the types of placement tests used, the cut scores associated with placement into non-remedial coursework, and the optional or mandatory nature of such placements. This variation exists within and between institutions within a state and across states.
- *Discipline-specific criteria for non-remedial course placement* – Within specific disciplines or college majors, specific criteria may be applied to determine whether a student needs remediation. For example, a solid grounding in trigonometry may be an essential element for preparation in a subset of fields of study. In some states, articulation agreements governing transfer credits from two-year to four-year institutions are specific to each discipline as well.
- *Eligibility for non-remedial course placement versus eligibility for admission to postsecondary education institutions* – Admissions selectivity criteria are separate from criteria used for non-remedial course placement.
- *Varying levels of admissions selectivity* – Two-year institutions and some four-year institutions are "open admissions" because they admit all applicants. Four-year institutions are more likely to impose criteria for the students they admit, and these criteria may reach beyond reading and mathematics knowledge and skills.
- *Diversity among two-year college institutions* – The mission of community colleges varies widely, with some functioning as a major source of job training within a community and others focusing on preparation for four-year institutions. Generalizing across all community colleges and treating them as a uniform set of institutions could hamper results of NAEP preparedness studies.

- *Lack of systematic course placement data* – Use of popular placement assessments such as ACCUPLACER, ASSET, and COMPASS varies dramatically across institutions. Each may house its own data and determine its own cut scores. Further, some institutions use locally developed placement instruments, which may even be specific to particular academic departments of the college.

The list of challenges here is not exhaustive, but it represents key considerations the Panel has examined while crafting study recommendations.

Responsive Strategy: Use Preparedness Indicators to Locate Points on the NAEP Scale

To address these issues, the Technical Panel has determined three critical resources necessary to support NAEP preparedness studies with respect to college preparedness:

- *Assessment instruments* widely recognized as indicators of college preparedness, such as ACT and SAT, as well as assessments used for course placement, such as ACCUPLACER, ASSET, and COMPASS, can be studied in relation to NAEP.
- *Preparedness standards and benchmarks* are available, and several of these sets of standards have been compiled through rigorous processes and are widely reflective of the reading and mathematics knowledge, skills, and abilities needed to be prepared for non-remedial college-level general education coursework.
- *Subject matter experts* across subject areas, i.e., reading and mathematics, as well as across specific postsecondary education roles, i.e., professionals specializing in remedial placement, will identify the appropriate academic performance standards and postsecondary education settings.

The study designs proposed by the Technical Panel for college preparedness make extensive use of these three resources. With an emphasis on placement in non-remedial college coursework, it is important to acknowledge one possible outcome of the studies focused on college preparedness: the composite range of cut scores on a particular assessment indicating non-remedial placement across various postsecondary education settings could cover a large range.

WORKPLACE PREPAREDNESS

Many of the jobs students can obtain immediately after high school graduation offer advancement potential via a career pathway and the eventual capacity to earn a wage sufficient to support a family after appropriate training. Because many of these jobs require considerable training, the National Commission on NAEP 12th Grade Assessment and Reporting proposed in 2004 that prospective NAEP preparedness reporting focus on job training. An array of occupational training options is available: on-the-job training, an in-house training program, a formal apprenticeship program, a training program in a community college, or training in a vocational institute or program. The Technical Panel's definition of workplace training encompasses all of these. Challenges to targeting such programs are listed below.

Issues

- *Addressing the diversity of occupationally oriented postsecondary education paths* – It is important to recognize that preparedness addresses training in an organization (e.g., on-the-job training and apprenticeship) as well as vocational training at a community college or institute.
- *Distinguishing qualifications for the job from qualifications for the job's training program* – Preparedness to be hired for the jobs of interest and preparedness to enter the training programs for the jobs of interest are not the same.
- *Identifying appropriate resources for studies* – Many resources that can be used for NAEP preparedness studies target qualifications for the job, which means these resources would require further refinement to target qualification for job training, which is the goal of NAEP preparedness research. Further, most research to develop academic standards for occupations focuses on higher performance levels or success in the occupation.
- *Identifying training programs with national scope* – For national reporting about preparedness, the qualifications, academic content standards, and

assessments for any given occupation that NAEP targets should be consistent nationally to ensure that preparedness for the occupation means the same thing across the country. Many occupations do not have a nationally consistent training core.

- *Identifying occupational training profiles equivalent in military and civilian sectors* – Equivalence between similar occupations in the military and civilian sectors cannot be determined without analyzing the occupational profiles in depth. A crosswalk for military and civilian jobs has been established using a common coding scheme. The equivalence of jobs from the two sectors should be confirmed for the job training programs selected for NAEP preparedness. For occupations that generally require the same qualifications across the military and civilian sectors, there may be some differences across the two sectors because of the different environments for the occupation.
- *Addressing occupations with differing academic emphases* – Some occupations require substantial geometry, whereas others may focus more heavily on algebra or simple numerical computations, for example. The NAEP scale incorporates all of these mathematics skills. If only a subset of such skills is required, this may not align with NAEP scaling procedures. The feasibility of setting preparedness standards for each of the major sub-domains (e.g., algebra, statistics, etc.) should be investigated.

Responsive Strategy: Use Exemplar Occupations

In addressing these challenges to targeting workplace preparedness, the Panel has crafted a research strategy to identify five to seven occupations to serve as exemplars, selected on the basis of criteria. These exemplars are the occupations deemed most informative for estimating the entry-level reading and mathematics requirements for mulitiple sectors of the labor force. Appropriate subject matter experts (SMEs) would identify points on the NAEP scale representing the entry-level reading and mathematics qualifications needed for training in the respective occupations. If, for example, air conditioning technician were one of the exemplar occupations, a training program for that occupation would be identified, and a group of appropriate SMEs would be engaged in standard-setting procedures to identify the cut scores on the NAEP scales. The greater the extent to which the exemplar occupations represent a broad range of critical occupational categories, the greater the potential of the NAEP preparedness statements to provide information about the preparedness of 12th graders for entry into job training.

Identification of these five to seven exemplar occupations should be considered the first phase of the process, with an incremental approach leading to more exemplar occupations that would eventually represent the full population of relevant occupations.

In line with this strategy to use exemplar occupations, the Technical Panel has identified critical resources to support NAEP preparedness studies with respect to workplace preparedness:

*The Occupational Information Network (O*NET) taxonomy* is the nation's primary source of occupational information. Central to the program is the O*NET database, containing information on hundreds of standardized and occupation-specific descriptors, including information regarding typical income of an occupation's workers and the typical educational requirements for qualification. The database is continually updated by surveying a broad range of workers from each occupation. For the purposes of NAEP preparedness research, the focus will be on two classification zones of O*NET: O*NET job zones 2 and 3. A dental assistant is an example of a zone 2 occupation requiring three months to one year of job training. A reading skill needed for this occupation is the ability to understand written sentences and paragraphs in work-related documents. A construction manager is an example of an O*NET zone 3 occupation requiring one to two years of job training. Mathematics requirements for this occupation include knowledge of arithmetic, algebra, geometry, calculus, statistics, and their applications. See the informational endnote on *O*NET Zones for NAEP Preparedness Research to learn more about zones 2 and 3.* [1]

Exemplar identification methodologies from previous research include models for this kind of exemplar-anchored research strategy. The data

needs for this type of research can be found in the American Diploma Project Benchmarks by Achieve (2004), which used an exemplar occupation approach for workplace preparedness.

Assessment instruments widely recognized as indicators of workplace preparedness, such as ACT WorkKeys and ASVAB, as well as assessments used to indicate qualification for specific training programs, can be studied in relation to NAEP.

Preparedness standards and benchmarks can be used as sets of standards compiled through rigorous processes and reflective of the knowledge, skills, and abilities needed to qualify for entry to job training programs.

Subject matter experts across subject areas, i.e., reading and mathematics, as well as across specific workplace contexts, i.e., front line supervisors and job training program instructors or administrators, will support strong connections to the appropriate academic performance standards and workplace settings.

Strategic Principle #1: Implement criteria for the features of occupations most useful for NAEP preparedness reporting.

In extensive discussion about the desired traits discussed by the Board and other traits of occupations, the Technical Panel proposes the following set of criteria to be used in selecting exemplar occupations.

- *O*NET zones 2 and 3* – These zones collectively represent occupations requiring three months to three years of training. The Panel judges that these educational requirements are most appropriate to the goals of NAEP preparedness reporting.
- *Availability of civilian and military counterparts* – To promote reporting across military and civilian sectors, it is necessary to select exemplar occupations that have cross-sector counterparts. Identifying counterpart occupations in the military sector would be a key consideration throughout the process of identifying exemplar occupations.
- *Coverage of industry sectors* – Exemplars should come from a broad range of industries to the extent possible.
- *Recognition of occupations* – To assure usefulness and relevance to the public, exemplars that are familiar to the public should be selected.
- *High employment level projected into the future* – In absolute numbers, the exemplars should cover a large proportion of jobs and job openings, and should not be projected to decline in the future. Focusing solely on high growth rates can be misleading when they represent a small proportion of jobs and fluctuate from year to year. Therefore, occupations with high growth rates should be considered only when they also represent a high proportion of jobs.
- *Coverage of reading and mathematics preparedness* – Taken together, the exemplars should represent a range of reading and mathematics skills along the NAEP scale. This may result by default from application of the previously mentioned proposed criteria, but it will be important to explicitly review this at some point in the exemplar selection process.
- *Representation of different training paths* – As noted earlier, it is important to represent apprenticeships as well as vocational training or community college programs.

