

National Assessment Governing Board Committee on Standards, Design and Methodology

February 28, 2014
10:00 am – 12:15 pm

AGENDA

10:00 – 10:05 am	Introductions and Review of Agenda <i>Lou Fabrizio, COSDAM Chair</i>	
10:05 – 10:40 am	Plans for two-stage adaptive testing in NAEP Mathematics and Science <i>Andreas Oranje, Educational Testing Service</i>	Attachment A
10:40 – 11:15 am	Results from NAEP-PIRLS linking study <i>Gary Phillips, American Institutes for Research</i>	Attachment B
11:15 – 11:20 am	Break	
11:20 am – 12:05 pm	[Closed Session] Participation and Engagement on 2013 NAEP Grade 12 Mathematics and Reading Assessments <i>Samantha Burg, NCES</i>	Attachment C
12:05 – 12:10 pm	Committee Questions on Information Items (see below) <i>Sharyn Rosenberg, NAGB Staff</i>	
12:10 – 12:15 pm	Other Issues and Questions <i>COSDAM Members</i>	
	Information Items: <ul style="list-style-type: none"> • NAEP 12th Grade Academic Preparedness Research: Phase 2 Research Updates • Update on TEL Achievement Levels Setting Procurement • Update on Implementation of SD/ELL Exclusion Policy • Update on Evaluation of NAEP Achievement Levels Procurement 	Attachment D Attachment E Attachment F Attachment G



Plans for Adaptive Testing in Mathematics and Science

There is little doubt that the National Assessment of Educational Progress (NAEP) has been subject to dramatic changes over the past two decades. The introduction of No Child Left Behind and the advent of Trial Urban District Assessments (TUDAs) have shifted the program to a source for increasingly divergent types of information. In this expanded role, the NAEP program serves as validation for state assessment results, reports academic performance in urban settings, describes trends in race/ethnicity and gender gaps, connects to post-secondary preparedness research, and provides detailed assessment content (items, scoring rubrics, and item level results). Implicitly, the expectations about what the assessment can tell about students have changed commensurately, mostly characterized by the interest in groups of students in jurisdictions that display performance at significantly different levels. For example, there is an interest in both white students in Massachusetts, who score more than half a standard deviation *above* the nation's average, and black students in Detroit, who score well over a full standard deviation *below* the nation's average. These changing interests and requirements are in addition to more traditional purposes for the NAEP program, including the comparisons of the knowledge and skills of U.S. students over time (i.e., reporting of trend results).

As the program is transitioning to technology-based testing in the next several years, new operational and psychometric requirements *and* opportunities surface, including the ability to adopt some type of adaptive testing in order to meet the aforementioned expanded reporting expectations. A comprehensive Mathematics Computer-Based Study (MCBS) was conducted in 2011 to assess the merits of two-stage adaptive testing within the NAEP group-score assessment context. This study primarily focused on two aspects: what gains in terms of measurement properties can be achieved (e.g., smaller standard errors for low or high performing groups) and what improvements can be made in terms of the student test taking experience (e.g., providing challenging yet achievable tasks to improve data quality and completion rates). The results, presented to the Committee on Standards, Design, and Methodology (COSDAM) at previous meetings, were highly encouraging and resulted in two concrete activities: (a) significant planning towards adaptive testing for Mathematics and Science as part of technology-based assessment transition plans, and (b) conduct of several (quasi-simulation) studies to determine optimal adaptive testing designs for NAEP.

Across these activities, a number of questions have been and, in some cases, have yet to be answered. At this COSDAM session, we will present plans for adaptive testing in Mathematics

and Science within the larger context of NAEP technology-based administration. Some key questions that we will review, and answer when possible, are:

- How will adaptive testing be implemented in terms of:
 - The type of adaptive testing
 - Design considerations (e.g., test length, decision points)
 - Framework coverage considerations (all topics covered for every jurisdiction)
 - Trend reporting both into the future and to past assessments
 - Pre- or pilot-testing and new item development
- What is the timeline for implementing adaptive testing?



Linking the 2011 National Assessment of Educational Progress (NAEP) to the 2011 Progress in International Reading Literacy Study (PIRLS) Study

Gary Phillips of the American Institutes for Research (AIR) will present his paper that describes a statistical linking between the 2011 National Assessment of Educational Progress (NAEP) in grade 4 reading and the 2011 Progress in International Reading Literacy Study (PIRLS) in grade 4 reading. The primary purpose of the linking study is to obtain a statistical comparison between NAEP (a national assessment) and PIRLS (an international assessment). By expressing both assessments in the same metric, the linking study can provide international benchmarks for the NAEP achievement levels. Dr. Phillips will compare the NAEP achievement levels with the PIRLS international benchmarks and he will discuss the validity of the NAEP-PIRLS linking results.



