

National Assessment Governing Board Committee on Standards, Design and Methodology

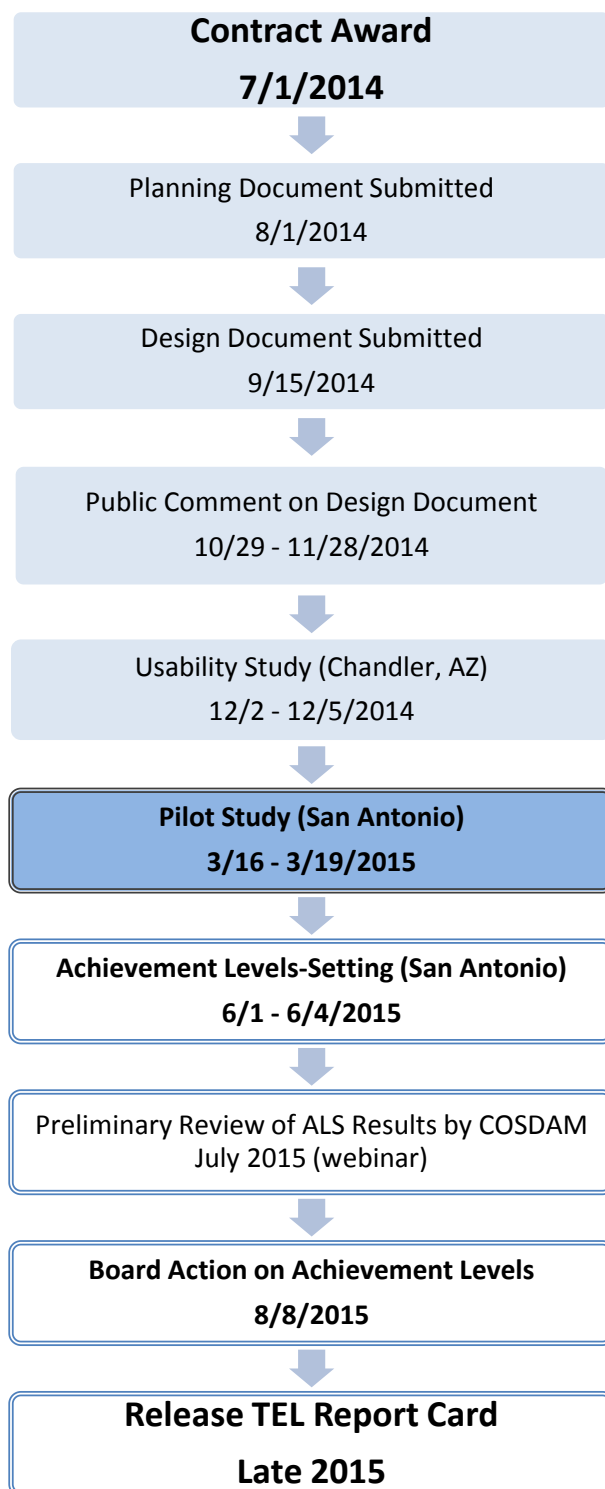
March 6, 2015
10:00 am – 12:30 pm

AGENDA

10:00 – 10:05 am	Introductions and Review of Agenda <i>Lou Fabrizio, COSDAM Chair</i>	
10:05 – 10:45 am	Project Update for Technology and Engineering Literacy (TEL) Achievement Levels Setting <i>Steve Fitzpatrick, Pearson</i>	Attachment A
10:45 – 11:35 am	Update on Transition to Digital-Based Assessment <i>Andreas Oranje, ETS</i>	Attachment B
11:35 – 11:55 am	Developing a Resolution on Maintaining Trend with Transition to Digital Based Assessment <i>Lou Fabrizio, COSDAM Chair</i>	Attachment C
11:55 am – 12:10 pm	Governing Board Strategic Planning Initiative <i>Lou Fabrizio, COSDAM Chair</i>	See Strategic Planning Tab in Board Materials
12:10 – 12:25 pm	Update on Academic Preparedness Research <i>Sharyn Rosenberg, Governing Board Staff</i>	Attachment D
12:25 – 12:30 pm	Other Issues and Questions <i>COSDAM Members</i>	
	Information Item: Update on Evaluation of NAEP Achievement Levels	Attachment E

DEVELOPING ACHIEVEMENT LEVELS FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS TECHNOLOGY AND ENGINEERING LITERACY (TEL) AT GRADE 8

Project Update – February 2015



Purpose: The purpose of this session is to provide an update to the Committee on Standards, Design and Methodology (COSDAM) regarding the development of achievement levels for the 2014 NAEP TEL and to present the plans for implementing the item mapping standard setting methodology.

In this session, Dr. Steve Fitzpatrick, NAEP TEL Achievement Levels-Setting (ALS) Project Director for Pearson, will provide an update on the project and a review of the usability study findings.

Legend:

Light shading: Completed

Dark shading: Current status

No shading: To be completed after March 2015

Project Overview: On July 1, 2014, the National Assessment Governing Board awarded a contract to Pearson to develop achievement levels for the National Assessment of Educational Progress (NAEP) Technology and Engineering Literacy (TEL) assessment. The computer-based 2014 NAEP TEL is based on the Board-adopted TEL Framework and consists of both scenario-based tasks and discrete items. The first-ever TEL assessment was administered to a nationally representative sample of more than 22,000 grade 8 students in 2014.

Working with Conference Solutions, EdCount and Measurement Incorporated as subcontractors, Pearson will conduct a usability study, a pilot study and an achievement levels-setting (ALS) meeting and produce a set of recommendations for the Governing Board to consider in establishing achievement levels for the NAEP TEL. Pearson will implement an item mapping methodology using software developed by Measurement Incorporated to collect panelists' ratings and present feedback. Conference Solutions will assist Pearson in planning and delivering meetings. Dr. Lori Nebelsick-Gullett from EdCount will serve as the process facilitator for the pilot and operational ALS meetings; Dr. Johnny Moyer will serve as the content facilitator for the pilot and operational ALS meetings; and Dr. Susan Cooper Loomis will serve as a consultant to Pearson.

For achievement levels setting, Pearson will use an item mapping process in which panelists will make criterion-referenced, content-based cut score recommendations. The content-based judgments will be made over three rounds. The process to be implemented for the standard setting meeting follows item mapping procedures used in previous NAEP standard setting studies. In addition, two studies will be completed prior to the pilot study: 1) a study of the functioning of the standard setting software used to collect panelists' item ratings and other judgments, and 2) a dual-computer usability study.

The Governing Board policy on Developing Student Performance Levels for NAEP requires appointment of a committee of technical advisors who have expertise in standard setting and psychometrics in general, as well as issues specific to NAEP. These advisors will be convened for 5 in-person meetings and up to 3 webinars to provide advice at every key point in the process. They provide feedback on plans and materials before activities are implemented and review results of the process and analyses. Six external experts in standard setting are serving on the Technical Advisory Committee on Standard Setting (TACSS):

Dr. Gregory Cizek

Professor of Educational Measurement, The University of North Carolina at Chapel Hill

Dr. Barbara Dodd

Professor of Quantitative Methods, The University of Texas at Austin

Dr. Kristen Huff

Senior Fellow, the University of the State of New York Regents Research Fund

Dr. Matthew Johnson

Associate Professor of Statistics and Education, Teachers College, Columbia University

Dr. Marianne Perie

Director, Center for Educational Testing and Evaluation, University of Kansas

Dr. Mary Pitoniak

Strategic Advisor for Statistical Analysis, Data Analysis, and Psychometric Research,
Educational Testing Service (NAEP Design, Analysis, and Reporting Contractor)

February 2015 Update:

Project Staffing

In February 2015, Dr. Steve Fitzpatrick became the Pearson Project Director for the TEL ALS. He replaces Dr. Paul Nichols, who has taken a new position outside of Pearson. Dr. Fitzpatrick is a Principal Research Scientist in Pearson's Assessment Solutions and Design group. Dr. Fitzpatrick has provided psychometrics, research support, and consulting for state and national testing programs delivered by Pearson. He has served as lead psychometrician on several large scale statewide assessment programs including Texas, Virginia, Arizona, and Puerto Rico. Dr. Fitzpatrick's areas of expertise include linking and scaling designs, equating and scaling based on IRT models, development of large-scale and standards-referenced assessment systems, standard setting, and psychometric and statistical software.