Strategic Principle #2: Focus on occupations with job training entry requirements that are interchangeable between military and civilian sectors.

The nationally consistent core of the military's training programs could prove invaluable to the NAEP preparedness studies. As noted above in the list of issues, national consistency in job training is a challenge, i.e., whether a training program is administered in a standardized manner across the country or whether its entry requirements are uniformly applied across the country. The Technical Panel suggests taking advantage of the vast resource of the military's training programs. However, the Panel has also noted that starting the exemplar selection process and analysis in the military sector (and then cross-referencing to the civilian sector) may be especially burdensome, given the limited time availability of military personnel who may serve as subject matter experts.

Suggestions for the Exemplar Identification Process

To identify exemplars, the Panel suggests starting with a listing of all occupations that require from three months to three years of training (collectively known in O*NET as zones 2 and 3). Then, remove occupations from the list if they do not have the desired growth rate and employment level (number of jobs in the national economy). Next, review the list of remaining occupations to identify whether there are occupations that have a nationally consistent set of training standards or training-program entry qualification criteria that can be used as targets in NAEP preparedness research. The group of occupations can be further reduced by selecting those that maximize diversity—across industries, for example—while also meeting other criteria of interest to the Board.

This set of exemplar occupations can be related to similar occupations in the same category through an analysis of the O*NET taxonomy of occupational classifications. Using this taxonomy, NAEP preparedness research could potentially generalize statements of preparedness for these selected occupations in relation to similar occupations outside of the final set of exemplars.

To assist exemplar selection and the operationalization of these exemplars into reference points or ranges on the NAEP scale, the Panel agrees that there should be an extensive review of the availability of industry training standards, such as those from the National Automotive Technicians Education Foundation. This review will help to identify training programs that are nationally standardized, which, as noted above, is a key challenge for workplace preparedness research. Other resources to identify industry training standards may include:

- Companies supplying temporary workers – Some companies place professional and permanent workers as well as temporary workers.
- Employer associations – Some companies provide services for employers who do not have in-house programs.
- Labor unions – Several unions provide training as well as assistance for new apprentices that cover the prerequisite knowledge and skills needed for an occupation.
- Professional associations – Some, such as the American Electrical Association, have conducted analyses of exemplar occupations within their respective fields, and their approach may serve as a model for identifying the types of data to collect in a study focused on exemplar occupations.
- Entities certifying job training programs (inside and outside of community college settings) – Some, such as the National Council for Continuing Education and Training, oversee many programs in community college settings that provide an associate's degree concurrently with job training certification.

Collaborations with the Department of Defense

The Technical Panel sees great need for military collaboration to maximize the success of the NAEP grade 12 preparedness research initiative. The Panel recommends asking senior leaders of both the U.S. Department of Education and the U.S. Department of Defense for support. The demands on military personnel and other resources related to war and

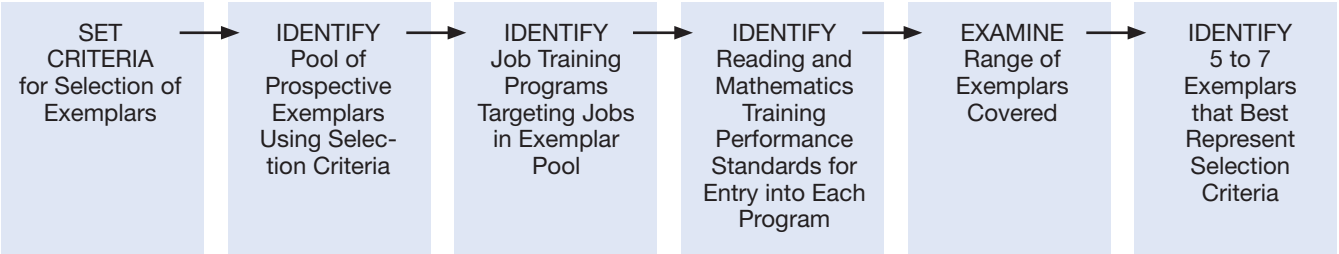


Figure 3-1. Process to Identify Exemplar Occupations. *This depicts the key parts of the process to identify exemplar occupations. Depending on the range of exemplars identified, deliberate efforts may be needed to widen the set of exemplars. For example, a national set of industry training standards may not be available for all potential exemplars. This may require developing appropriate reading and mathematics knowledge and skills statements via subject matter expert panels. These statements would then be used to set cut scores on NAEP.*

national security may be a formidable challenge to their involvement in NAEP preparedness research and must be taken into account.

SPECIAL CHALLENGES TO THE RECOMMENDED METHODOLOGIES

The Governing Board has indicated a special interest in studies to establish statistical relationships between NAEP and other external indicators of preparedness. The Technical Panel has devoted extensive discussion to design issues, and has invited company representatives and state assessment staff members to provide details that would further clarify those issues. A major concern has been assessment organizations' protocols regarding confidentiality of test questions and related assessment materials. This concern makes it difficult to maintain a rigorous and consistent study design across multiple assessments. The Technical Panel has met with representatives of ACT and the College Board to discuss feasibility issues in great detail. Board staff members will continue efforts to identify the feasibility of collaborations for the proposed studies. The Panel encourages the Governing Board to continue to identify sources of data that can be shared for NAEP preparedness research.

The following is a list of challenges that were identified in the Panel's internal deliberations and conversations with external partners.

Collaboration Opportunities for Accessing Data

Materials from Other Assessments – Agreements are needed to secure materials required for content alignment studies, for example. Although the Technical Panel has called for consistency across content alignment studies to the extent possible, there are some fundamental challenges. Assessment companies are protective of their items, test specifications, and blueprints, citing proprietary considerations or long standing institutional philosophy and practice.

Score Data for Other Assessments – Collaborative partnerships may be difficult to establish, but are essential to accessing score data for other assessments to enable the statistical relationship studies. A comprehensive longitudinal database, such as Florida's, is one way to access the data needed for the statistical relationship studies to relate NAEP performance for 12th grade examinees with scores on other assessments. These other assessments of interest could have been taken either before (college admissions tests) or after (course placement tests) high school was completed. Alternatively, collaborative partnerships can be formed directly with testing companies housing the needed score data. Both of these routes may be difficult, but they are necessary for statistical relationship studies.

Incorporating a Variety of Assessments – Although one study proposed by the Technical Panel will identify SAT examinee matches across the 12th grade national NAEP sample, relatively few students in the central states of the nation take the SAT. Although a parallel national study for ACT data is not yet feasible, the Florida database provides the potential for establishing a state-level statistical relationship between NAEP and the ACT. A full-scale content alignment study in which an independent organization would compare ACT specifications and items to the NAEP framework and items is not yet feasible, but the Panel recommends proceeding with a statistical study involving these ACT data for Florida. The Technical Panel has also recommended that Governing Board staff members determine if any of the grade 12 NAEP pilot states have large samples of students taking ASVAB.

Technical Challenges in Study Design

Content Alignment and the Newer Context of Test-to-Test Alignment – Traditionally, content alignment studies have been conducted to evaluate the relationship between an assessment and a set of content specifications (or frameworks). Examining the alignment between two tests represents new territory. To identify the best methodology for NAEP preparedness research, the Technical Panel recommended convening an advisory group to identify the key elements of content alignment studies and how they should be standardized to assure procedural validity.[2] The Panel considered a wide range of issues specific to NAEP preparedness research, such as the areas of design specifications that should be standardized to support consistency across content alignment studies. The

Panel also noted the need to avoid over-specifying the design and to allow flexibility so that studies can yield information about degrees of alignment. A related issue is the need to ensure consistent usage and operationalization of terms that tend to be used in a variety of ways across education contexts. [3]

Content Alignment and the Newer Context of Alignment with Computer Adaptive Tests – Computer adaptive assessments are often used to efficiently gather information about the knowledge, skills, and abilities of entering college students. These assessment instruments involve a large number of test items, and there is limited research addressing how to conduct alignment studies with these instruments. The Panel recommends that the Board solicit focused advice in this area to ensure a sound and rigorous design.

Statistical Relationship Studies and Differences in Assessments – As previously stated, NAEP is unlike other assessments of U.S. students. NAEP includes constructed response items in all subject assessments, whereas assessments such as those developed by ACT and the College Board are largely or solely based on multiple choice items. Only NAEP has a nationally representative sample of students overall. Some assessments, such as the ACT and SAT, are national in scope, but usage is concentrated geographically and generally limited to college bound students. In some cases, these tests are mandated for all high school students. Despite these cited differences, it should be possible to establish meaningful statistical relationships.