Participation and Engagement of Twelfth Grade Students

The quality of scores obtained from NAEP depends in part on sufficient student engagement with the task of answering the assessment questions and on having a very large proportion of the sampled schools and students participate in the assessment. The credibility of 12th grade NAEP results can be impaired not only by insufficient participation and engagement, but also by public skepticism about student engagement. In the November 2010 meeting of the Governing Board, Andrew Kolstad presented several types of evidence related to student participation and engagement. In a series of presentations to the Assessment Development Committee (ADC) during 2010, Dave Freund of Educational Testing Service presented information about student non-response at each grade. In this update for the Committee on Standards, Design, and Methodology (COSDAM), a short review of previous studies as well as new evidence from the 2013 assessment relating to student participation and engagement will be presented.

This presentation will display trends in school and state participation. In addition, brief reviews of private school and student participation and of student participation in the grade 12 pilot states will be presented.

The presentation will also include various measures of student engagement. One measure is a self-report, derived from a question asking the participating students how hard they tried on this test compared to other tests. Results will be compared with performance and will include an analysis of trends in effort on NAEP. Another measure is behavioral—the proportion of test questions left blank. This presentation will display trends in the proportion of uncompleted test questions on NAEP.

NAEP 12th Grade Academic Preparedness Research

Phase 1 Research

The first phase of the Governing Board's research on academic preparedness is now complete; results from more than 30 studies are available at: <http://www.nagb.org/what-we-do/preparedness-research.html>. During the August 2013 meeting, the Board voted on a motion to use the phase 1 research on academic preparedness for college in the reporting of the 2013 grade 12 national results for reading and mathematics. The approved motion and supporting validity argument also appear on the aforementioned website.

Phase 2 Research

The second phase of the Governing Board's research on academic preparedness currently consists of the following studies that are planned or underway:

Study name	Sample	February 2014 Update
Statistical linking of NAEP and ACT	National; FL, IL, MA, MI, TN	Page 7
Longitudinal statistical relationships: Grade 12 NAEP	FL, IL, MA, MI, TN	Page 8
Statistical linking of NAEP and EXPLORE	KY, NC, TN	Page 9
Longitudinal statistical relationships: Grade 8 NAEP	KY, NC, TN	Page 10
Content alignment of NAEP and COMPASS		Page 11
Content alignment of NAEP and EXPLORE		Page 12
College Course Content Analysis		Pages 14-17
Evaluating Reading and Mathematics Frameworks and Item Pools as Measures of Academic Preparedness for College and Job Training (Research with Frameworks)		Pages 18-21

Brief overviews and informational updates are provided for each study.

National and State Statistical Linking Studies with the ACT

The Governing Board is planning to partner with ACT, Inc. to conduct a statistical linking study at the national level between NAEP and the ACT in Reading and Mathematics. Through a procedure that protects student confidentiality, the ACT records of 12th grade NAEP test takers in 2013 will be matched, and through this match, the linking will be performed. A similar study at the national level was performed with the SAT in 2009. There will not be a statistical linking study performed for NAEP and the SAT in 2013.

In addition, the state-level studies, begun in 2009 with Florida, will be expanded with 2013 NAEP. Again using a procedure that protects student confidentiality, ACT scores of NAEP 12th grade test takers in the state samples in partner states will be linked to NAEP scores. We are in the planning stages with five states to be partners in these studies at grade 12: Florida, Illinois, Massachusetts, Michigan, and Tennessee. In three of these states (IL, MI, TN), the ACT is administered to all students state-wide, regardless of students' intentions for postsecondary activities.

Draft Research Questions for National and State Statistical Linking Studies with the ACT:

1. What are the correlations between the grade 12 NAEP and ACT student score distributions in Reading and Math?
2. What scores on the grade 12 NAEP Reading and Math scales correspond to the ACT college readiness benchmarks? (concordance and/or projection)
3. What are the average grade 12 NAEP Reading and Math scores and interquartile ranges (IQR) for students below, at, and at or above the ACT college readiness benchmarks?
4. Do the results differ by race/ethnicity or gender?

February 2014 Update: Data sharing agreements with the ACT and the five partner states at grade 12 are in development.