Technical Advisory Committee on Standard Setting (TACSS) meeting (December 18-19, 2014)

The TACSS met December 18-19, 2014, to discuss preparations for the pilot study, review findings from a usability study and discuss the collection of public comment on the achievement levels-setting outcomes:

1. The TACSS recommended that comments on the ALS outcomes should be collected using a discussion format and two small groups where each group consists of 10 to 15 participants. The two groups would meet at the end of the National Conference on Student Assessment but each group would meet separately. The TACSS expressed concern that the context in which the comments were collected from these two groups be considered when interpreting these comments.
2. After reviewing the results of the usability study, the TACSS recommended that the standard setting software should be configured to reduce the cognitive load imposed on the standard setting panelist as much as possible. This included the recommendation that the ordered item book screen show the item number and the title of the scenario.

3. The TACSS noted that uninterrupted wireless connectivity will be critical to the success of the standard setting meeting. The TACSS recommended that Pearson provide an independent wireless source for the meeting room and not depend on the wireless service provided by the hotel.
4. The TACSS expressed concern that the tables used by the standard setting panelists must be large enough to accommodate two computers for five panelists at a table. The TACSS recommended that Pearson explore the layout of the standard setting meeting room and the spacing of the computers at the tables.
5. Copies of the panelist questionnaires were provided to the TACSS members several days before the TACSS meeting and again during the meeting. The TACSS observed that the response options appeared to show a positive bias. The TACSS recommended that the labels be revised to eliminate any positive bias. In addition, the TACSS recommended that open-ended questions be included in the questionnaire and that the questions be worded in a manner that did not imply the panelist was failing or had failed in some way.

The TACSS will meet in February and is scheduled to meet again at the end of April, 2015. During the February meeting, the TACSS will discuss preparations for the March 2015 pilot test, including a demonstration of the software used to collect cut score recommendations and a review of facilitator training materials.

Findings from the Usability Study

On December 2-4, 2014, Pearson conducted a usability study to learn more about the dual computer set-up planned for the pilot study and ALS meeting. The usability study was designed to collect data to answer the following questions:

- Can panelists navigate the items, including viewing them within their scenarios, and complete the bookmark process?
- What kind of training is needed to train the panelists to navigate the dual-computer setup?
- Given the dual-computer setup, how long do panelists need to complete the ALS ratings?

A group of five grade eight science teachers from the Phoenix area participated in the study. Each teacher participated individually. Each teacher completed a four-hour session with the following activities:

- Completion of a form of the NAEP TEL assessment;
- Review of the knowledge, skills and abilities required by the items in the assessment form;

- Review of the achievement levels descriptions and creation of borderline descriptions; and,
- Completion of two rounds of standard setting.

The findings from the study indicated that all participants were comfortable moving between computers. The most difficult activity was mastering the features of the software on the two computers. Panelist training for the pilot study and ALS meeting should include directed practice on navigating the software on the two computers. The study findings suggested that the time provided for completing standard setting rounds was sufficient. However, panelists required additional time to review the knowledge, skills and abilities required by the items. Overall, the usability study findings indicated that the use of two computers was a comfortable arrangement for standard setting panelists.

Public Comment on the Design Document

The Design Document is intended to provide the foundation for all achievement levels-setting activities. The Design Document for the TEL achievement levels-setting process includes discussion of the methodology, procedures, and documentation of the entire project.

The Design Document was distributed for public comment from October 29 - November 28, 2014. Information about the public comment period appeared on both the Governing Board and Pearson websites and was sent to over 100 interested organizations as part of the panelist nomination effort. No comments were received.

Update on Panelist Recruitment

A group of 51 panelists were recruited to participate in the ALS activities consisting of 15 panelists for the pilot study, 30 panelists for the ALS meeting and six panelists as replacements. Furthermore, the 51 panelists for the TEL assessment standard setting should consist of 28 grade 8 classroom teachers, 15 members of the general public and eight non-teacher educators.

Panelists recruited for standard setting did meet the targets for recruitment and consist of 28 classroom teachers currently engaged in TEL instruction at grade eight (55 percent), 15 members of the general public (30 percent), and eight non-teacher educators (15 percent). This group reflects an overall balance of gender, race/ethnicity, geographic location, type of TEL experience and institutional affiliation.

Next Steps for COSDAM:

The May 2015 update to COSDAM will include a report on the results of the pilot study.



Update on Transition to Digital Based Assessment (DBA)

The Digital Based Assessment *start-up* administration is wrapping up in the field, signaling an important step in the transition away from paper-based testing and towards tablet-based assessments. The start-up has been designed to answer a number of questions, primarily about the relationship between tablet-based and paper-based assessment, with the goal to maintain trends across the transition. In order to communicate design principles and goals and to engage various audiences in discussions about the DBA transition design, ETS developed a white paper. This document was subsequently reviewed by a subset of experts from NAEP's advisory committees (the Design and Analysis Committee, the NAEP Validity Studies panel, and the Quality Assurance Technical Panel), has become part of ongoing discussions with DAC, and also shared and discussed with COSDAM at the most recent Governing Board meeting.

Based on all these discussions and reviews, ETS has revised the white paper and proposed some changes to the 2017 design and is considering some additional studies in line with comments and suggestions made by the various discussion partners. Some of those changes and studies are directly implementable, whereas others have budgetary implications that need to be weighed and prioritized against other assessment activities. In this update presentation, we will discuss the following topics, which represent the key themes distilled from all the rich discussion and commentary:

1. The current design for math and science combines the transition to DBA and the transition from linear to multi-stage testing. Can these effects be disentangled?
2. Digital equity is expected to be a very important factor and getting a good handle on this factor will be critical. How are we approaching this?
3. There is significant interest in evaluating the transition at the state level, which is limited in the current design. This evaluation could be conceived of as a validation to aid interpretation or even as a way, if technically defensible, to combine paper-based assessment results and digitally-based results and effect a slower transition. Can a design with a paper to digital direct comparison at the state level and within a single year be developed and implemented?
4. Looking ahead at 2017, the introduction of Scenario Based Tasks to yield more interactive tasks that hold the promise of eliciting and gathering evidence about cognitive processes involved in responding, is a notable addition. In what ways are these exactly used and what value do we expect from them?
5. What evaluation criteria do we plan to apply at what points along the transition? What outcomes would lead to what decisions?

As part of the discussion of these themes, we will present the directions we are taking to address the questions above and discuss possible and probable solutions. We welcome COSDAM's perspectives as we further refine the designs and build the foundation that gives us excellent odds for robust and meaningful trend reporting in the face of rapid technological advancement.

As part of the briefing materials for this session, we plan to send the revised white paper to COSDAM members via email. This is currently an embargoed document.

Resolution on Maintaining Trend with Transition to Digital Based Assessments (DBA)

Over the past year, the Board has had several discussions about the importance of maintaining trend with the shift towards digital based assessment (DBA). The following is an excerpt from the COSDAM minutes from November 21, 2014:

“COSDAM members emphasized the critical importance of the 2017 NAEP results and maintenance of trend, given all of the changes occurring in state assessments. Terry Holliday stated that if the cost of moving to TBA is that we lose the trend, then NAEP’s gold standard will be undermined. There was consensus that everything possible must be done upfront to maintain the trend, and that the question should be reframed as *how* rather than *whether* trend can be maintained. There was considerable discussion about the extent to which the trend decision is a policy issue. It is unlikely that the data from the bridge studies will be definitive, and the narrative around the trend decision (including any caveats) will be as important as the trend decision itself.”

The Board staff proposes that a Resolution be developed to formally document and articulate the Board’s position on the importance of maintaining trend during the transition to DBA. We are seeking input on the development of such a Resolution, which we would bring to COSDAM for review during the May 2015 Board meeting.