Judgmental Standard-Setting Studies and Commitments of SME Panelists – NAEP preparedness studies will depend on collaboration from postsecondary education course placement professionals; business community and military personnel representatives; and content experts. For some of the studies, multiple expert perspectives may be needed on the same SME panel. Military personnel representatives may be least likely to be available for these studies. (If exemplar occupations are in the Air Force or the Navy, that could be beneficial as these branches are potentially more available than the Army during this time of war.) In considering potential SMEs knowledgeable of entry requirements related to ASVAB, instructors at military schools and former ASVAB technical panel members could conduct the standard setting on NAEP if those choices are acceptable to ASVAB leadership. ■

Section 4

Summary of Proposed Research Studies and Prioritization

OVERVIEW OF THE RECOMMENDED STUDIES

The individual studies proposed for each recommended study design are presented below. An abbreviated listing of studies can be found in *Appendix E: Recommended Studies for 2009 NAEP 12th Grade Preparedness Reporting*. A priority label of "high" or "medium" is affixed to each list, and the full set of prioritization criteria follows.

The Technical Panel recommends seven assessments for analysis in NAEP preparedness research. These assessments are the primary focus of the proposed content alignment studies and statistical relationship studies below. In each case, *only the mathematics and reading portions of the assessments would be used.*

- ACCUPLACER is a computer adaptive test used for college course placement decisions in two-year and four-year institutions. It is produced by the College Board, and includes assessments of sentence skills; reading comprehension; computational skills in arithmetic, elementary algebra, and college-level math; and essay-writing skills.
- The ACT assessment is a college admissions test used by colleges and universities to determine the level of examinees' knowledge and skills in subject areas including reading, English, and mathematics. (*College Readiness Standards and Benchmarks* connect reading or mathematics knowledge and skills—and probabilities of a college course grade of "C" or higher (65%) or "B" or higher (50%)—with particular ACT score ranges.)
- ACT WorkKeys is an assessment employers use to evaluate the knowledge and skills of a prospective employee relative to an occupation's profile. WorkKeys, a product of ACT, covers communication (business writing, listening, reading for information, writing) as well as problem solving (applied technology, applied mathematics, locating information, observation). There is also an interpersonal skills section of WorkKeys.
- Armed Services Vocational Aptitude Battery (ASVAB) is an assessment the military uses to evaluate the knowledge and skills of prospective recruits. The assessment tests 11 areas: general science; arithmetic reasoning; word knowledge; paragraph comprehension; numerical operations; coding speed; auto and shop information; mathematics knowledge; mechanical comprehension; electronics information; and sum of word knowledge and paragraph comprehension. Two areas are being phased out: numerical operations and coding speed.
- ASSET is a course placement assessment designed for use in two-year institutions. ACT produces ASSET, which covers basic skill measures (writing, numeracy, reading), advanced mathematics (elementary algebra, intermediate algebra, college algebra, and geometry), and additional skills (specifically requested by institutions in areas such as chemistry or for portions of the ACT assessment).
- COMPASS is a computer adaptive test used for college course placement decisions. Produced by ACT, its scores can be related to ACT's *College Readiness Standards and Benchmarks* for English (writing), reading, and mathematics subtests.
- The SAT reasoning test is a college admissions assessment produced by the College Board. Colleges and universities use it to evaluate the knowledge and skills of applicant pools in critical reading, mathematics, and writing. (The College Board has recently conducted an item mapping exercise to extrapolate performance descriptors that relate to various score ranges on the SAT. This will benefit NAEP content

alignment research because it will enable analysis with the assessment overall, as well as with the distribution across various score ranges.

Content Alignment

Goal of This Methodology

The goal of content alignment studies is to examine the structure and content of other assessments relative to NAEP to determine whether the scores on NAEP and the other assessments convey similar meaning in terms of the knowledge and skills of examinees.

Features of Content Alignment Study Designs for NAEP Preparedness Research

Content alignment studies will involve convening panels of appropriate subject matter experts to identify and describe the content similarities and differences between grade 12 NAEP and relevant assessment instruments. The subject matter experts may draw their expertise from the subject area (reading or mathematics) and from their experience with the requirements for the postsecondary activity of interest (college or workplace). Key recommendations from the Technical Panel regarding this study design follow:

- Studies should be conducted by a *third party* (an independent organization separate from the assessment companies themselves) to increase the credibility of the results.
- *Preliminary content alignment studies* should be conducted to provide an early signal of the feasibility of a full-scale content alignment study. (One such study, between NAEP and SAT in mathematics and reading, has already been completed, and indicated that a full-scale study is warranted.)
- *A pilot study* should be conducted for at least one subject area of each testing program (ACCUPLACER, ACT, etc.) to work out methodological issues in advance and thereby to improve consistency, rigor, and efficiency of the content alignment.
- To the extent possible, one *consistent methodology* should be used for the alignment studies to facilitate comparisons across different studies and decisions about how each study's findings are employed to report preparedness on the NAEP scale.
- *Degrees of alignment* should be identified to provide stronger context for what score relationships mean and to support the Board's possible decision to use another assessment as a tool by which preparedness can be reported on NAEP.
- *A two-directional process* should be part of full-scale content alignment studies because it maximizes information regarding content overlap and measurement coincidence.
- *A modified version of the Norman Webb alignment methodology* appears appropriate for NAEP preparedness research. The Panel considered other methods and evaluated their suitability for NAEP preparedness research. Porter and Smithson's (2001) method was developed to produce an overall index of alignment for documents (standards, assessments, and curricula), and has been applied across schools, districts, and states. The Panel concluded that the comparison framework was not ideally suited for the NAEP preparedness research. Brown and Conley (2007) used a modified Webb methodology, and the Panel saw several useful features of these modifications for NAEP preparedness research. The Technical Panel examined alignment issues and concluded that NAEP alignment studies would require modifications of any current methodology. Given that, the Panel recommended the Webb method because it appears to include all the necessary features and is most widely used. Additional discussion related to the alignment methodology appropriate for NAEP preparedness research can be found in *Appendix F: Summary Report on Recommendations of Expert Content Alignment Group*.

  The Governing Board has commissioned Norman Webb to draft a design document that can be used as the basis for designing all content alignment studies in NAEP preparedness research. The Panel recommends that this document be vetted by a variety of alignment experts in the field before it is used for the studies. This will ensure that the latest advances and refinements in the content alignment methodology can be applied in NAEP preparedness research, and will assist in overall procedural validity.

The Panel is aware that tradeoffs are likely. If it is necessary to obtain critical data for key areas of preparedness research, these recommendations may be

relaxed to support coverage of these key topics. This step should be taken on a case-by-case basis when limited informational materials are available for a full-scale content alignment study. The recommended features above are the goal for all studies, however.

COLLEGE

Recommended for NAEP 2009 (High Priority)	Feasible for NAEP 2009
NAEP-ACT Content Alignment [4]	✓
NAEP-SAT Content Alignment	✓
NAEP-ACCUPLACER Content Alignment	✓
NAEP-ASSET Content Alignment	
NAEP-COMPASS Content Alignment	

WORKPLACE

Recommended for NAEP 2009 (High Priority)	Feasible for NAEP 2009
NAEP-ASVAB Content Alignment	
NAEP-WorkKeys Content Alignment [5]	✓

Figure 4-1. Content Alignment Studies for Reading and Mathematics. *This shows content alignment studies that are recommended and feasible.*

Additional Details for Feasible 2009 Content Alignment Studies

A third party has not yet been identified for the NAEP-ACCUPLACER and NAEP-SAT content alignment studies, but the College Board has recommended and fully supports this approach.

ACT will conduct a preliminary content alignment study for the comparison between NAEP and WorkKeys.

ACT will also conduct a modified content alignment of NAEP in relation to ACT College Readiness Benchmarks in reading and mathematics, involving an item classification study in which NAEP items are classified into one of seven score bands on the ACT scale:

- "College Ready"
- Three score bands above "College Ready"
- Three score bands below "College Ready"

After the classification study, the items will be mapped into ranges on the NAEP scale in order to describe where these categorized items fall on the NAEP scales.

Statistical Relationships

Goal of This Methodology

The goal of statistical relationship studies is to link NAEP to assessments "that are indicators of preparedness." To the extent that NAEP can be linked to existing assessments that have interpretations related to preparedness (e.g., college placement testing programs), NAEP scores or score intervals can be linked to these existing interpretations.

Features of Statistical Relationship Study Designs for NAEP Preparedness Research

The initial recommended study design called for embedding other assessments in the administration of grade 12 NAEP. In developing the specifications for this study design, there were hurdles that prompted the Panel to recommend an alternative plan that instead collects previous scores on other assessments for NAEP examinees. The Panel concurred that this alternative to the embedded forms study is better because it does not alter standard testing conditions of either assessment.

The recommended statistical studies for NAEP preparedness research are focused at the national level and at the state level for one state with a longitudinal database housing postsecondary information about students leaving high school. The state longitudinal database provides access to student scores for several assessments of interest to NAEP preparedness research. The research design will connect NAEP records to the state's longitudinal database to capture postsecondary data including scores on other relevant assessments.

The goal is to eventually identify more states with comprehensive longitudinal databases. A state-representative NAEP sample for grade 12 reading and mathematics is needed for this study design, and the NAEP pilot study of grade 12 NAEP at the state level in 11 states provides additional opportunities for connections to longitudinal databases. Florida has indicated a willingness to share its data for NAEP preparedness research, and Florida is one of these pilot states.