Longitudinal Statistical Relationships: Grade 12 NAEP

In addition to the linking of ACT scores to NAEP 12th grade test scores in partner states, the postsecondary activities of NAEP 12th grade test takers will be followed for up to six years using the state longitudinal databases in Florida, Illinois, Massachusetts, Michigan, and Tennessee. These studies will examine the relationship between 12th grade NAEP scores and scores on placement tests, placement into remedial versus credit-bearing courses, GPA, and persistence.

Draft Research Questions for Longitudinal Statistical Relationships, Grade 12 NAEP:

1. What is the relationship between grade 12 NAEP Reading and Math scores and grade 8 state test scores?
2. What are the average grade 12 NAEP Reading and Math scores and interquartile ranges (IQR) for students with placement in remedial and non-remedial courses?
3. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students with a first-year GPA of B- or above?
4. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who remain in college after each year?
5. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who graduate from college within 6 years?

February 2014 Update: Data sharing agreements with the five partner states at grade 12 are in development.

State Statistical Linking Studies with EXPLORE

In 2013, linking studies between 8th grade NAEP in Reading and Mathematics and 8th grade EXPLORE, a test developed by ACT, Inc. that is linked to performance on the ACT, are planned with partners in three states: Kentucky, North Carolina, and Tennessee. In all three of these states, EXPLORE is administered to all students state-wide during grade 8.

Draft Research Questions for State Statistical Linking Studies with EXPLORE:

1. What are the correlations between the grade 8 NAEP and EXPLORE scores in Reading and Math?
2. What scores on the grade 8 NAEP Reading and Math scales correspond to the EXPLORE college readiness benchmarks (concordance and/or projection)?
3. What are the average grade 8 NAEP Reading and Math scores and the interquartile ranges (IQR) for students below, at, and at or above the EXPLORE college readiness benchmarks?

February 2014 Update: The data sharing agreement for North Carolina has been signed; data file preparation is underway. Data sharing agreements with Tennessee and Kentucky are in development.

Longitudinal Statistical Relationships: Grade 8 NAEP

In 2013, the Governing Board will also expand the state-level studies by partnering with a few states at grade 8. Again using a procedure that protects student confidentiality, secondary and postsecondary data for NAEP 8th grade test takers in the state samples in partner states will be linked to NAEP scores. These studies will examine the relationship between 8th grade NAEP scores and scores on state tests, future ACT scores, placement into remedial versus credit-bearing courses, and first-year college GPA.

Three states will be partners in these studies at grade 8: Kentucky, North Carolina, and Tennessee.

Draft Research Questions for Longitudinal Statistical Relationships, Grade 8 NAEP:

1. What is the relationship between NAEP Reading and Math scores at grade 8 and state test scores at grade 4?
2. What are the average NAEP Reading and Math scores and the interquartile ranges (IQR) at grade 8 for students below the ACT benchmarks at grade 11/12? At or above the ACT benchmarks?
3. What are the average NAEP Reading and Math scores and the interquartile ranges (IQR) at grade 8 for students who are placed in remedial and non-remedial courses in college?
4. What are the average NAEP Reading and Math scores (and the IQR) at grade 8 for students who obtain a first-year college GPA of B- or above?
5. What is the relationship between grade 8 NAEP Reading and Math scores and grade 12 NAEP Reading and Math scores? (contingent on feasibility of sampling the same students in TN, NC, and KY)

February 2014 Update: The data sharing agreement for North Carolina has been signed; data file preparation is underway. Data sharing agreements with Tennessee and Kentucky are in development.

Content Alignment Study of Grade 12 NAEP Reading and Mathematics and COMPASS

Content alignment studies are a foundation for the trail of evidence needed for establishing the validity of preparedness reporting, and are, therefore, considered a high priority in the Governing Board's Program of Preparedness Research. The alignment studies will inform the interpretations of preparedness research findings from statistical relationship studies and help to shape the statements that can be made about preparedness. Content alignment studies were recommended to evaluate the extent to which NAEP content overlaps with that of the other assessments to be used as indicators of preparedness in the research.

We plan to conduct an alignment study of grade 12 NAEP Reading and Mathematics and ACT COMPASS.

February 2014 Update: Details of our agreement with ACT are still being worked out. We intend to have an independent third party conduct the study, and we have budgeted for the procurement to be awarded during fiscal year 2014.

Content Alignment Study of Grade 8 NAEP Reading and Mathematics and EXPLORE

Content alignment studies are a foundation for the trail of evidence needed for establishing the validity of preparedness reporting, and are, therefore, considered a high priority in the Governing Board's Program of Preparedness Research. The alignment studies will inform the interpretations of preparedness research findings from statistical relationship studies and help to shape the statements that can be made about preparedness. Content alignment studies were recommended to evaluate the extent to which NAEP content overlaps with that of the other assessments to be used as indicators of preparedness in the research.