Discussion Questions

1. What information should be included in the rationale for the importance of maintaining trend during the DBA transition?
2. How should the Board be involved in the maintenance of trend and interpretation of results from the DBA transition?

NAEP 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, released on May 14, 2014. The motion, validity argument, and phase 1 final report are now available 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 2015 Update
Statistical linking of NAEP and ACT	National; FL, IL, MI, TN	Page 11
Statistical linking of NAEP and SAT	MA	Page 12
Longitudinal statistical relationships: Grade 12 NAEP	FL, IL, MA, MI, TN	Page 13
Statistical linking of NAEP and Explore	KY, NC, TN	Page 14
Longitudinal statistical relationships: Grade 8 NAEP	NC, TN	Page 15
Content Alignment Studies of the 2013 National Assessment of Educational Progress for Grade 8 Reading and Mathematics with ACT Explore Assessments of These Subjects		Pages 16 - 18
Evaluating Reading and Mathematics Frameworks and Item Pools as Measures of Academic Preparedness for College and Job Training		Pages 19 – 38
College Course Content Analysis		Page 39

Brief overviews and project 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 national 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 working with four states to be partners in these studies at grade 12: Florida, Illinois, 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.

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 2015 Update: Data have been received from MI and TN, and data analyses are underway. Data sharing agreements with ACT, FL and IL have not yet been finalized.

State Statistical Linking Study with the SAT

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

We have partnered with Massachusetts to conduct a state-level linking study for 2013 NAEP and the SAT. Again using a procedure that protects student confidentiality, SAT scores of NAEP 12th grade test takers in Massachusetts will be linked to NAEP scores.

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

1. What are the correlations between the grade 12 NAEP and SAT student score distributions in Reading and Math?
2. What scores on the grade 12 NAEP Reading and Math scales correspond to the SAT 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 SAT college readiness benchmarks?
4. Do the results differ by race/ethnicity or gender?

February 2015 Update: The data have been received from MA; data analyses are underway.

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.

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 2015 Update: The data sharing agreements have been finalized for MA, MI, and TN; longitudinal data files will be prepared and transmitted when available. Data sharing agreements with FL and IL are still being worked out.

State Statistical Linking Studies with ACT Explore

In 2013, linking studies between 8th grade NAEP in Reading and Mathematics and 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 was administered to all students state-wide who were in grade 8 during the 2012-13 school year.

Research Questions for State Statistical Linking Studies with ACT 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 2015 Update: The data have been received from all three states, and data analyses are currently underway.

Longitudinal Statistical Relationships: Grade 8 NAEP

In 2013, the Governing Board will also expand the state-level studies by partnering with two 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.

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

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 and NC)

February 2015 Update: The data sharing agreements are complete; analyses are currently underway (to address the first research question). Additional data will be transmitted when they become available over the next several years.

Content Alignment Studies of the 2013 National Assessment of Educational Progress for Grade 8 Reading and Mathematics with ACT Explore Assessments of These Subjects**Project Status Update
February 16, 2015
Contract ED-NAG-14-C-0002****Project Overview**

In September 2014, NORC at the University of Chicago, along with its subcontractor, the Wisconsin Center for Education Products and Services (WCEPS), were awarded a contract to conduct content alignment studies with the ACT Explore assessments in reading and mathematics and the 2013 National Assessment of Educational Progress (NAEP) Reading and Mathematics assessment at grade 8. The purpose of this research is to evaluate the extent to which 8th grade NAEP is aligned in content and complexity with the Explore assessment. For each subject area, the studies will compare the two assessments (NAEP and ACT Explore) to the NAEP framework, and also to the ACT College and Career Readiness framework. Using the content alignment methodology designed by Dr. Norman Webb for the Preparedness Research Program commissioned by the Governing Board, these studies will measure and describe the degree of alignment between the grade 8 NAEP math and reading assessments and ACT Explore assessments in these same subjects. The results of these NAEP-Explore content comparisons will also inform interpretations from statistical linking studies of 2013 results of NAEP and Explore in grade 8 reading and mathematics.

To support the provision of ACT proprietary Explore data, the Governing Board also issued a sole source contract with ACT, Inc. NORC is working with ACT to receive data and materials that will be used in the content alignment studies, and is consulting with ACT assessment staff to support the work and analyses.

February 2015 Update:

Project leaders at NORC at the University of Chicago, along with the subcontractor, Wisconsin Center for Education Products and Services (WCEPS), focused efforts during the quarter on preparation for the February 2015 Content Alignment Institute (CAI) to be convened at NORC facilities in Bethesda, MD. Several key milestones (see table below) were met and preparation activities continue.

During the months of October and November 2014, the NORC-WCEPS team secured a data use agreement with ACT, submitted the final Planning Document to the Governing Board, and completed the NAEP and ACT Framework Analysis Reports for reading and mathematics. These reports are a critical step in establishing the coding sheets and procedures and panelist instructions for the Content Alignment Institute, February 9-13, 2015. In addition to the

framework reports, NORC convened in November a meeting with invited statistical and psychometric experts for the purpose of study Design Review and Strengthening (DRS). This meeting served to vet and discuss the approach to the CAI to ensure its success, and it was attended by the study staff and Michelle Blair, the Governing Board Contracting Officer Representative.

With the input from the DRS meeting, NORC-WCEPS staff incorporated suggestions to the plans for the Content Alignment Institute, and in December of 2014, began to recruit the 32 educators (16 math, 16 reading) that would serve on the content analysis panels. Facilitator participation was also confirmed at this time. In December, NORC sent an invitation and recruitment letter to many national professional associations, all state departments of education, and school districts in the mid-Atlantic region. Efforts were made to contact and invite qualified educators at the middle grades level across the nation and to ensure that educators of varying backgrounds and experience could apply for participation as panelists. An effort was made to ensure representation of educators by age, gender, race/ethnicity, and region. From more than 100 applicants who provided a snapshot of their qualifications and experience, the NORC-WCEPS team selected and received confirmation of participation by 32 highly qualified educators. Four consultants with extensive experience with the content alignment methodology and procedures will serve as panel facilitators. NORC promptly began the process of securing travel and lodging arrangements.

Tasks recently underway were related primarily to the preparation and planning for the CAI meeting, and finalizing the coding sheets and other materials that will be used at the Institute. NORC-WCEPS received comments from the Governing Board staff, NCES staff, and ACT staff on the Framework Analysis Reports that have contributed to the materials and organization for the Institute. Copies of the assessment items to be reviewed have been received by NORC and are being maintained with security. The CAI meeting was conducted from February 9 – 13, 2015.

Milestones

There are several major milestones over the course of this project. The highlighted items are complete (milestone has been met). To briefly summarize, the milestones include preparatory work, data collection, and alignment analysis and reporting. The following table includes the major milestones for completing this work:

<i>Milestone</i>	<i>Date</i>
Kickoff Meeting	9/29/14 - <i>Complete</i>
Submit Planning Document	10/31/14 - <i>Complete</i>
Conduct Framework Analyses	10/16/14 – 11/4/14 - <i>Complete</i>
Recruit Panelists for Content Alignment Institute	11/1/14 – 12/19/14 - <i>Complete</i>
Convene Design Review and Strengthening Meeting	11/12/14 - <i>Complete</i>
Conduct Content Alignment Institute	2/9/15 – 2/13/15- <i>Complete</i>
Conduct Data Analysis	2/9/15-2/27/15
Prepare Draft Reports	3/8/15-4/15/15
Prepare Final Reports	5/21/15-7/1/15
Present final reports to COSDAM	7/15/15-8/7/15

COSDAM will receive ongoing updates as the study progresses.