Longitudinal state data, such as Florida's, provide the potential to estimate points on the NAEP scale to represent preparedness for several aspects of postsecondary education and employment. As a result of these studies, it may be possible for NAEP Report Cards to relate specific scores or score ranges to preparedness for specific postsecondary activities related to college and the workplace.

COLLEGE

Recommended for NAEP 2009 (High Priority)	Feasible for NAEP 2009
NAEP-ACT Statistical Linking	✓
NAEP-SAT Statistical Linking	✓
NAEP-ACCUPLACER Statistical Linking	✓
NAEP-ASSET Statistical Linking	
NAEP-COMPASS Statistical Linking	

WORKPLACE

Recommended for NAEP 2009 (High Priority)	Feasible for NAEP 2009
NAEP-ASVAB Statistical Linking	
NAEP-WorkKeys Statistical Linking	✓

Figure 4-2. Statistical Relationship Studies for Reading and Mathematics. *This shows statistical relationship studies that are recommended and feasible.*

Additional Details for Feasible 2009 Statistical Relationship Studies

NAEP in Relation to Other Assessments

National sample analyses for the NAEP-SAT study will be possible based on a data partnership with the College Board. The College Board has indicated a willingness to support a NAEP preparedness study with ACCUPLACER data, but a study design has not yet been developed to address the challenges related to a computer adaptive test. National samples for other assessments (ACT, ASSET, ASVAB, COMPASS, WorkKeys) are not yet possible, as data partnership agreements have not been secured.

State sample analyses for Florida will be possible for statistical linking of NAEP-ACCUPLACER, NAEP-ACT, and NAEP-WorkKeys in reading and mathematics. The link to the SAT data will be part of the national study including representative samples of 12th graders in Florida and 10 other states.

NAEP in Relation to Postsecondary Outcome Indicators

National sample analyses will be possible based on the High School Transcript Study (SAT scores) routine data collection as well as the College Board's data on SAT examinees.

State sample analyses for Florida will be possible for NAEP performance relative to college transcript data and employment data.

Judgmental Standard Setting

Goal of This Methodology

The goal of the criterion-based judgmental standard-setting studies with subject matter experts is to use existing sets of academic performance standards in conjunction with NAEP assessments to set cut scores on the NAEP score scale. Separate judgmental studies may also be needed to establish sets of academic performance criteria if pre-existing standards are not available for postsecondary activities of interest. For each assessment program, these judgmental standard-setting studies are to be preceded by pilot studies to support successful implementation.

Features of Judgmental Standard-Setting Study Designs for NAEP Preparedness Research

Using NAEP data, a feasibility or pilot phase for each study would assess the relationship between NAEP assessments and other sets of standards as the criteria for preparedness. Pilot studies would also help identify necessary modifications in the planned study design.

If the research pilot results indicate that the study design is feasible, a standard-setting process would be conducted to set "preparedness" cut scores on the NAEP scale by using the specified sets of standards as the criteria. For some studies, the standards would need to be developed. This qualitative research will provide comparative data for the results of other NAEP preparedness studies.

One example of an academic training performance standard for auto mechanics is the ability to use a variety of technological and information resources (e.g., libraries, databases, computer networks, video) to gather and synthesize information and to create

and communicate knowledge. Subject matter experts should have critically relevant experience with and exposure to the type of postsecondary activity at hand, whether it is for an occupation, such as auto mechanic, or for a particular postsecondary education setting, such as entry-level general education courses at a four-year college. Because academic qualifications for entry into job training are to be targets of a judgmental standard-setting procedure, the SME panel should include members with expertise clearly related to these entry requirements. Front line supervisors, for example, would have an appropriate perspective to include in standard-setting studies focused on workplace.

In noting the time commitment this would require for SME panelists, the Technical Panel acknowledged that support for this type of research is increasingly difficult to obtain from prospective SME panelists in the military. As noted earlier, the Technical Panel suggested that instructors at military schools and former ASVAB technical panel members could conduct the standard setting on NAEP if those choices are acceptable to ASVAB leadership.

The Technical Panel expressed concern that enlisting subject matter experts who are also experts in either NAEP or a study's other assessment of interest may be a source of bias within SME alignment panels. The Technical Panel considered several alternatives for the composition of the alignment panels and ultimately suggested that the panels be selected to include subject matter experts who do not have extensive ties to either NAEP or the other assessment. Assessment expertise would be represented in the facilitation staff members for the process.

Comparing information from different studies within college preparedness and within workplace training preparedness is central to the Panel's recommended research strategy. To provide comparative information for judgmental standard-setting studies where no statistical relationship studies are yet feasible, replicate panels should be an additional feature of the judgmental studies. This recommendation should be followed for both workplace training judgmental studies as well as college course placement judgmental studies because these two areas will not be subjects of a statistical relationship study based on a national sample. Hence, these areas can be reinforced via replicate panels of subject matter experts in the judgmental standard-setting studies.

COLLEGE

Recommended for NAEP 2009 (Medium Priority)	Feasible for NAEP 2009
Standard Setting: ACT College Readiness Standards	✓
Standard Setting: College Board Standards for College Success	✓
Development of Performance Standards: Course Placement Standards	✓
Standard Setting: Course Placement Standards	✓

WORKPLACE

Recommended for NAEP 2009 (High Priority)	Feasible for NAEP 2009
Identification of 5–7 Exemplar Occupations	✓
Development of Training Performance Standards: 1–7 Exemplar Occupations	✓
Standard Setting: 5–7 Exemplar Occupations	✓

Figure 4-3. Judgmental Standard-Setting Studies for Reading and Mathematics. *This shows judgmental standard-setting studies that are recommended and feasible.*

Additional Details for Feasible 2009 Judgmental Standard-Setting Studies

Preparedness for Workplace Training – The judgmental standard-setting studies related to workplace are centered on exemplar occupations. The studies listed in *Figure 4-3* will identify occupations and related job training programs in the civilian and military spheres that would be useful as exemplars in reporting grade 12 NAEP results in reading and mathematics (the assumption being that a job for an occupation in the civilian sphere can be determined to be like another job for the same occupation in the military). Twenty occupations will be identified, from which the Governing Board will select five to seven exemplars. The Technical Panel recommends that WorkKeys scores be collected for WorkKeys job profiles associated with the 20 occupations identified in this research.

A separate complementary study will use a judgmental standard-setting procedure to identify the score locations (i.e., cut scores) on the grade 12 NAEP reading and mathematics scales that represent the skills and abilities needed to qualify for the respective job training programs. The five to seven selected exemplar occupations will be used in this phase of the project. Subject matter experts will develop the criteria to use in these studies.

Preparedness for College – The Governing Board hopes to be able to compute cut scores related to the ACT College Readiness Benchmarks in reading and mathematics based on the item classification data resulting from the preliminary content alignment study outlined above.

Detailed specifications on the *College Board Standards for College Success* are still being developed, as is the study design for the judgmental process to devise academic performance standards for course placement. These details can be part of the contracting process.

National Survey

Goal of This Methodology

The goal of the survey of postsecondary education and job training institutions is to identify cut scores used for placement in non-remedial courses and exemption from placement tests.

Features of the Survey Design for NAEP Preparedness Research

This survey will focus on a nationally representative sample of two-year and four-year colleges. Data will be collected regarding the use of course placement assessments, such as ACCUPLACER, ACT, ASSET, COMPASS, and SAT. The respondents in the college sample will provide information about the assessments and the cut scores used for placement in credit-bearing courses in mathematics and reading, as well as placement in remedial or developmental courses. The survey will ask whether other assessments are used for placement decisions.

The Panel has identified several specifications for this survey study that will enhance its value:

- Include urbanicity, size, and other institutional characteristics (such as public/private, two-year/four-year, and level of selectivity) as stratification variables.
- Collect course placement information related to earning certification in a vocational/technical program, as well as earning an Associate of Arts (AA) or Associate of Applied Sciences (AAS) degree.
- Include a section in the survey to collect information about certification programs in target occupations. (Possibly an entire survey could focus solely on placement criteria, assessments, and cut scores used for students entering training programs for target occupations.)
- Perhaps send the survey to recipients in conjunction with a letter of support from the National Association for Developmental Education to encourage higher response rates.
- Add a small field trial to the study design to identify assessments other than ACCUPLACER, ACT, ASSET, COMPASS, or SAT that may inform placement decisions. This should be used to finalize the operational survey and ensure that it gathers helpful information about all relevant placement instruments.

In addition, the Panel has noted the difficulty in designating vocational or technical certification courses as either “higher education” or “workplace training” postsecondary activities. The academic requirements of some certification programs may exceed those of the AA or AAS degree, although the degree is not conferred upon completion of the certification program. It will be useful to conduct a small field trial for certification programs to learn more about the academic requirements for their course placement. Without such an exploration, it may be necessary to focus on students entering AA or AAS programs, because their entry requirements are more well-defined.

The Technical Panel acknowledges that the range of placement scores is likely to be great, even within a single institution. The survey study has potential for providing information regarding the set of plausible scores that may result from the judgmental standard-setting studies. The Panel cautioned that the results of this survey will be difficult to interpret; within and across institutions, large variation in placement requirements is expected.