We plan to conduct an alignment study of grade 8 NAEP Reading and Mathematics and ACT EXPLORE. Results from this content alignment study will be particularly important for interpreting the findings from the NAEP-EXPLORE statistical linking studies.

February 2014 Update: Details of our agreement with ACT are still being worked out. We intend to have an independent third party conduct the study, and we have budgeted for the procurement to be awarded during fiscal year 2014.

OVERVIEW OF REFERENCED ASSESSMENTS

For additional background information, the following list presents a brief description of the assessments referenced in the phase 2 academic preparedness research studies. In each case, only the mathematics and reading portions of the assessments are the targets for analysis, although analyses with the composite scores may be conducted.

- ACT – The ACT assessment is a college admissions test used by colleges and universities to determine the level of knowledge and skills in applicant pools, including Reading, English, and Mathematics tests. ACT has *College Readiness Standards* that connect reading or mathematics knowledge and skills and probabilities of a college course grade of “C” or higher (75%) or “B” or higher (50%) with particular score ranges on the ACT assessment.
- ACT EXPLORE – ACT EXPLORE assesses academic progress of eighth and ninth grade students. It is a component of the ACT College and Career Readiness System and includes assessments of English, Mathematics, Reading, and Science. ACT EXPLORE has *College Readiness Standards* that connect reading or mathematics knowledge and skills and probabilities of a college course grade of “C” or higher (75%) or “B” or higher (50%) by the time students graduate high school with particular score ranges on the EXPLORE assessment.
- COMPASS – ACT Compass is a computer-adaptive college placement test. It is produced by ACT and includes assessments of Reading, Writing Skills, Writing Essay, Math, and English as a Second Language.
- SAT – The SAT reasoning test is a college admissions test produced by the College Board. It is used by colleges and universities to evaluate the knowledge and skills of applicant pools in critical reading, mathematics, and writing. The SAT has calculated preparedness benchmarks are defined as the SAT scores corresponding to a 65% probability of earning a first-year college grade-point average of 2.67 (B-) or better.

COLLEGE COURSE CONTENT ANALYSIS

The College Course Content Analysis (CCCA) study is one of a series of studies contributing to the National Assessment of Educational Progress (NAEP) Program of 12th Grade Preparedness Research conducted by the National Assessment Governing Board (NAGB). The purpose of the CCCA study is to identify a comprehensive list of the reading and mathematics knowledge, skills, and abilities (KSAs) that are pre-requisite to entry-level college mathematics courses and courses that require college level reading based on information from a representative sample of U.S. colleges. The Educational Policy Improvement Center (EPIC) is the contractor working for the Board to conduct this study.

Another goal of the CCCA study is to extend the work of the two previous preparedness studies—the Judgmental Standards Setting (JSS)¹ study, implemented in 2011 and the Job Training Program Curriculum (JTPC) study, implemented in 2012. The CCCA study is designed so the results can be compared to the JSS and JTPC studies, reporting on how this new information confirms or extends interpretations of those earlier studies. The design of the CCCA study is based on the JTPC study but with modifications based on the lessons learned.

The CCCA study addresses four core research questions:

1. What are the prerequisite KSAs in reading and mathematics to qualify for entry-level, credit-bearing courses that satisfy general education requirements?
2. How do these prerequisite KSAs compare with the 2009 and 2013 NAEP reading and mathematics frameworks and item pools?
3. How do these prerequisite KSAs compare with previous NAEP preparedness research (i.e., the descriptions of minimal academic preparedness requirements produced in the JSS research)?
4. How can these prerequisites inform future NAEP preparedness research?

METHODOLOGY

The Design Document for the CCCA study is complete. It provides guidance for the study by describing:

- Criteria for collecting courses and artifacts;
- A sampling plan to comprise a representative sample of institutions;

¹ National Assessment Governing Board. (2010). *Work Statement for Judgmental Standard Setting Workshops for the 2009 Grade 12 Reading and Mathematics National Assessment of Educational Progress to Reference Academic Preparedness for College Course Placement*. (Higher Education Solicitation number ED-R-10-0005).

- Review and rating processes, including a training plan and process for ensuring reviewer effectiveness and consistency; and
- The process for ensuring reliability across reviewers providing artifact analysis.

This study comprises three primary phases:

1. Identification and collection of course artifacts,
2. Review of course artifacts by Review Teams, and
3. Analysis and reporting.

February 2014 Update: The Design Document is complete.