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:** In its June 2009 report, *Making New Links: 12th Grade and Beyond*, the Technical Panel on 12th Grade Preparedness Research recommended that content alignment studies be conducted to examine the structure and content of various assessments relative to NAEP. The purpose of such content alignment would be to determine whether the scores on NAEP and the other assessments convey similar meaning in terms of the knowledge and skills of examinees. In fact, the panel specifically recommended that content alignment studies be conducted between NAEP and WorkKeys to determine the correspondence between the content domain assessed by NAEP and that of WorkKeys. If the alignment is relatively high, or even moderately high in some cases, then statistical relations between NAEP and WorkKeys may allow for the interpretation of NAEP results in terms of how WorkKeys would typically be interpreted. Using WorkKeys as a measure of job training preparedness allows the comparison of findings from this research to findings from previous content alignment studies with WorkKeys.

HumRRO extended prior analysis of the relation of NAEP to WorkKeys by including the NAEP grade 8 assessments and by expanding the method for assessing content alignment. ACT provided operational WorkKeys items in support of the study. The study method followed the Governing Board content alignment design document for preparedness research studies, with some modifications. The two-pronged approach included alignment of: (a) WorkKeys to the NAEP frameworks, and (b) NAEP items to the framework from which WorkKeys was developed.

The final executive summary of this report follows this Project Status overview.

2. **O*NET Linkage Study:** This study a) identified relevant linkages between the National Assessment of Educational Progress (NAEP) and training performance requirements for selected occupations, and b) compared the levels of knowledge, skills, and abilities (KSAs) required for the relevant NAEP content to the levels of KSAs required for the relevant job training content. For this study, tasks (i.e., performance requirements) for each occupation were extracted from O*NET. The O*NET, or Occupational Information Network, is the U.S. Department of Labor's occupational information database. The final executive summary from this study was included in the July 2014 Governing Board briefing materials.
3. **Technical Advisory Panel (TAP) Symposium:** 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. The TAP met in Washington D.C. in late October 2013. This brainstorming session included presentations by Governing Board and HumRRO staff describing findings from previous studies and descriptions of other studies currently underway, followed by an open discussion of issues and possible additional areas of investigation. Each panelist was asked to use this information to propose a paper that he/she could develop. TAP members submitted nine proposals from which Governing Board staff commissioned five papers. Panelists developed three of these papers and presented them in a TAP Symposium on August 20, 2014:

- *Using 8th and 12th Grade NAEP to Measure Student Readiness for Careers*, Barbara Plake, University of Nebraska – Lincoln
- *Grit: A Useful Concept in College and Career Preparedness?* Ann Marie Ryan, Michigan State University
- *Relating NAEP to Commercial Off-the-Shelf Measures*, Nancy Tippins, Corporate Executive Board – Valtera Corporation

The final proceedings document summarizing the commissioned papers and discussion was submitted to the Governing Board in February 2015.

In addition to these three tasks, HumRRO is producing a comprehensive project report at the conclusion of the contract. The draft report is currently under review by the Governing Board.

February 2015 Update:

Evaluation of Alignment of Grade 8 and 12 NAEP to an Established Measure of Job Preparedness: The final report was submitted to the Governing Board in December 2014. The executive summary is included with this update.

O*NET Linkage: This task was completed in April 2014. See July 2014 Governing Board status update for details.

TAP Symposium: Final proceedings from the TAP Symposium were submitted to the Governing Board in February 2015.

A comprehensive report has been finalized and is included with this update.

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

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

The Content Alignment between the NAEP and WorkKeys Assessments

Human Resources Research Organization (HumRRO)

Executive Summary

The National Assessment Governing Board (Governing Board), which sets policy and provides general oversight and direction for the National Assessment of Educational Progress (NAEP), is conducting a program of research to determine the feasibility of using NAEP to report on the preparedness of U.S. 12th grade students for entry into postsecondary education and job training.

The purpose of the current study is to explore the relationships between the NAEP Frameworks and item pools (for mathematics and reading) and the WorkKeys assessments for: Applied Mathematics, Applied Technology, Reading for Information, and Locating Information. This study expands upon prior research on the content alignment between NAEP and WorkKeys in three major ways (ACT, 2010a; 2010b). First, this study includes the Frameworks and items from the NAEP grade 8 assessments in order to address concerns raised that grade 8 Frameworks may provide a better match to the academic content expectations of job training programs (Kilpatrick, 2012; Loomis, 2012). Second, additional WorkKeys assessments (Applied Technology and Locating Information) are included in the study to determine the extent to which NAEP mathematics and informational reading content may relate to other WorkKeys assessments. Finally, the NAEP Frameworks are directly compared to the WorkKeys targets to determine the degree of overlap between the two content domains.

Several key results do not support the use of NAEP for determinations related to the academic preparedness of U.S. 12th grade students for entry into job training:

- NAEP items do not adequately represent the WorkKeys content domain, as evidenced by the percentages of WorkKeys' mathematics and reading targets (52% and 72%, respectively) that were not matched to any NAEP item.
- Sixteen of the 24 content strands within the NAEP Math Framework and one of the three cognitive targets within the NAEP Reading Framework were not matched to any WorkKeys item.
- A direct comparison of the content domains for the two assessments indicated that the majority of the elements of the NAEP Math Framework, WorkKeys math targets, and WorkKeys applied technology targets reflected unique content.

These results are not all that surprising given the differing purposes of the two assessments. While NAEP has been designed to provide evidence of what students in the United States know and can do in a broad academic sense, WorkKeys assessments provide information about job-related skills that can be used in the selection, hiring, training, and development of employees. It should be noted that while this study found that much of the content assessed by WorkKeys and NAEP did not overlap, lack of overlap is not evaluative of either assessment. The lack of overlap in content likely reflects substantial differences in purpose and design. Both assessments may function very well for their specified purposes without exhibiting great similarity in the content they measure.

Finally, while there is some indication that there is more overlap between grade 8 NAEP items and WorkKeys targets, which is consistent with information provided in related research studies

(Loomis, 2012; Sinclair, Becker, McCloy, & Thacker, 2014; Educational Policy Improvement Center, 2013), the results of this study suggest that including grade 8 NAEP does not improve the level of alignment between NAEP and WorkKeys, nor would the 8th grade NAEP assessments be an appropriate measure of academic preparedness for job training.

2015 No. 004



Evaluation of NAEP 12th Grade Reading and Mathematics Frameworks and Item Pools as Measures of Academic Preparedness for College and Job Training

Comprehensive Report

Prepared for: National Assessment Governing Board
800 North Capitol Street, N.W., Suite 825
Washington, DC 20002-42330

Authors: D. E. (Sunny) Becker
Emily R. Dickinson
Rodney A. McCloy
Andrea L. Sinclair
Arthur A. Thacker

Prepared under: National Assessment of Educational Progress (NAEP)
Reading and Mathematics Frameworks: 12th Grade
Preparedness
ED-NAG-13-C-0001

Date: February 9, 2015

Evaluation of NAEP 12th Grade Reading and Mathematics Frameworks and Item Pools as Measures of Academic Preparedness for College and Job Training

Executive Summary

The National Assessment Governing Board (hereafter, “Governing Board”) contracted with the Human Resources Research Organization (HumRRO) to conduct a multi-method evaluation of the 12th grade National Assessment of Educational Progress (NAEP) reading and mathematics frameworks and item pools as a measure of preparedness for college and job training. This effort included three distinct investigations; 1) a content alignment study between NAEP and the WorkKeys assessments, 2) comparisons between NAEP and U.S Department of Labor’s occupational information network, or O*NET, for five target occupations, and 3) exploratory papers commissioned from a joint panel of education and industrial/organizational psychology experts, the NAEP Framework Evaluation Technical Advisory Panel (TAP). This report summarizes the entire project and includes highlights from each study.

In summary, findings from both studies—the content alignment study between NAEP and WorkKeys and the study comparing NAEP content and O*NET content—call into question the validity of inferences that can be made about using NAEP to report on the preparedness of U.S. 12th grade students for entry into job training. Moreover, during the TAP symposium many challenges were discussed with regard to conducting research on academic preparedness for job training such as the variability in training programs across and even within occupations. Challenges such as these, coupled with research findings that have provided converging evidence that NAEP measures reading and math content that is broader and at a higher level than the reading and math knowledge, skills, and abilities (KSAs) required for entry into job training, call into question whether the Governing Board should continue to conduct research on using NAEP as an indicator of academic preparedness for entry into job training.