COLLEGE and WORKPLACE	
Recommended for NAEP 2009 (High Priority)	Feasible for NAEP 2009
National Survey of Postsecondary Education Assessments and Cut Scores	✓

Figure 4-4. Survey Studies for Reading and Mathematics. *This shows the survey study that is recommended and feasible.*

Additional Details for the Feasible 2009 Survey Study

Based on the five to seven exemplar occupations selected from the workplace-related judgmental study above, some information should be collected on job training programs focused on these exemplar occupations.

ADVISORY GROUPS FOR IMPLEMENTATION AND ANALYSIS PHASES OF STUDIES

The Technical Panel recommends convening one or more advisory groups to advise Governing Board staff members regarding the specifics of the design of the studies and analyses. Working with such advisory groups would support the procedural validity of the studies and assist the Board in developing appropriate interpretations of results of the preparedness studies. Such advisory groups should be convened to detail the designs and vet the methodologies of the judgmental and statistical studies; specify appropriate methodologies for the content alignment studies; and recommend analyses of data to maximize for NAEP the information available about student preparedness.

CRITERIA USED TO ASSIGN PRIORITY FOR THE STUDIES

The Panel has urged that its recommendations draw appropriately and in a balanced fashion from the results of the qualitative and quantitative studies. This will help ensure the appropriateness and defensibility of the findings and recommendations, as discussed earlier. This balanced approach will help the Board to utilize the quantitative empirical evidence in ways capable of withstanding the scrutiny that is likely for this type of research initiative. The full set of criteria the Panel used to assign priority to each study included:

- Prospective utility and strength of the results
- Data collection requirements and potential for fulfillment
- Mix of postsecondary targets, i.e., college, military, civilian, or a combination
- Balance of quantitative and qualitative study designs
- Availability of alternative study designs to produce same or similar results
- Whether the study represents a preparatory step for future studies or produces results for direct reporting on the NAEP scale
- When study results could be available

These prioritization criteria, and the special challenges to the study designs, determined the priority labeling for each study presented above. As noted in the discussion of the Technical Panel's Charge, financial cost was not a major factor in the Panel's deliberations, although Panel members were sensitive to the likely costs of some studies. In adherence to the Charge to the Panel, the focus remained on the technical features of the studies. Other types of cost were considered, e.g., time, logistics, and efforts to establish collaboration. ■

Section 5

Additional Recommendations for Preparedness Research and Reporting

When results from NAEP preparedness studies become available and when decisions are being made about NAEP preparedness reporting, the Panel believes that a set of key considerations must be kept in mind to ensure a stronger outcome. Several issues have been noted earlier: using multiple study designs; focusing squarely on reading and mathematics skills; and avoiding representing NAEP reports of preparedness as constituting the sole authoritative definition of preparedness. The following subsections present additional recommendations for reporting preparedness in NAEP Report Cards.

Encourage Parallel Studies by External Organizations

Complementary studies by external agents could possibly address, for example, important aspects of preparedness for the workplace and college that NAEP cannot assess. This would help to distinguish what NAEP can and cannot do—which, in turn, could increase acceptance of NAEP's unique capabilities by focusing on the benefits of results that NAEP can report about preparedness of 12th graders.

Similarly, if Trial Urban District Assessment cities (TUDAs) are willing, and if sufficient existing and available data can be utilized, external funding could support a special study to administer grade 12 NAEP in interested TUDAs. This could provide 1) validity evidence for urban student populations and 2) information about the postsecondary choices of students, such as going immediately into full-time versus part-time postsecondary education, enlisting in the military, and the like.

Be Mindful of the Evolving Context of Preparedness

In recent years, attention to postsecondary preparedness has increased substantially as a policy issue, at the state level and now at the federal level. The July 2008 report from Achieve "Out of Many, One: Toward Rigorous Common Core Standards from the Ground Up" states (p.1):

> *For the first time in the history of American education, educators and policymakers are setting their sights on reaching this goal. . . . Since 2005, states have made rapid progress in raising standards to align with the real-world expectations of employers and postsecondary faculty [members] in the increasingly competitive global market place. To date, 22 states have aligned their high school standards with these real-world goals.*

As further evidence of a dramatic, emerging policy shift, the National Governors Association (NGA), Council of Chief State School Officers (CCSSO), and Achieve, Inc., have joined to provide states a roadmap for benchmarking their K-12 education systems to those of top-performing nations. As of August 2008, 14 states had joined the Algebra II partnership and 12 of these states had already administered the newly developed Algebra II exam to their high school students. Also, as noted earlier in this report, 11 states have volunteered for the grade 12 state NAEP pilot, which will, for the first time, provide detailed state-level NAEP results for 12th graders in reading, mathematics, and science.

State education policy officials are striving to understand the relationship between their high school content standards and assessments and college preparedness. They are confronted by the reality that in almost all states, high school standards and assessments were developed without explicit reference to postsecondary success criteria and that this relationship has not been examined until very recently. Given the substantial investment most states have in their high school standards and assessments, this creates an environment within which state officials may be tempted to resist any information suggesting that a state's adopted high school standards and assessments are not well aligned with college preparedness or to dispel the notion that the two should be related. In either case, introducing information about how well the state's students perform on NAEP that can be related, even in a limited fashion, to college preparedness will likely have serious political ramifications. While this may be positive overall and serve NAEP's purpose as the Nation's Report Card, the potential reactions should nevertheless be borne in mind and anticipated, to the degree it is possible to do so.

Additionally, as mentioned earlier, this report has focused on NAEP's role in the more narrowly defined arena of preparedness. It is worth noting that many policy initiatives are directed toward the broader notion of readiness. The definition of readiness and the means to measure readiness are evolving rapidly in a number of states. Texas, for example, has adopted a formal set of college readiness standards that includes, along with content standards in four subject areas, a set of cross-disciplinary skills that emphasize mastery of key cognitive strategies necessary for learning in postsecondary education. Six states require all students to take a college admissions test, either the ACT or SAT. Placement tests are under scrutiny in a number of states. A few states are beginning to investigate more complex forms of assessment that take into account student performance along a wider range of dimensions than is typically measured on tests such as NAEP. Future state high school testing seems to be trending toward increased use of end-of-course exams that will delve more deeply into subject matter knowledge. Classroom-based assessment is also receiving greater emphasis, with a number of new development projects for large-scale systems of this nature on the horizon. These systems utilize performance tasks embedded within classroom teaching to measure more comprehensively what is actually being taught.

If states begin to assess readiness in broader terms, even if only for formative assessments that augment state tests, NAEP may want to consider whether or how to adapt to these developments. This may require moving beyond preparedness to readiness as the operational framework within which NAEP is analyzed. Assessing an expanded definition of readiness would require additional studies of the type described in this report, the addition of item types in NAEP assessments, or other strategies to capture a wider range of student skills and abilities. Given current measurement technologies and NAEP's policy goals, the path recommended in this report appears to be the most feasible and productive way forward.

All of this activity across the nation is leading to the rapid evolution and specification in greater detail of the notion of readiness itself. As noted previously, readiness has come to encompass multiple dimensions beyond reading and mathematics knowledge and skills. State education policy may begin to require that high schools ensure that all students develop a broader range of learning and self-management strategies, particularly during the final two years of high school.

The Panel recognizes that NAEP preparedness reports will be issued within this dynamic context and believes that state preparedness initiatives and current national and state policy dialogues about preparedness definitions should prompt continued development of NAEP preparedness reference points to be responsive to this changing policy context.

Address Caveats to NAEP's Reporting on 12th Graders' Preparedness

The Technical Panel recommends that each NAEP report include a section that explains the definition of preparedness as used in NAEP, including the limits of that definition and the aspects of preparedness that NAEP is *not* designed to measure. Aspects of preparedness that NAEP is not designed to measure include individual student persistence, time manage-

ment, study skills, and student knowledge of the postsecondary context. In addition, this section of each report should acknowledge the validation research that has been undertaken to specify preparedness more fully, the complex nature of postsecondary education and employment in the United States, and NAEP's limitations in addressing preparedness within this broader context.

Proceed with Reporting Workplace Training Preparedness on NAEP if Supported by the Research

The Panel thought that workplace training preparedness reporting should proceed with the 2009 NAEP reading and mathematics Report Cards, if the Governing Board decides to report preparedness in 2009 based on preparedness research studies. Relative to college preparedness, the number of feasible and proposed studies for workplace training preparedness is lower. Therefore, the Panel recommends that the Governing Board exercise special care and caution in what is reported relative to workplace training preparedness of 12th graders. Little is currently being reported for workplace-related preparedness in the field. Hence, these preparedness research studies and future NAEP preparedness reports could provide a valuable source of information.

Consider a Research Approach to Enable Affirmative Statements About Examinees Scoring Below the Basic Achievement Level

In accordance with the Board's request, the Panel has considered how the studies recommended could enable reporting of the preparedness of students scoring below the *Basic* achievement level on NAEP. The Panel urges a research strategy for collecting information across a range of settings. Those data could show whether there is a range of preparedness levels for a particular postsecondary activity, and whether the lower part of this range overlaps with NAEP scores below the *Basic* achievement level.