OVERVIEW OF ACTIVITIES BY PHASE

Phase 1: Identification and collection of course artifacts

In the CCCA study, a *course artifact* is defined as a syllabus, a non-textbook based assignment or assessment, and textbook excerpt. In mathematics, there are some instances where the only specifically identified assignments were listed in the syllabus and were from the textbook. In those cases, a textbook based assignment or assessment was allowed. The CCCA sample of artifacts is derived from extant artifacts and combined with newly gathered course artifacts. Extant artifacts contributing to the CCCA sample were extracted from EPIC's repository of artifacts compiled during previous research on entry-level curricula at postsecondary educational institutions. Project staff solicited new course artifacts as needed to create a complete and nationally representative sample.

EPIC identified a set of inclusion criteria that courses must meet to be included in the CCCA study as well as a set of institutional characteristics of which the final CCCA Artifact Bank must be representative. The final CCCA Artifact Bank comprises a set of courses and artifacts that are to be used as the basis for the content reviews to be conducted by mathematics and reading content review teams in the second phase of the study.

Phase 1 preparatory work also included the convening of NAEP advisory panels, for reading and mathematics respectively, to obtain content-based guidance and recommendations. In these meetings, preliminary coding schemas, training materials and decision rules were reviewed. NAEP advisors also reviewed all of the course packets to be used in validation data analyses, training sessions, and determining sufficient reviewer competence (qualifying). Guidance from these NAEP advisory panels was integrated into the implementation of the study.

February 2014 Update: Identification and collection of course artifacts have been completed.

Phase 2: Review of course artifacts by Review Teams

In Phase 2, content reviewers were recruited and training materials were developed in preparation for reviewing course artifacts and conducting content reviews. Content reviewers were first trained to review the course packets from a “holistic” perspective and identify prerequisite mathematics and reading KSAs. In the second independent review training, the NAEP frameworks for grade 12 reading and mathematics were used as a basis for coding the packets. All additional KSAs beyond the NAEP frameworks were documented and included in all successive reviews, comparisons and data analyses. The overarching goal of the CCCA study is to identify all prerequisite KSAs, not just those KSAs associated with the NAEP frameworks.

Subsets of the course artifact packets were set aside to serve as training packets and qualifying packets. These packets were annotated by the NAEP advisory panel members for use as exemplars of expert coding. After the holistic reviews, content reviewers were trained with respect to the NAEP frameworks, and as part of the training process, the reviewers coded the training packets with respect to the NAEP frameworks in small groups. Then, the reviewers coded qualifying packets independently. EPIC project staff then compared and scored this coding with respect to the exemplars provided by the NAEP advisory panel. If a reviewer scored below a certain threshold, retraining was provided. Reviewers who received a second low score were not invited to participate in the study. Qualified reviewers proceeded to the next stage: coding 28 course artifact packets independently. Group review meetings were then held to discuss discrepancies identified in independent reviews.

NAEP experts attended the group review meetings as on-site assistance, answering questions about the NAEP framework as they arise. Validity checks were also embedded in the group review process. Validity packets were annotated by the NAEP experts at the advisory panel meetings to be used as reference coding. Those packets were reviewed by all content reviewers without the knowledge that the packets were for validity purposes. This provided the opportunity for evaluating the reliability of the review team coding. The percent agreement between the four review teams’ group consensus coding on the validation packets and the NAEP reference coding as reliability evidence was calculated within each course title and across course titles.

In summary, the CCCA Study’s Phase 2 combines independent individual judgments with panel processes. The primary goal of the second, or group, review was to adjudicate differences where possible in coding of the packets completed during the independent review and to produce group-level coding of the additional prerequisite KSAs that were not found in the NAEP frameworks. The final result of this two-part review process is a comprehensive list of prerequisite KSAs, answering the Board’s first research question: what are the prerequisite KSAs in reading and mathematics to qualify for entry-level, credit-bearing courses that satisfy general education requirements?

The final step in Phase 2 was for the NAEP experts to review the results of the KSA prerequisite data collected from the content reviewers, which was summarized in content maps. The NAEP experts' primary task was to compare these data with the 12th grade NAEP 2009 and 2013 items, achievement level descriptions, and minimal academic preparedness descriptions (from the JSS studies) in both mathematics and reading.

February 2014 Update: The review of course artifacts by content review teams and NAEP experts is complete. The review process yielded reliable data on the prerequisite KSAs for entry-level college courses. The review process was conducted with respect to four courses each for mathematics and reading. For the mathematics analyses, the courses were: (1) college algebra, (2) finite mathematics, (3) introduction to calculus, and (4) introduction to statistics. For the reading analyses, the courses were: (1) English literature, (2) introduction to psychology, (3) U.S. government, and (4) U.S. history.