If, however, the Governing Board decides to move forward with preparedness research for job training, several avenues for future research were provided. These avenues fall into essentially two categories. First, modify NAEP itself or NAEP reporting to more directly address job preparedness. This could mean reporting subscores based on relevant parts of the frameworks, re-administering portions of the grade 8 NAEP to high school students, or adding background questions to measure additional constructs that are more relevant to job preparedness (e.g., grit). The second avenue for research involves monitoring longitudinal performance of students with known NAEP scores. This would allow linking to job training performance or job performance. This avenue might help NAEP establish performance categories more directly relevant to job training performance or to select specific frameworks statements on which to base a job preparedness indicator.

Evaluation of NAEP 12th Grade Reading and Mathematics Frameworks and Item Pools as Measures of Academic Preparedness for College and Job Training

Table of Contents

Executive Summary	i
Introduction	3
Background.....	3
This Report	5
Content Alignment Between the NAEP and WorkKeys Assessments	6
Comparisons Between NAEP and O*NET on Academic Preparedness for Job Training for Five Target Occupations	7
NAEP Framework Evaluation Technical Advisory Panel	9
Summary of HumRRO Investigations.....	11
References	13

Evaluation of NAEP 12th Grade Reading and Mathematics Frameworks and Item Pools as Measures of Academic Preparedness for College and Job Training

Introduction

The National Assessment Governing Board (hereafter, “Governing Board”) contracted with the Human Resources Research Organization (HumRRO) to conduct a multi-method evaluation of the 12th grade National Assessment of Educational Progress (NAEP) reading and mathematics frameworks and item pools as a measure of preparedness for college and job training. This study advances the Governing Board’s preparedness research agenda, which has been underway since 2010.

The HumRRO evaluation included three distinct investigations. Detailed reports for each of these efforts are listed in the References section. This report summarizes the entire project and includes highlights from each study. To provide context, the following section summarizes the history of the Governing Board’s preparedness research preceding this work.

Background

In March 2003, the Governing Board established a National Commission on National Assessment of Educational Progress (NAEP) 12th Grade Assessment and Reporting. Co-chaired by Mark Musick (Southern Regional Education Board) and Michael Nettles (Educational Testing Service), the commission comprised 18 members representing K-12 education, postsecondary education, business and industry, and the military. The commission focused on improvements needed for 12th grade NAEP, as opposed to reviewing the status of secondary education or evaluating the quality and rigor of high school curricula (U.S. Department of Education, 2004).

Questions addressed by the commission included: (a) What does America need to know about 12th graders soon to graduate from high school? (b) What can NAEP do to provide that information? (c) What should be the content of 12th grade NAEP? (d) How can 12th graders be motivated to do their best on NAEP? and (e) How can the Governing Board and the National Center for Education Statistics promote 12th grade NAEP participation (Sellman, Wise, Schultz, & Schantz, 2003)?

In a March 2004 report entitled, *Twelfth Grade Student Achievement in America: A New Vision for NAEP*, the commission concluded that 12th grade NAEP should be redesigned. One commission recommendation was that NAEP should report 12th grade student preparedness for postsecondary learning (i.e., transition to higher education, training for employment, and entrance into the military) (U.S. Department of Education, 2004). Reporting postsecondary preparedness would be a new direction for NAEP that would require researching whether NAEP results could be used to predict preparedness for college, civilian employment training, and military enlistment.

For the next several years, the Governing Board explored potential approaches for implementing the commission’s recommendations. In June 2007, the Governing Board convened a Technical Panel on 12th Grade Preparedness Research to assist in planning research and validity studies that would enable NAEP to report on the preparedness of 12th

graders for education and job training after high school graduation. In particular, the panel was to identify and recommend research designs and priorities for the use of 12th grade NAEP as a predictor of postsecondary preparedness. Composed of seven members with a wide variety of expertise in higher education policy, the civilian and military worlds of work, and psychological and educational measurement, the panel was chaired by Michael Kirst (Stanford University).

After a year of deliberations, the panel concluded: (a) reporting performance on 12th grade NAEP in relation to preparedness was feasible and should be pursued, (b) a multipronged approach was the way to gain information regarding the complex relationship between NAEP and postsecondary school success, (c) NAEP's reporting on preparedness should not be construed as the single authoritative definition or conception of preparedness, and (d) the evolving national context related to preparedness policies and practices was a dynamic that required special consideration. The panel recommended that preparedness should represent the academic knowledge and skills in reading and mathematics necessary to enroll in a credit-bearing undergraduate course without need for remediation or to be qualified for placement into a civilian or military job training program (U.S. Department of Education, 2009). This definition of preparedness is meant to be distinct from "readiness," which includes other important aspects of student performance that may mediate post-high school outcomes such as time management, persistence, and interpersonal skills, among others.

In November 2008, Dr. Kirst presented the panel's findings to the Governing Board. The panel recommended several types of studies, to include:

- Content alignment studies to evaluate the extent of content overlap between NAEP and other assessments.
- Criteria-based judgmental standard setting studies by subject matter experts to set cut scores on NAEP via a performance-based standard setting procedure.
- Statistical relationship studies to project preparedness indicators onto the NAEP scale by relating NAEP to performance on other assessments.
- Survey studies to collect data regarding cut scores on other assessments used for college admissions and job training placement decisions.

Later, the Governing Board added a fifth type of study – benchmarking – in which 12th grade NAEP is administered to a student reference group of interest, e.g., freshmen college students, military recruits, or job trainees. NAEP scores are then mapped onto relevant preparedness criteria.

In 2009, the Governing Board approved four categories of research to support setting and reporting 12th grade preparedness levels. The first group of studies involved analyses of the alignment of NAEP content to the content covered by other assessments (e.g., SAT, AccuPlacer, and WorkKeys). Judgmental standard setting studies, designed and conducted for work preparedness, comprise the second group of studies. Five exemplar jobs were selected by the Governing Board for inclusion in these studies. They are: (a) Automotive Master Mechanic; (b) Computer Support Specialist; (c) Heating, Ventilation, and Air Conditioning Technician (HVAC); (d) Licensed Practical Nurse (LPN); and (e) Pharmacy Technician. Statistical linkage studies comprise the third group of studies, while surveys of two- and four-year colleges, covering admissions and placement practices, make up the fourth group. A fifth group of studies pertaining to benchmarking 12th grade NAEP scores against reference groups of interest (e.g.,

college students, military recruits, and civilian job trainees) was also approved by the Governing Board.

Over the past several years, more than 30 12th grade NAEP preparedness studies have been completed by the Governing Board. Central to the validity of reporting 12th grade preparedness for reading and mathematics is confirmation that the assessments actually measure the knowledge, skills, and abilities for students to be academically prepared for college course work and/or for entry into job training programs. Content alignment studies for 12th grade NAEP reading and mathematics assessments with college admissions tests (i.e., SAT, ACT) found similar content in NAEP and the college admission examinations SAT and ACT, and somewhat less with the ACCUPLACER.

On the other hand, alignment between 12th grade NAEP reading and mathematics and tests designed to measure knowledge, skills, and abilities supposedly related to the workplace (e.g., WorkKeys) have led to less clear results. Subject matter experts participating in judgmental standard setting studies reported that NAEP reading and mathematics items were less relevant for predicting success in job training programs than in entry level college courses (Loomis, 2012).

At its November-December 2012 meeting, the Governing Board discussed plans for additional research to be conducted during the coming year. This would include content analysis and linking of 8th grade NAEP and 8th grade EXPLORE, an ACT developed test that is linked to the college-level ACT. Analysis of course content prerequisites for freshman college courses and job training programs, and efforts to obtain information about military job training also were on the drawing board.

After discussing the challenges of job training preparedness, the Governing Board decided to commission this project as a final effort to explore the feasibility of this line of research (Sellman, Wise, & Schultz, 2012).