It is important to note, however, that the types of occupations that the Board selects as exemplar occupations will greatly determine the likelihood of reference points in the level below *Basic*. If exemplar occupations represent education and experience levels well above the minimum score for the *Basic* achievement level, there is less chance that reference points will be identified below the *Basic* level. Further, the Panel has noted that occupations at the minimum levels of education and experience may not have sufficiently detailed reading and mathematics requirements to support implementation of a judgmental standard-setting study for NAEP.

Evaluate the Feasibility of Using the Newly Identified NAEP Preparedness Reference Points to Report Preparedness

Given the number and variety of preparedness studies, the results will need to be synthesized into a set of coherent recommendations. A panel of professionals from various methodological areas represented by NAEP preparedness research—and, perhaps, some professional representatives of stakeholder groups who have a critical interest in preparedness issues—could probably contribute to this process. Such a panel could provide a critical evaluation of all the study results and could render a professional judgment about the feasibility of using the available NAEP preparedness research findings to report preparedness on NAEP. This would assist the Board by providing valuable input from a variety of viewpoints, and it would support the procedural validity of this research and reporting initiative. ■

Section 6

Future Studies Recommended by the Technical Panel

To build on the foundation set by the NAEP preparedness studies for NAEP 2009 reading and mathematics, and to address the evolving national context of preparedness, the Panel has proposed additional studies for NAEP preparedness research beyond 2009.

Studies to Incorporate Additional Exemplar Occupations

As noted earlier, identification of the original five to seven exemplar occupations should be considered the first phase of this line of NAEP workplace training preparedness research. In the next phases, more exemplar occupations should be added to eventually represent the full population of relevant occupations. If not included in the first round of studies, the Technical Panel recommends that National Automotive Technicians Education Foundation (NATEF) standards be used in a judgmental standard-setting study to produce related reference points on NAEP. The NATEF industry training standards represent an especially rich source of information for the National Institute for Automotive Service Excellence. These standards outline the knowledge, skills, and abilities needed to perform various functions as an automotive service provider.

Benchmarking Studies

As noted earlier in this report, the Technical Panel sees great value in the information that can be provided by benchmarking studies. Further, they have the potential to complement earlier NAEP preparedness research studies by providing reference points related to key postsecondary settings. Due to some of the challenges (enumerated in *Section 2*) in administering these studies, the Panel recommends that they be done after the first set of studies for the 2009 NAEP reading and mathematics assessments. The proposed design for these studies is to administer NAEP as soon as possible after graduation from high school in the postsecondary settings of interest. These settings could be job training programs for the exemplar occupations or colleges in the fall semester with students in entry-level general education courses who graduated from high school the previous spring. The Panel also recommends conducting surveys with postsecondary instructors of the NAEP examinees. The instructors will be asked to use a rating scale to evaluate preparedness of these newly entering students. The survey data can then be examined relative to the NAEP scores of the students. These studies have the potential to augment both college preparedness and workplace training preparedness research.

Studies with Other State Longitudinal Databases

The richness of the longitudinal data available in the state of Florida will provide a valuable set of lessons learned that can be operationalized in later studies to be used for NAEP preparedness research. The Panel recommends research with other states, after the Florida research findings become available for NAEP 2009 reading and mathematics.

Reference Course Studies

The reference course study design is a new methodology currently being employed by the Texas Higher Education Coordinating Board (THECB) to validate the standards it developed for college readiness. The methodology involves the collection of a set of representative syllabi, course materials, and instructor ratings from courses of the same title within the state. These courses are then synthesized into a single course that can serve as the reference when developing assessments or high school materials geared toward

college preparedness. The reference course consists of a detailed syllabus with supporting course materials such as assignments and tests. Its purpose is to demonstrate what college preparedness looks like in practice for that particular course title. The reference course can then be used to judge the relationship of any given assessment to college preparedness relative to that course title. Such a process allows for a more complete and in-depth determination of the alignment between a test and what students would be expected to do in a set of courses within the subject area the test purports to measure.

The reference course design process begins with instructors from a carefully selected cross-section of entry-level courses of the same or similar title (e.g., College Composition, Biology 101). Via a Web-based application, the instructors submit their syllabi and indicate the degree to which college preparedness standards are important for success in their class. Instructor responses are then validated by external reviewers who analyze the course documents (syllabi, assignments, tests) for evidence that the college readiness standards are actually important to success in the class. This process helps to confirm the plausibility of the instructor responses and functions as a cross-check between espoused and actual practices. The results are then given to a team of highly trained content experts who also teach entry-level courses in the subject area. This team then crafts a composite course representing the knowledge and skills identified by instructors. The resulting course is designed to reflect a consensus of what is required in such a course across multiple institutions.

The Panel has discussed the strong potential this newly developed research design has for NAEP preparedness research. It represents an efficient way for NAEP to obtain broad-based input from postsecondary education institutions. ■

End Notes

[1] ***O*NET Zones for NAEP Preparedness Research.***

Job Zone 2:

Overall Experience: Some previous work-related skill, knowledge, or experience may be helpful in these occupations, but usually is not needed. For example, a teller might benefit from experience working directly with the public, but an inexperienced person could still learn to be a teller with little difficulty.

Training and Education: Job Training Employees in these occupations need anywhere from a few months to one year of working with experienced employees. These occupations usually require a high school diploma and may require some vocational training or job-related coursework. In some cases, an associate's or bachelor's degree could be needed.

Job Zone Examples: These occupations often involve using your knowledge and skills to help others. Examples include sheet metal workers, forest fire fighters, customer service representatives, pharmacy technicians, salespersons (retail), and tellers.

Job Zone 3:

Overall Experience: Previous work-related skill, knowledge, or experience is required for these occupations. For example, an electrician must have completed three or four years of apprenticeship or several years of vocational training, and often must have passed a licensing exam, in order to perform the job.

Training and Education: Employees in these occupations usually need one or two years of training involving both on-the-job experience and informal training with experienced workers. Most occupations in this zone require training in vocational schools, related on-the-job experience, or an associate's degree. Some may require a bachelor's degree.

Job Zone Examples: These occupations usually involve using communication and organizational skills to coordinate, supervise, manage, or train others to accomplish goals. Examples include funeral directors, electricians, forest and conservation technicians, legal secretaries, interviewers, and insurance sales agents.

[2] An expert panel was convened in June 2008 to provide focused advice for content alignment studies in the context of NAEP preparedness research. This expert panel also concluded that a modified version of the Webb methodology would be useful to NAEP preparedness research. See *Appendix F* for a full summary of their advice.

[3] One example of how key terms can be interpreted in different ways across educational contexts is the term "analysis." What constitutes "analysis" in an academic performance standard may be interpreted in several ways. Managing this variability as well as variability in the judgmental standard-setting expert panel raters will be important. This is also related to the degree of stringency in their overall judgments will be a challenge that should be addressed in technical work plans for NAEP preparedness studies. Maintaining detailed records of subject matter experts' reasons for their ratings would be a helpful post-meeting check, though front end procedures are also needed.

[4] The NAEP-ACT Content Alignment Study is a preliminary study. A full-scale study may be possible at a later time.

[5] The NAEP-WorkKeys Content Alignment Study is a preliminary study. A full-scale study may be possible at a later time.

References

American Diploma Project (2004). Ready or Not: Creating a High School Diploma That Counts. Washington, DC: Achieve, Inc.

Brown, R. and Conley, D. (2007). Comparing state high school assessments to standards for success in entry-level university courses. *Journal of Educational Assessment*, 12(2), 137–160.

National Center for O*NET Development (2008). Occupational Information Network: O*NET Online. Retrieved September 29, 2008, from O*NET Web site: http://online.onetcenter.org.

Porter, A. and Smithson, J. (2001). CPRE Research Report Series RR-048: Defining, Developing, and Using Curriculum Indicators. Philadelphia, PA: Consortium for Policy Research in Education.

Appendices

Appendix A

LIST OF PANEL MEETINGS AND OBJECTIVES

Meeting 1: June 5-6, 2007
Washington, DC

- Received briefings on the NAEP assessment and recent activities related to grade 12 NAEP.
- Reviewed briefing on the history of linking studies with NAEP.
- Discussed the utility of benchmarking studies for NAEP.
- Developed a preliminary plan for how to identify recommendations and priorities.

Meeting 2: August 13-14, 2007
Denver, CO

- Identified the potential anchors (other assessments used to indicate preparedness) most likely to be feasible for NAEP preparedness studies.
- Identified the advantages and disadvantages of each potential anchor.
- Identified linking studies based on the advantages and disadvantages of potential anchors.
- Evaluated the feasibility of operationalizing a definition of preparedness in the NAEP context.
- Evaluated the feasibility of a "single standard" model of preparedness for postsecondary education and workplace training.

Teleconference 1: September 6, 2007

- Reviewed draft study descriptions to confirm study ideas developed to date.
- Identified organizations, states, and individuals to convene for discussion.

Meeting 3: October 2-3, 2007
Washington, DC

- Examined the NAEP item pool and test specifications.
- Met with representatives from ACT, Inc., the College Board, and the Florida state assessment office.
- Discussed potential study ideas with the invited groups to gain a better understanding of the availability of data and other resources.
- Determined the feasibility and usefulness of these data and resources for the goals of the Governing Board.