Phase 3: Analysis and reporting

Phase 3 includes processing and analyzing the judgments collected during the review of course artifacts by review teams, and preparing the data to be reported in ways that are directly responsive to research questions in accordance with the analysis plan specified within the Design Document. Standard statistical methods will provide evidence of validity and reliability, and both conceptual (information processing/document analysis) and technical (quantitative) analyses will be conducted. The CCCA study is structured to provide a fully crossed, three factor design to ensure that results can be reviewed in statistical generalizability analyses, which will allow us to evaluate the reliability of the study design.

Final results will include narrative summaries of the prerequisite knowledge, skills, and abilities in mathematics and reading. Summary analyses will also address all aspects of the CCCA study design.

February 2014 Update: The analysis and reporting are underway with the final report due on April 30, 2014.

EVALUATING READING AND MATHEMATICS FRAMEWORKS AND ITEM POOLS AS MEASURES OF ACADEMIC PREPAREDNESS FOR COLLEGE AND JOB TRAINING

Project Status Update

Contract ED-NAG-13C-0001

The National Assessment Governing Board contracted with the Human Resources Research Organization (HumRRO) in June 2013 to conduct three tasks related to research on 12th grade preparedness:

1. **Evaluation of the Alignment of Grade 8 and Grade 12 NAEP to an Established Measure of Job Preparedness:** This study will extend prior analysis of the relation of NAEP to measures such as WorkKeys by including the NAEP grade 8 assessments and by expanding the method for assessing content alignment. The study method will follow the Governing Board content alignment design document for preparedness research studies, with some modifications. The two-pronged approach includes alignment of: (a) the training preparedness measure to the NAEP frameworks; and (b) NAEP items to the framework from which the training preparedness measure was developed.
2. **O*NET Linkage Study:** This study is a content validity investigation. Major duties (MDs) for the five target occupations will be identified. The occupations are automotive master technician, computer support specialist, HVAC technician, licensed practical nurse, and pharmacy technician. Expert raters will link NAEP content to MDs; NAEP content to O*NET knowledge, skills, and abilities (KSAs); and O*NET KSAs to MDs. This study will identify any disconnects between the level of constructs measured by NAEP and the level of those constructs required for entry into job training programs.
3. **Technical Advisory Panel (TAP) Symposium:** As part of the current contract, HumRRO assembled a technical advisory panel (TAP) of five experts in educational measurement and five experts in industrial-organizational (I-O) psychology to review extant research and to generate ideas for commissioned papers on preparedness. Each panelist is being asked to propose a paper that he/she could develop. Governing Board staff will review the proposals and select up to 10 papers. Panelists will have several months to develop the papers, after which the TAP will reconvene in a late 2014 symposium. Authors will present their papers and the entire panel will discuss implications for preparedness research. HumRRO will produce a proceedings document summarizing the commissioned papers and discussion. (A list of TAP members is included at the end of this document.)

In addition, HumRRO will produce a comprehensive report at the conclusion of the contract in December 2014.

Work completed as of February 2014:

Evaluation of Alignment of Grade 8 and 12 NAEP to an Established Measure of Job

Preparedness: With support from the National Center for Education Statistics (NCES), contacted NAEP State Coordinators to identify educators well-suited to serve on alignment workshop panels. Recruitment is underway.

O*NET Linkage: HumRRO developed lists of major duties (MDs) for each occupation based on O*NET task lists and course objectives from training curricula. Content experts in each occupation reviewed and vetted the MDs. We also obtained NAEP items for the O*NET Linkage Study. HumRRO staff were trained and completed linking NAEP content to O*NET knowledge, skills, and abilities (KSAs). The next step is to analyze the linkage rating data from the HumRRO raters.

Technical Advisory Panel (TAP) Symposium: The initial brainstorming meeting of the TAP was convened in October 2013. Following the meeting, panelists submitted nine proposals for papers related to 12th grade preparedness research. Governing Board staff reviewed the proposals and selected a number of papers to be developed by TAP members. Draft papers are due in July 2014; final versions will be presented at the second TAP meeting in August 2014.

Technical Advisory Panel (TAP) Members

John Campbell

Professor of Psychology
University of Minnesota
(Member, NAGB Technical Panel on 12th
Grade Preparedness Research, 2007-2008)

Michael Campion

Herman C. Krannert
Professor of Management
Purdue University

Gregory Cizek

Professor of Educational Measurement
and Evaluation
University of North Carolina at Chapel Hill

Brian Gong

Executive Director of Center for Assessment
National Center for the Improvement of
Educational Assessment, Inc.