This Report

HumRRO conducted a multi-method evaluation of the 12th grade NAEP reading and mathematics frameworks and item pools as a measure of preparedness for college and job training. Given that the college preparedness research has progressed further than that for job training, HumRRO focused its efforts on expanding the investigation of job training preparedness. Based on findings from research to date, we incorporated 8th grade NAEP reading and mathematics frameworks and items pools into our investigations to determine whether some 8th grade NAEP content is better suited than the 12th grade NAEP for assessing academic preparedness for job training as well as college.

We extended the prior NAEP preparedness research with two studies designed to dive more deeply into job training preparedness issues not yet clearly resolved. One study investigated the content alignment between NAEP and a job skills assessment that helps employers select, hire, train, develop, and retain a high-performance workforce (ACT WorkKeys). A second study compared the levels of knowledge, skills, and abilities (KSAs) required for relevant NAEP reading and mathematics content to the levels of KSAs required for relevant job training content, as indicated in the U.S Department of Labor's occupational information network, or O*NET.

A separate, parallel effort commissioned thoughtful guidance from leading experts in industrial-organizational (I-O) psychology and educational research. Panelists were asked to consider

completed NAEP research and develop white papers discussing the potential of the 12th grade NAEP reading and mathematics frameworks and item pools as a measure of academic preparedness for college and job training.

Content Alignment between the NAEP and WorkKeys Assessments

The first study and its results are documented in detail in Dickinson, Smith, Deatz, Thacker, Sinclair, and Johnston-Fisher (2014). HumRRO recruited 48 educators from 13 states as panelists to determine the content alignment between NAEP and WorkKeys. Panelists were divided among two workshops, conducted separately to facilitate the assessment of the reliability of results. Panelists participated in a variety of alignment tasks. This study built upon prior research on the content alignment between NAEP and WorkKeys in three major ways (ACT, 2010a; 2010b). First, this study included the Frameworks and items from the NAEP grade 8 assessments in order to address concerns raised that grade 8 Frameworks may provide a better match to the academic content expectations of job training programs (Kilpatrick, 2012; Loomis, 2012). Second, additional WorkKeys assessments (Applied Technology and Locating Information) were included in the study to determine the extent to which NAEP mathematics and informational reading content may relate to these additional WorkKeys assessments. Finally, the NAEP Frameworks were directly compared to the WorkKeys targets (standards) to determine the degree of overlap between the two content domains.

This study helped to identify (a) the extent to which NAEP assessments measure the content and cognitive complexity reflected in the WorkKeys targets, (b) the extent to which WorkKeys assessments measure the content and cognitive complexity reflected in the NAEP Mathematics Framework and the informational component of the NAEP Reading Framework, and (c) the amount of overlap between the NAEP Frameworks for mathematics and informational reading and the WorkKeys targets for Applied Mathematics, Applied Technology, Locating Information, and Reading for Information. Specifically, the following comparisons were made:

- NAEP Mathematics Framework and WorkKeys targets for Applied Mathematics
- NAEP Mathematics Framework and WorkKeys targets for Applied Technology
- NAEP Reading Framework (Informational component only) and WorkKeys targets for Reading for Information
- NAEP Reading Framework (Informational component only) and WorkKeys targets for Locating Information

Several key results do not support the use of NAEP for determinations related to the academic preparedness of U.S. 12th grade students for entry into job training:

- NAEP items do not adequately represent the WorkKeys content domain, as evidenced by the percentages of WorkKeys' mathematics and reading targets (52% and 72%, respectively) that were not matched to any NAEP item.
- Sixteen of the 24 content strands within the NAEP Math Framework and one of the three cognitive targets within the NAEP Reading Framework were not matched to any WorkKeys item.

- A direct comparison of the content frameworks for the two assessments indicated that the majority of the elements of the NAEP Math Framework, WorkKeys math targets, and WorkKeys applied technology targets reflected unique content. Unique mathematics elements were calculated for grade 12 NAEP math framework (85%), grade 8 NAEP math framework (75%), WorkKeys math targets (61%), and WorkKeys applied technology targets (100%). Unique reading elements included grade 8 and 12 informational reading framework (50%), WorkKeys reading targets (46%), and WorkKeys locating information targets (50%).

These results are not all that surprising given the differing purposes of the two assessments. While NAEP has been designed to provide evidence of what students in the United States know and can do with respect to a broad academic curriculum, WorkKeys assessments provide information about job-related skills that can be used in the selection, hiring, training, and development of employees. It should be noted that while this study found that much of the content assessed by WorkKeys and NAEP did not overlap, lack of overlap is not evaluative of either assessment. The lack of overlap in content likely reflects substantial differences in purpose and design. Both assessments may function very well for their specified purposes without exhibiting great similarity in the content they measure.

Finally, while there is some indication that there is more overlap between grade 8 NAEP items and WorkKeys targets, which is consistent with information provided in related research studies (Loomis, 2012; Sinclair, Becker, McCloy, & Thacker, 2014; Educational Policy Improvement Center, 2013), the results of this study suggest that including grade 8 NAEP content does not significantly improve the level of alignment between the NAEP and WorkKeys assessments, nor would the 8th grade NAEP assessments be an appropriate measure of academic preparedness for postsecondary job training.

Comparisons between NAEP and O*NET on Academic Preparedness for Job Training for Five Target Occupations

Details and results from the second study are documented in Sinclair, Becker, McCloy, & Thacker (2014). This study identified NAEP content (8th and 12th grade) that is relevant to training performance requirements for each of the five target occupations, and, conversely, the training performance requirements that are relevant to NAEP content. The job training content was based on performance requirements (tasks) adapted from O*NET, the U.S. Department of Labor's occupational information network. The study also compared the levels of academically-relevant KSAs required for proficiency on the job-relevant NAEP content to the levels of KSAs required for the NAEP-relevant job training content. The KSAs included in this study were a subset of academically-relevant KSAs from the O*NET covering reading and mathematical related skills (e.g., written comprehension, mathematical reasoning, critical thinking, complex problem solving, deductive reasoning, etc.)

An overview of the findings is as follows:

- The range of reading and mathematics skills required by NAEP (both grade 8 and grade 12) is broader than the range of reading and mathematics skills required by job training. This was demonstrated by the finding that considerably more content on NAEP was rated as irrelevant to job training than was job training content rated as irrelevant to NAEP.
- The NAEP reading objectives most relevant to job training content are the objectives associated with the Locate/Recall cognitive target for NAEP informational reading.

- The NAEP reading objectives that were least relevant to job training content were the objectives associated with the Critique/Evaluate cognitive target.
- The NAEP mathematics objectives most relevant to job training content were the objectives associated with the Number Properties and Operations content area and the Measurement content area (except for Computer Support Specialists). This was true for both grade 8 and grade 12 NAEP.
- The NAEP mathematics objectives that were least relevant to job training content were the objectives associated with Geometry (except for HVAC) and Algebra (except for LPNs). This was true for both grade 8 and grade 12 NAEP.
- The percentage of the NAEP mathematics objectives linked to job training requirements for specific occupations decreased considerably from grade 8 to grade 12, indicating that as the complexity of the NAEP objectives increased from grade 8 to grade 12 their relevance to job training decreased¹.
- Disconnects were found between the levels of KSAs required for proficient performance on NAEP and the levels of KSAs required for entry into job training such that higher levels of the KSAs were required in the NAEP assessments than for job training. The largest disconnects occurred between grade 12 NAEP mathematics and job training. Disconnects also occurred between grade 12 reading and job training. The disconnects in required levels of KSAs tended to be smaller when comparing grade 8 content to job training content, particularly for grade 8 reading, which demonstrated several “matches” with KSA levels for training content (most notably with Written Comprehension).

As a result of the above set of findings, HumRRO offered the following recommendations for the Governing Board’s consideration:

- Given that there is converging evidence across studies that the Number Properties and Operations content area for mathematics and the Locate/Recall cognitive target from NAEP informational reading are most relevant to job training, consider the possibility of using subscores from these content areas to report on students’ academic preparedness for job training.
- Given the greater correspondence between grade 8 content and job training content in reading and mathematics, consider the possibility of administering the grade 8 assessments to 12th grade students to make determinations about their academic preparedness for entry into job training.
- Consider the possibility of updating the working definition of job training preparedness to include trainee outcomes, such as trainee performance in job training. Actual performance in job training is at a level that is somewhat beyond “just qualified” for placement into job training. Including training outcomes in the working definition of job training preparedness might potentially lead to evidence that is more supportive of grade 12 NAEP as an indicator of job training preparedness. Furthermore, including training outcomes as elements of the working definition of job training preparedness would expand opportunities for future research investigations.