Teleconference 2: December 3, 2007

- Reviewed the November 16, 2007 presentations by Michael Kirst to the Committee on Standards, Design, and Methodology and to the Governing Board, and discussed feedback to the Panel from each session.
- Updated participants on study designs under consideration and on additional information collected to assist the Panel in assessing feasibility and potential collaborations.
- Collected Panel input on the draft list of criteria for prioritizing studies.
- Gathered preliminary recommendations from the Panel regarding studies that will impact NAEP operational procedures for 2009.

Meeting 4: February 6-7, 2008
Washington, DC

- Reviewed updates regarding the Panel's study proposals in order to discuss and recommend additional modifications that may be needed.
- Prioritized the Panel's study proposals developed to date.
- Discussed possible recommendations to the Governing Board regarding definitions of preparedness.
- Recommended criteria for selecting target occupations to serve as exemplars of military and civilian workplace preparedness.
- Discussed options for reporting preparedness and achievement levels for grade 12 NAEP.

Meeting 5: April 16-17, 2008
San Francisco, CA

- Reviewed updates regarding the Panel's study proposals in order to discuss and recommend additional modifications that may be needed.
- Discussed course placement issues in higher education related to 12th graders' preparedness.
- Reviewed findings of the preliminary content alignment studies for grade 12 NAEP reading and mathematics in relation to the SAT, and discussed methodological implications for other alignment studies.
- Reviewed draft work statements for the ACT studies.
- Reviewed methodological updates for the reference course study design.
- Reviewed timelines for studies.

Teleconference 3: September 15, 2008

- Reviewed updates on study collaborations and related contracts.
- Deliberated on definitions of preparedness to recommend to the Governing Board.
- Reviewed and revisited earlier discussion points regarding benchmarking studies as discussed in the August 2008 quarterly Board meeting.
- Reached a final determination of how the reference course study design should be incorporated into the Panel's formal recommendations.

Meeting 6: September 17-18, 2008
Washington, DC

- Briefed former U.S. Secretary of Education Margaret Spellings and senior staff members.
- Reviewed a draft of the Panel's final report and outlined next steps.
- Constructed recommendations for future studies.

Appendix B

GUIDING PRINCIPLES FOR PREPAREDNESS

Several principles have shaped the Panel's approach to its work. These principles have been used to help the Panel align its recommendations with the purpose and function of NAEP as well as the Governing Board's goals related to grade 12 NAEP preparedness research.

From the Assessments and their Frameworks:

1. NAEP frameworks for the 2009 assessments in grade 12 reading and mathematics have been revised to allow reporting of preparedness for postsecondary endeavors in college and the workplace.
2. NAEP does not report individual student scores.
3. The identity of students and schools sampled in NAEP is confidential.

From the Ad Hoc Committee on Planning for NAEP 12th Grade Assessments in 2009:

1. If found to be technically, operationally, and economically feasible, NAEP reports should include statements about 12th grade student preparedness, beginning with the reading and mathematics assessments to be conducted in 2009.
2. The Governing Board should develop and implement a plan for setting a final policy definition of 12th grade student preparedness, decide on the statements about "preparedness" to include in NAEP reports, and conduct associated research and validity studies.
3. The term "12th grade student preparedness" should be limited to postsecondary education and postsecondary training for occupations (including occupations in the military).
4. The definition of 12th grade student preparedness should focus on academic qualification without remediation for postsecondary education and postsecondary training for occupations; it should not include non-academic personal attributes.
   Academic qualification: This recommendation emphasizes "qualification to enter" rather than "success in" or "completion of" postsecondary education and training.
   Without remediation: A key factor in defining 12th grade student preparedness.
5. The reporting of 12th grade student preparedness in NAEP should be kept as simple as possible to promote public understanding, consistent with available validity evidence.
6. The reporting of 12th grade student preparedness should be done in conjunction with the Governing Board's achievement levels—*Basic*, *Proficient*, and *Advanced*—rather than by setting separate preparedness performance standards. The degree of preparedness of students whose achievement is in the range below *Basic* should be fully reported as well.

From the Charge to the Technical Panel:

1. It is important to find technically appropriate, accurate, and valid ways to use grade 12 NAEP as a measure of preparedness and to report such results.

Appendix C

CONSIDERATIONS FOR TERMS RELATED TO PREPAREDNESS

When NAEP preparedness research studies are completed, the Governing Board will be able to formulate preparedness statements validated by the research findings. The Technical Panel has recommended in its final report a statement of preparedness referring to NAEP as an indicator of academic preparedness for college and workplace training. Research may reveal a statement that will fit more coherently with the research findings. Given the possible statements and terms that can be used in statements of preparedness for NAEP Report Cards, the Panel has identified some key terms that need to be considered when developing statements of 12th graders' preparedness.

Remedial has some negative connotations that should be avoided. Further, remediation is relative to the context, such as the selectivity of the postsecondary institution and the structure of the state's postsecondary education system. For example, transfer-credit courses at two-year institutions (applicable to a four-year undergraduate degree program) are by definition non-remedial courses for the most part, but there are non-remedial courses at two-year institutions that do not count as transfer credits. ***Developmental*** is a term often used as a synonym for *remedial*, but it may be too general conceptually given its various applications in postsecondary education and in education more broadly. Whether *remedial* or *developmental* course placement is referenced in a prospective statement regarding preparedness, both terms will lead to a statement in the negative (defining what preparedness is not), while a statement in the affirmative (defining preparedness as what students can do) would be preferable for NAEP. For these reasons, the Technical Panel recommends that the statement(s) of preparedness in NAEP report cards not include the word *remedial* or *developmental*.

Focusing on ***credit-bearing*** coursework seems an alternative, but it too presents challenges. Across institutions, the same course content may be offered as a for-credit course or a non-credit course. There could be institution-level or state-level distinctions in what constitutes a remedial course versus a course that counts toward a degree. Further, as intimated in the above discussion of the term *remedial*, courses that count for an associate's degree may not count for a bachelor's degree. Therefore, it may be desirable for a statement of preparedness to specify at the outset whether *credit-bearing* is defined in the context of a two-year or four-year degree program.

The term ***transfer credit*** is one way to qualify the term *credit-bearing* for the purpose of NAEP preparedness reporting. Courses offering *transfer credits* enable students in two-year institutions (or community colleges) to earn credits that count toward a bachelor's degree at a four-year institution. Transfer credits would be applicable if, for example, this student decides to seek a bachelor's degree after attending a two-year institution. Courses that only count for terminal associate's degrees are excluded from this set of credit-bearing courses.

The term ***freshman*** is less straightforward than one might imagine, because any first-time enrollee is considered a freshman, regardless of how many years have elapsed since his or her K-12 education. The average age of community college students is higher than that of students in four-year institutions, and this higher average age is also represented among freshmen. Therefore, an age range may be needed, e.g., ages 17 to 19, to avoid this quandary.

Alternatively, "*recent* high school graduates" could be helpful phrasing, or "high school completers enrolling immediately after graduation."

For related reasons, ***first-year students*** also need to be specified carefully if they are included in a preparedness statement, because students may elect to take their courses on a part-time basis, thereby making them a "first-year" student for longer than one school year.

General education courses refer to courses in a variety of subject areas covering physical sciences (inclusive of mathematics), social sciences, and the humanities. General education course titles and the content within courses of similar titles vary greatly, and the Panel recognizes that *general education courses* are diverse across institutions. These courses are typically "introductory" courses in the core areas. More specificity is needed for NAEP. Therefore, for the NAEP mathematics assessment, the focus could be *general education* courses in mathematics, whereas for the NAEP reading assessment the focus could be the reading knowledge and skills for a *general education* course such as psychology. Many colleges and universities specify the number and distribution of credit hours in these *general education* courses required for a bachelor's degree. *General education* courses typically satisfy the distributional requirements that establish a breadth of academic content. One benefit of using *general education* is that it does not refer to courses on the pathway to a particular college major that would likely require a level of competency well beyond the general entry-level course. This is also desirable for the context of NAEP, because the Panel has understood the Governing Board's conceptualization of preparedness to mean "eligibility to enter" rather than "prepared to succeed."

The Panel has noted that preparedness statements could represent a composite of an institution-centered approach focused on types of institutions and a student-centered approach focused on paths from the student perspective.

Appendix D

TIMELINES FOR COMPLETING THE STUDIES

Key milestones in the timelines for the studies are listed below. While there are many additional milestones, this timeline is meant to provide an overview based on the Panel's deliberations at its last meeting in September 2008. As item-level data become available from the 2009 NAEP reading and mathematics assessments, the content alignment studies and preliminary work for the judgmental studies with subject matter experts can begin.

Two major milestones for the Board will be making reporting decisions based on the studies' results in May 2010 and planning for the report's release in approximately October 2010. The Panel has suggested it may be advisable to delay reporting to incorporate all of the findings of the NAEP preparedness studies.