Ronald Hambleton

Distinguished University Professor,
Educational
Policy, Research, & Administration
Executive Director, Center for Educational
Assessment
University of Massachusetts at Amherst

Suzanne Lane

Professor, Research Methodology
University of Pittsburgh School of
Education

Kenneth Pearlman

Independent Consultant in Industrial-
Organizational Psychology
Sarasota, FL

Barbara Plake

University Distinguished Professor,
Emeritus
University of Nebraska-Lincoln

Ann Marie Ryan

Professor of Psychology
Michigan State University

Nancy Tippins

Senior Vice President
CEB Valtera

Technology and Engineering Literacy (TEL) Achievement Levels Setting

The achievement levels setting for the 2014 grade 8 NAEP Technology and Engineering Literacy Assessment will take place in two stages via separate procurements. First, the achievement levels descriptions (ALDs) for grade 8 must be finalized. The preliminary achievement levels descriptions for each assessment area (Technology and Society; Design and Systems; and Information and Communication Technology) were developed as part of the Technology and Engineering Literacy Framework development project¹. The Governing Board requires expert evaluation of these descriptions in order to develop final descriptions for use in the achievement levels setting process. The final descriptions will be for the overall Technology and Engineering Literacy Assessment rather than separately by assessment area.

A statement of objectives was developed, and a Request for Quotations was issued in early February. Responses are due in early March, and the expected award date is April 1, 2014. The final achievement levels descriptions will be presented to COSDAM and the full Board for approval at the July 31 – August 2, 2014 meeting.

The second procurement is for the 2014 Technology and Engineering Literacy achievement levels setting at grade 8. The solicitation is scheduled to be posted in late February, with a projected award date of June, 2014.

The Statement of Work (SOW) for the TEL achievement levels setting will include the requirement for a pilot study and an operational achievement levels setting study. The Technical Advisory Committee on Standard Setting (TACSS) will need to include at least five members, one of whom must be a state assessment director or coordinator. The methodology for setting achievement levels for Technology and Engineering Literacy will not be specified. However, computers will need to be incorporated into the achievement levels setting process. Offerors will be provided with the opportunity to adapt the Governing Board's prior standard setting tools used in the writing standard setting process and the Judgmental Standard Setting Studies—Body of Work Technological Integration and Enhancements (BOWtie) or Computer-Aided Bookmarking (CAB), developed under previous contracts with the Governing Board².

¹ <http://www.nagb.org/publications/frameworks/technology/2014-technology-framework.html>

² As part of the contract closeout activities for the 2011 NAEP Writing achievement levels setting at grades 8 and 12 (contract #ED-NAG-10-C-0003) and the NAEP grade 12 preparedness research project judgmental standard setting (JSS) studies (contract #ED-NAG-10-C-0004), user manuals and technical documentation are being developed for CAB and BOWtie. The software, user manuals, and technical documentation are intended to be made available to the public in April, 2014.

Update on Implementation of SD/ELL Exclusion Policy

The March 2010 Governing Board policy on NAEP Testing and Reporting on Students with Disabilities (SD) and English Language Learners (ELL) was intended to reduce exclusion rates and provide more consistency across jurisdictions in which students are tested on NAEP to promote sound reporting of comparisons and trends. The policy limits the grounds on which schools can exclude students from NAEP samples to two categories – for SD, only those with the most significant cognitive disabilities, and for ELL, only those who have been in the U.S. schools for less than a year. Previously, schools excluded students with Individualized Education Programs (IEPs) that called for accommodations on state tests that NAEP does not allow, primarily the read-aloud accommodation on the Reading assessment. Under the current Board policy, schools should no longer decide to exclude students whose IEPs for state tests specify an accommodation not allowed on NAEP. Instead, such students should take NAEP with allowable accommodations. Additionally, parents and educators should be encouraged to permit them to do so, given that NAEP provides no scores and causes no consequences for individuals but needs fully representative samples to produce the valid results for the groups on which it reports. By law, individual participation in NAEP is voluntary and parents may withdraw their children for any reason.

During the December 2013 Board meeting, the Committee on Standards, Design, and Methodology and the Reporting and Dissemination Committee met in joint session to discuss the 2013 participation data for grades 4 and 8 Reading and Mathematics. There have been large increases in inclusion rates over the past ten years, and the Board’s first inclusion rate goal—95 percent of all students in each sample—was met in almost all states in 2013. However, 11 states and eight districts failed to meet the Board’s second goal of testing at least 85 percent of students identified as SD or ELL. Contrary to Board policy, NCES has continued to permit schools to exclude students whose IEPs called for accommodations that NAEP does not allow. NCES believes changing this practice could possibly be detrimental to students, increase refusals, change NAEP’s target population, and be counter to current statistical procedures.