¹ A comparable statement about whether including grade 8 reading resulted in more linked content is not possible given that NAEP reading objectives are the same for grade 8 and for grade 12.

NAEP Framework Evaluation Technical Advisory Panel

The final HumRRO study involved convening an expert panel as described in McCloy and Day (2014). HumRRO assembled a 10-person technical advisory panel (TAP) to consider the research conducted to date, as well as the two studies described above, produce ideas for future work, and also to provide input on whether the Governing Board should continue to perform research on using NAEP as an indicator of academic preparedness for job-training programs. To draw on a wide range of relevant expertise we recruited 10 experts: 5 experts in the area of industrial-organizational psychology and 5 experts in educational measurement. HumRRO's experience in both these fields led us to recognize the potential of contributions from both disciplines. The members of the Technical Advisory Panel (TAP) were:

Industrial-Organizational Psychologists

- John P. Campbell, University of Minnesota;
- Michael A. Campion, Purdue University;
- Kenneth Pearlman, Kenneth Pearlman, LLC
- Ann Marie Ryan, Michigan State University
- Nancy T. Tippins, Corporate Executive Board – Valtera Corporation

Educational Researchers

- Gregory J. Cizek, University of North Carolina, Chapel Hill
- Brian Gong, Center for Assessment, National Center for the Improvement of Educational Assessment;
- Ronald K. Hambleton, University of Massachusetts, Amherst;
- Suzanne Lane, University of Pittsburgh;
- Barbara S. Plake, University of Nebraska—Lincoln, retired, private consultant;

Panelists attended a brainstorming session in October 2013. In this session panelists were briefed on the Governing Board's preparedness research conducted to date and engaged in discussions of possible future studies to further the preparedness research agenda. After the meeting, panelists submitted proposals for white papers. Governing Board staff reviewed these proposals and commissioned three papers. The commissioned authors presented their papers for discussion at a second meeting, the TAP Symposium, in August 2014.

During this symposium, many challenges were discussed with regard to conducting research on academic preparedness for job training. For example, the variability in training programs across occupations—and even the variability across training programs within the same occupation—makes it exceedingly difficult to determine a reference point on the NAEP scale that signifies academic preparedness for job training. Even if arrangements could be made to administer a special NAEP test, it would likely be difficult to obtain participation from organizations to support this special administration. It

also would be challenging to obtain scores from the NAEP respondents on some validated measure of job training. Challenges such as these, coupled with research findings that have provided converging evidence that NAEP measures reading and math content that is larger and broader than the reading and math KSAs required for entry into job training, call into question whether the Governing Board should continue to conduct research on using NAEP as a measure of academic preparedness for job-training programs.

If, however, the Governing Board decides to move forward with preparedness research for job training, the next agenda items to consider, based on the TAP symposium, would be the following:

1. Build upon the idea to use NAEP to measure student job readiness

Dr. Plake suggested a modification of prior expert judgment research—namely, asking job training experts to evaluate the importance for training of items mapped onto the NAEP Achievement Levels as a group, rather than simply rating discrete items as was done in a previous standard setting study (WestEd & Measured Progress, 2011; 2012). The items might be organized by content strand, so that importance might be rated differentially by strand. The general idea is to get experts to respond to more general descriptions of academic preparedness, rather than to individual items.

2. Collect information on additional aspects of preparedness through student (and possibly also teacher) questionnaires

Dr. Ryan suggested that further research on academic preparedness for job training might employ measures of the grit construct as potential moderators.

3. Analyze/conduct longitudinal research relating academic performance at the end of high school to subsequent career success

The National Center for Education Statistics (NCES) has conducted several research studies following students from high school into post-high school activities. The National Longitudinal Surveys conducted by the Department of Labor contain similar data. The TAP discussed how results from prior analyses of these data, together with new analyses, could inform the relation between academic skills in high school and persistence and success in different career fields.

4. Consider studies linking employment tests to the NAEP scale

Dr. Tippins suggested studies to link employment tests to the NAEP scale. The approach would likely entail the following: (a) recruit a few large companies that provide entry-level jobs not requiring prior job knowledge or experience, (b) ask them to administer a number of NAEP items along with commonly used employment tests so that the employment test and NAEP scales can be linked, (c) collect information on key decision points,

and (d) summarize findings on the relation between the employment test score levels and subsequent success².

5. Embed job training preparedness items into NAEP

Dr. Campion recommended embedding items from a measure of job training preparedness (e.g., WorkKeys) into the NAEP assessment. This could offer an alternative to trying to get job trainees and/or individuals in associate degree programs to complete a special administration of NAEP.

Summary of HumRRO Investigations

In summary, findings from both HumRRO studies—the content alignment study between NAEP and WorkKeys and the study comparing NAEP content and O*NET content—call into question the validity of inferences that can be made about the preparedness of U.S. 12th grade students for entry into job training using NAEP results. Moreover, during the TAP symposium many challenges were discussed with regard to conducting research on academic preparedness for job training such as the variability in training programs across occupations and even the variability across training programs within the *same* occupation. Challenges such as these, coupled with research findings that have provided converging evidence that NAEP measures reading and math content that is broader and at a higher level than the reading and math KSAs required for entry into job training, call into question whether the Governing Board should continue to conduct research on using NAEP as an indicator of academic preparedness for entry into job training.

If, however, the Governing Board decides to move forward with research on preparedness for job training, several avenues for future research were suggested. In summary, based on findings from the studies summarized in this paper and from the TAP symposium, the following have been offered as next agenda items for consideration:

- Consider the possibility of using subscores from the Number Properties and Operations content area from mathematics and the Locate/Recall cognitive target from informational reading to report on students' academic preparedness for job training.
- Consider the possibility of administering the grade 8 assessments to 12th grade students to make determinations about their academic preparedness for entry into job training³.
- Modify the previous standard setting study (WestEd & Measured Progress, 2011; 2012) such that job training experts evaluate the importance for training of items mapped onto the NAEP achievement levels as a group, rather than simply rating discrete items as was done in the aforementioned standard setting study. The items might be organized

² Note that implementing this recommendation would require expanding the current working definition of job training preparedness to include outcome measures.

³ It should be noted that evidence from the O*NET study was more supportive of this recommendation than the evidence from the alignment study with WorkKeys.

by content strand, so that importance might be rated differentially by strand. The general idea is to get experts to respond to more general descriptions of academic preparedness, rather than to individual items.

- Continue explorations already underway by NAEP to use contextual questions to develop a measure of the grit construct as an indicator of preparedness for success in both college and job training.
- Analyze/conduct longitudinal research relating academic proficiencies at the end of high school to subsequent career success using, for example, data from the National Longitudinal Surveys conducted by the Department of Labor. Results from prior analyses of these data, together with new analyses, could inform the relation between academic skills in high school and persistence and success in different career fields.
- Link employment tests to the NAEP scale. This approach would entail recruiting a few large companies that provide entry-level jobs not requiring prior job knowledge or experience, asking them to administer a number of NAEP items along with commonly used employment tests so that the employment test and NAEP scales could be linked, collecting information on key decision points, and summarizing findings on the relation between the employment test score levels and subsequent success.
- Embed items from a measure of job training preparedness (e.g., WorkKeys) into the NAEP assessment. This could offer an alternative to trying to get job trainees and/or individuals in associate degree programs to complete a special administration of NAEP.
- Consider the possibility of updating the working definition of job training preparedness to include trainee outcomes, such as trainee performance in job training. Actual performance in job training is at a level that is somewhat beyond “just qualified” for placement into job training. Including training outcomes in the working definition of job training preparedness might potentially lead to evidence that is more supportive of grade 12 NAEP as an indicator of job training preparedness. Furthermore, including training outcomes as elements of the working definition of job training preparedness would expand opportunities for future research investigations.