Content Alignment Studies
December 2008:
- Begin 9 to 12 month period of implementing studies

Statistical Relationship Studies
January - March 2009:
- Administer 2009 NAEP in reading and mathematics
- Implement strategies to collect data for matching NAEP national or state performance data to those of other assessments for studies to establish statistical relationships

October 2009:
- Begin statistical analyses with operational NAEP data

Judgmental Standard-Setting Studies with Subject Matter Expert Panels
October 2008:
- Begin studies to develop preparedness standards in areas for which preparedness indicators (eligibility criteria or performance standards) are not available

October 2009:
- Initiate pilot studies of the judgmental standard-setting studies
- Begin full-scale judgmental standard-setting studies with operational data

Survey of Postsecondary Education and Job Training Institutions
October 2008 – September 2009:
- Conduct the survey study and report results for use in other preparedness studies

Appendix E

RECOMMENDED STUDIES FOR 2009 NAEP 12TH GRADE PREPAREDNESS REPORTING

(A) Content Alignment Studies for NAEP and Assessments of Postsecondary Preparedness

- Comparison with college admissions and placement examinations (ACCUPLACER, ACT, ASSET, COMPASS, SAT)
- Comparison with workplace eligibility and placement examinations (WorkKeys and ASVAB)

(B) Statistical Relationship Studies for NAEP and Assessments of Postsecondary Preparedness

- Linking national NAEP scores with preparedness indicator scores from other assessments
- Linking 12th grade NAEP performance with longitudinal databases (score data for college admission and course placement; transcript data; and workplace data)

(C) Judgmental Studies to Set NAEP Cut Scores for Workplace Preparedness (Military and Civilian)

- Identification of five to seven target occupations across various sectors
- Identification and development of eligibility criteria for each target occupation's job training programs
- Setting NAEP reading and mathematics job training program cut scores

(D) Judgmental Studies to Set NAEP Cut Scores for College Preparedness

- Setting NAEP reading and mathematics college preparedness cut scores using:
  - ACT College Readiness Standards
  - College Board Standards for College Success
  - Standards developed by subject matter experts specializing in college course placement

(E) National Survey of College Course Placement Assessments and Cut Scores

Appendix F

SUMMARY REPORT ON RECOMMENDATIONS OF EXPERT CONTENT ALIGNMENT GROUP

June 26, 2008

The Technical Panel on 12th Grade Preparedness Research recommends content alignment studies as central to establishing validity evidence for reporting preparedness of 12th graders on NAEP. The Panel recommends content alignment for each assessment to be used in relation to NAEP to guide the interpretation of results and to bring greater credibility and validity to preparedness reporting. The Panel sees these studies as an important first step on the trail of evidence to establish validity of measures and indicators.

The Technical Panel views the studies as prerequisites to other studies, particularly those statistically relating NAEP to other assessments. The Panel has reviewed various methodologies commonly used in content alignment studies and recommended the following:

1. Use the Webb methodology, with appropriate modifications for the test-to-test context.
2. Standardize the alignment studies from one assessment to another.
3. Secure the services of a third party to conduct the studies to avoid the potential for bias or the appearance thereof.
4. Implement bidirectional alignment studies: alignment of NAEP to the other assessment and alignment of the other assessment to NAEP.

These recommendations have been made on the basis of the Panel members' collective knowledge and expertise regarding sound research design, and not on the basis of specific expertise in content alignment research. Given the importance of the studies, the Panel suggested that a small group of experts on content alignment be convened to advise staff members on several issues related to the design of the studies.

The five persons convened on June 26 represent expertise in content alignment, standard setting, equating, and psychometrics. Members and their affiliations are:

Robert Brennan, Director
Center for Advanced Studies in Measurement and Assessment
The University of Iowa

Mary J. Pitoniak, NAEP Associate Director
Research and Development
ETS

Marianne Perie, Senior Associate
National Center for the Improvement of Educational Assessment

Stanley Rabinowitz, Director
Assessment and Standards Development Services
WestEd

Norman J. Webb, Senior Research Scientist
Wisconsin Center for Education Research and The National Institute for Science Education

The recommendations of the Expert Content Alignment Group (hereafter referred to as "alignment experts") are presented here, along with key points of their discussion. Their advice generally confirms that of the Technical Panel. Based on the extensive field experience of some alignment panel members, the criteria for some alignment study features were made more explicit and the necessary relationships across features and studies made clearer.

1. Role of Content Alignment Studies

The alignment experts discussed the proposed role of content alignment studies for reporting preparedness on NAEP. Content alignment studies were judged to be necessary to provide the "qualifications to circumscribe the statistical results." The alignment studies will indicate what can and cannot be said about the statistical relationships between two assessments. Alignment studies were characterized as typically showing the upper limit of statistical relationships, although statistical relationships may exist without content alignment.

2. Degree of Alignment

Content alignment studies for NAEP should be considered as showing the level or degree of alignment. States conduct alignment studies to determine if an assessment is aligned with their curriculum standards, and the results are "aligned" or "not aligned." Alignment with NAEP should be evaluated as continuous, not dichotomous.

3. Bidirectional Alignment Studies

The alignment studies should be bidirectional in order to maximize information for reporting preparedness indicators on NAEP derived through statistical predictions or other statistical linking methods (excluding equating). The alignment experts discussed this point extensively because of the resource and cost implications of this feature. Both the informational materials needed and the amount of time needed for the alignment study present greater resource demands. The information on content overlap and coverage provided by a bidirectional design allows evaluation of the relative weights of the key components of each assessment, and that information seems essential. It would be possible to evaluate the alignment of the other assessments to NAEP through less direct means if highly detailed materials—specifications for item development, forms construction, and so forth—can be obtained from the testing companies.

4. Alignment Study Panel Facilitation and Composition

The alignment experts concurred with the Technical Panel that the alignment studies should be conducted by a third party and not the test development company staff. The group discussed the composition of the alignment study panels extensively. They recommended that each alignment panel include approximately eight members with approximately equal representation of both content alignment and subject matter expertise. It is possible to have only subject matter experts on the panels, but more time would be needed for training. The practitioners in the group reported that this mix of expertise works especially well. They noted that each individual in the cadre of trained content alignment personnel also represents knowledge and skills training in a content area.

5. Alignment Study Materials

Materials for the studies and the ability to represent the content domain of each assessment under consideration were discussed at length. The alignment experts noted that the larger the scale or "grain size" of the analysis, the higher the probability of finding alignment between assessments. While item-to-item alignment was not recommended, the need to include test items in the alignment studies was emphasized. The alignment experts specified materials to be provided to both the alignment facilitation staff members and to the alignment panel members.

6. Standardization Across Studies

Standardization across studies should be maximized to the extent practicable. More than one alignment contractor is needed to conduct the volume of studies planned for the NAEP preparedness research. Standardization across facilitation is critical, and scripts for facilitators were recommended.

The alignment experts anticipated considerable variability in the materials available for each assessment of interest. Standardization of methodology should be maintained, but adjustments will be needed to reflect the differences in materials available for alignment studies with specific assessments.

7. Viability of Non-alignment as Study Outcome

Panelists should be made aware of the purpose of NAEP and how this purpose differs from that of the other assessment in each alignment study. Instructions to panelists should clarify that the assessments are known to be different and that complete alignment would be anomalous. Panelists should understand the value of delineating non-alignment and how that information will be used in interpreting and reporting results. The instructions must be documented and panelists' understandings of instructions should be documented through evaluation questionnaires.

8. Reliability Issues

Reliability issues were discussed in some detail. Inter-rater and intra-rater consistency in alignment classifications are typically collected and reported as reliability evidence in alignment studies. One of the alignment experts advocated strongly for replicate panel studies to provide the confirmatory evidence of alignment. The potential for implementing a replicate panel alignment study design was also discussed. The alignment experts agreed that replication studies are desirable, but the costs may be prohibitive. They suggested the possibility of replication with one subject for each assessment as a compromise.

Note: The Technical Panel has recommended a pilot alignment study for each type of assessment. The pilot study and operational study may be considered as replicate studies, although it will be necessary to document and account for differences in procedures between the two implementations.

9. Relative Value of Judgmental and Statistical Studies

During the wrap-up, the alignment experts discussed the array of research studies recommended for reporting preparedness on NAEP. The Technical Panel has placed highest priority on the studies to establish statistical relationships, and the alignment experts clearly agreed with that position. A cautionary message was delivered regarding the value of judgmental studies without statistical studies to accompany them. The value of preparedness reference points derived through judgmental studies alone would perhaps not warrant the expenditure of resources. If collaborative agreements and partnerships cannot be forged to provide data needed to produce statistical predictions in one area of preparedness for reporting in 2009, then the Governing Board may find it advisable to delay research studies for that particular area.

10. Sequencing Studies

Finally, the alignment experts considered the advisability of conducting statistical studies first to determine the need for content alignment studies. The requirements of statistical analyses are more clearly known and established than those of content alignment. Thus, the existence of a statistical relationship would indicate the need for the content alignment studies to circumscribe the interpretation of results. The timelines associated with the availability of data, however, made this suggestion impractical. Content alignment studies can be conducted with 2008 field trial data during the year prior to the availability of data for statistical analyses, whereas statistical studies cannot begin before the 2009 operational data are available.

National Assessment Governing Board
800 North Capitol Street, N.W., Suite 825
Washington, DC 20002–4233
www.nagb.org