At the end of the December 2013 joint session, the Committees asked the staffs of NAGB and NCES to consider possible policy and operational changes and what their impact might be, as well as a timeline for possible Board action.

February 2014 Update: The staffs of NAGB and NCES have had several conversations about the implementation of the SD/ELL policy, which have included the following possible next steps:

- It would be helpful to look at the universe of students who were sampled for NAEP and receive an accommodation on their state tests that is not allowed on NAEP, to examine

the percentages of such students who participate in NAEP even without that accommodation. Unfortunately, the data collection procedures for 2013 did not enable this question to be explored. We have recommended that the SD/ELL questionnaire be modified for 2015 to better differentiate between allowable NAEP accommodations (i.e., reading aloud the test directions) and non-allowable NAEP accommodations (i.e., reading aloud the reading passages and/or items) used on state tests.

- The policy could be clarified by revising the language about converting excluded students to refusals. The fourth implementation guideline for students with disabilities states, *“Students refusing to take the assessment because a particular accommodation is not allowed should not be classified as exclusions but placed in the category of refusals under NAEP data analysis procedures.”* NCES asserts that it is technically incorrect to apply a weight class adjustment that combines students who did not participate due to receiving accommodations on their state tests that are not allowed on NAEP with students who refused for other reasons. The former group cannot be assumed to be randomly missing, which is a necessary assumption for the current NAEP statistical procedures. However, NCES will explore other methods for imputing scores for such students, so that their lack of participation can be considered appropriately when calculating the NAEP scores.
- NCES has agreed to provide information about total participation rates in a manner that is transparent and more prominently displayed than the current approach.

We expect to convene a follow-up joint session during the May 2014 Board meeting, with possible steps for Board action at the August 2014 Board meeting.

Update on Evaluation of NAEP Achievement Levels Procurement

Objective To receive a brief informational update from NCES on the current status of the procurement being planned to evaluate NAEP achievement levels. Ongoing updates will be provided at each COSDAM meeting.

Background

The NAEP legislation states:

The achievement levels shall be used on a trial basis until the Commissioner for Education Statistics determines, as a result of an evaluation under subsection (f), that such levels are reasonable, valid, and informative to the public.

In providing further detail, the aforementioned subsection (f) outlines:

(1) REVIEW-

- A. IN GENERAL- The Secretary shall provide for continuing review of any assessment authorized under this section, and student achievement levels, by one or more professional assessment evaluation organizations.
- B. ISSUES ADDRESSED- Such continuing review shall address--
 - (i) whether any authorized assessment is properly administered, produces high quality data that are valid and reliable, is consistent with relevant widely accepted professional assessment standards, and produces data on student achievement that are not otherwise available to the State (other than data comparing participating States to each other and the Nation);
 - (ii) whether student achievement levels are reasonable, valid, reliable, and informative to the public;-
 - (iii) whether any authorized assessment is being administered as a random sample and is reporting the trends in academic achievement in a valid and reliable manner in the subject areas being assessed;
 - (iv) whether any of the test questions are biased, as described in section 302(e)(4); and

- (v) whether the appropriate authorized assessments are measuring, consistent with this section, reading ability and mathematical knowledge.

(2) REPORT- The Secretary shall report to the Committee on Education and the Workforce of the House of Representatives and the Committee on Health, Education, Labor, and Pensions of the Senate, the President, and the Nation on the findings and recommendations of such reviews.

(3) USE OF FINDINGS AND RECOMMENDATIONS- The Commissioner for Education Statistics and the National Assessment Governing Board shall consider the findings and recommendations of such reviews in designing the competition to select the organization, or organizations, through which the Commissioner for Education Statistics carries out the National Assessment.

Responsively, a procurement has been planned to administer an evaluation of NAEP achievement levels. The last update COSDAM reviewed on this topic was in August 2013.

In the following brief written update, NCES provides the Committee with a summary of the status of this procurement.

Evaluation of NAEP Achievement Levels

The National Center for Education Evaluation and Regional Assistance (NCEERA), part of the Institute for Education Sciences (IES), will administer the Evaluation of the NAEP Achievement Levels. NCEERA and the Department of Education's Contracts and Acquisitions Management (CAM) office will begin this procurement during fiscal year 2014. Tentatively, the scheduled award date is June, 2014. This will be a full and open competition.