We believe that the above topics are likely to yield the most fruitful avenues for future research on the use of NAEP as an indicator of academic preparedness for job training. However, these recommendations for future research should be tempered with the caveats described in the three HumRRO reports (Dickinson et al., 2014; McCloy & Day, 2014; Sinclair et al., 2014), and summarized in this final report—namely, that there is mounting evidence indicating that 12th grade NAEP may not be a suitable indicator of academic preparedness for job training.

References

- ACT, Inc. (2010a). The alignment of the NAEP grade 12 mathematics assessment and the WorkKeys applied mathematics assessment. Iowa City, IA: Author.
- ACT, Inc. (2010b). The alignment of the NAEP grade 12 reading assessment and the WorkKeys reading for information assessment. Iowa City, IA: Author.
- Dickinson, E. R., Smith, E. A., Deatz, R. C., Thacker, A. A., Sinclair, A. L., & Johnston-Fisher, J. (2014). The content alignment between the NAEP and WorkKeys assessments (2014 No. 054). Alexandria, VA: Human Resources Research Organization.
- Educational Policy Improvement Center (2013). National Assessment of Educational Progress Grade 12 Preparedness Research Project Job Training Programs Curriculum Study. Eugene, OR: Author.
- Kilpatrick, J. (2012, April). The standard for minimal academic preparedness in mathematics to enter a job-training program. Paper presented in the *Setting Academic Preparedness Standards for Job Training Programs: Are We Prepared?* symposium at the annual meetings of the National Council on Measurement in Education, April 14, 2012, Vancouver, British Columbia, Canada.
- Loomis, S.C. (2012, April). A Study of “Irrelevant” Items: Impact on Bookmark Placement and Implications for College and Career Readiness. Paper presented in symposium, Setting academic preparedness standards for job training programs: Are we prepared? at the annual conference of the National Council on Measurement in Education, Vancouver, British Columbia, Canada.
- McCloy, R. A. and Day, T. C. (2014). NAEP technical advisory panel symposium proceedings (2014 No. 062). Alexandria, VA: Human Resources Research Organization.
- Sellman, W.S., Wise, L.L., Schultz, S.R., & Schantz, L.B. (2003, March). Notes from the March 2003 meeting of the National Assessment Governing Board, March 6-8, 2003 (SR-03-15). Paper prepared for the National Center for Education Statistics. Alexandria, VA: Human Resources Research Organization.
- Sellman, W.S., Wise, L.L., & Schultz, S.R. (2012, December). Notes from the November/December 2012 meeting of the National Assessment Governing Board, November 29-December 1, 2012. Paper prepared for the National Center for Education Statistics. Alexandria, VA: Human Resources Research Organization.
- Sinclair, A. L., Becker, D. E., McCloy, R. A., & Thacker, A. A. (2014). Linkage between O*NET and NAEP grade 8 and NAEP grade 12 reading and mathematics (2014 No. 012). Alexandria, VA: Human Resources Research Organization.

U.S. Department of Education, National Commission on NAEP 12th Grade Assessment and Reporting. (2004). Twelfth grade student achievement in America: A new vision for NAEP. Washington, DC: National Assessment Governing Board.

U.S. Department of Education, Technical Panel on 12th Grade Preparedness Research. (2009). Making new links: 12th grade and beyond. Washington, DC: National Assessment Governing Board.

WestEd & Measured Progress. (2011). National Assessment of Educational Progress Grade 12 preparedness research project judgmental standard setting (JSS) studies: Process report. San Francisco, CA.

WestEd & Measured Progress. (2012). National Assessment of Educational Progress Grade 12 Preparedness Research Project Judgmental Standard Setting (JSS) Studies: Technical Report. San Francisco, CA.

COLLEGE COURSE CONTENT ANALYSIS

Project Status Update Contract ED-NAG- 12C-0003

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)⁴ 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 project is now complete (see May 2014 COSDAM materials for Executive Summary). The final report is now available on the Governing Board’s website at:

http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/judgmental-standard-setting-studies/College_Course_Content_Analysis.pdf.

4 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).

OVERVIEW OF REFERENCED ASSESSMENTS

For additional background information, the following list presents a brief description of the assessments referenced in the phase two 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, Mathematics, and Science 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 (0.75) or “B” or higher (0.50) 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 and mathematics knowledge and skills and probabilities of a college course grade of “C” or higher (0.75) or “B” or higher (0.50) by the time students graduate high school with particular score ranges on the Explore assessment.
- 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 0.65 probability of earning a first-year college grade-point average of 2.67 (B-) or better.

Evaluation of NAEP Achievement Levels

Objective To receive a brief informational update on the current status of the independent evaluation of NAEP achievement levels that is being performed by the National Center for Education Evaluation and Regional Assistance (NCEE), part of the Institute for Education Sciences (IES). 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.

Evaluation of NAEP Achievement Levels Contract

The National Center for Education Evaluation and Regional Assistance (NCEE), part of the Institute for Education Sciences (IES), will administer the Evaluation of the NAEP Achievement Levels. On September 29, 2014, NCEE awarded a contract to The National Academy of Sciences to perform this work.

Objectives for the evaluation include the following:

- Determine how "reasonable, valid, reliable and informative to the public" will be operationalized in this study.
- Identify the kinds of objective data and research findings that will be examined.
- Review and analyze extant information related to the study's purpose.
- Gather other objective information from relevant experts and stakeholders, without creating burden for the public through new, large-scale data collection.
- Organize, summarize, and present the findings from the evaluation in a written report, including a summary that is accessible for nontechnical audiences, discussing the strengths/ weaknesses and gaps in knowledge in relation to the evaluation criteria.
- Provide, prior to release of the study report, for an independent external review of that report for comprehensiveness, objectivity, and freedom from bias.
- If the optional tasks are authorized by ED, plan and conduct dissemination events to communicate the conclusions of the final report to different audiences of stakeholders.

Design:

This study will focus on the achievement levels used in reporting NAEP results for the reading and mathematics assessments in grades 4, 8, and 12. Specifically, the study will review developments over the past decade in the ways achievement levels for NAEP are set and used and will evaluate whether the resulting achievement levels are "reasonable, valid, reliable, and informative to the public." The study will rely on an independent committee of experts with a broad range of expertise related to assessment, statistics, social science, and education policy. The project will receive oversight from the Board on Testing and Assessment (BOTA) and the Committee on National Statistics (CNSTAT) of the National Research Council.

Members of the interdisciplinary review committee are expected to be selected by early 2015, and the committee is expected to meet over the course of 2015. The report from the evaluation is expected to be released in 2016 and will be announced on <http://ies.ed.gov/ncee/>.

February 2015 Update:

The list of Committee members was released for public comment in early January 2015:

Name	Affiliation
Dr. Christopher F. Edley, Jr. (Chair)	University of California, Berkeley
Dr. Peter Afflerbach	University of Maryland, College Park
Dr. Sybilla Beckmann	University of Georgia
Dr. H. Russell Bernard	University of Florida
Dr. Karla Egan	National Center for the Improvement of Educational Assessment
Dr. David J. Francis	University of Houston
Dr. Margaret E. Goertz	University of Pennsylvania
Dr. Laura Hamilton	The RAND Corporation
Dr. Brian W. Junker	Carnegie Mellon University
Dr. Suzanne Lane	University of Pittsburgh
Ms. Sharon J. Lewis	Retired
Dr. Bernard L. Madison	University of Arkansas
Dr. Scott Norton	Council of Chief State School Officers
Dr. Sharon Vaughn	The University of Texas at Austin
Dr. Laurrell L. Wise	HumRRO

Additional information about the Committee is available at:

<http://www8.nationalacademies.org/cp/CommitteeView.aspx?key=49677>. The first Committee meeting will take place in Washington, DC on February 19-20, 2015. Governing Board staff will attend the open sessions and will make a presentation to the Committee.