National Assessment Governing Board
Committee on Standards, Design and Methodology

May 17, 2013
10:00 am – 12:30 pm

AGENDA

Joint Session with Reporting and Dissemination Committee *Hershey/Crocker Room, 2nd Floor*		
10:00 – 11:00am	Welcome and Introductions *Lou Fabrizio, COSDAM Chair* NAEP Testing and Reporting of Students with Disabilities and English Language Learners *Grady Wilburn, NCES*	Attachment A
COSDAM Meeting		
11:00 – 11:35 am	Preliminary Discussion on Setting Achievement Levels on the 2014 NAEP Technology and Engineering Literacy (TEL) Assessment *COSDAM Members*	Attachment B
11:35 – 11:40 am	Update on Evaluation of NAEP Achievement Levels Procurement *Peggy Carr, Associate Commissioner, NCES*	Attachment C
11:40 am – 12:10 pm	NAEP 12th Grade Academic Preparedness Research ▪ NCES Research on Academic Preparedness *William Ward, NCES* ▪ Committee Questions on the Phase 2 Research Updates (see Information Item) *Cornelia Orr, Executive Director, NAGB* *Michelle Blair, Senior Research Associate, NAGB*	Attachment D
12:10 – 12:25 pm	Interpreting NAEP Proficient using Preparedness Research Findings *Lou Fabrizio, COSDAM Chair*	Attachment E *(Will be sent separately)*
12:25 – 12:30 pm	Other Issues or Questions *COSDAM Members*	
Information Item	NAEP 12th Grade Academic Preparedness Research: Phase 2 Research Updates ▪ Course Content Analysis Research ▪ National and State Partnerships ▪ Research with Frameworks	Attachment F

Interpreting NAEP Proficient Using Preparedness Research Findings

For almost a decade, the Governing Board has been thoughtfully and deliberately working to determine the feasibility of NAEP reporting on the academic preparedness of 12th grade students for college and job training. Accordingly, the Governing Board is conducting a comprehensive program of preparedness research. The first phase of the research involved more than 30 studies in 5 areas: content alignment, statistical relationship, standard-setting, benchmarking, and a survey of higher education.

On the basis of the research results, Governing Board staff propose the following inferences for use in reporting NAEP 12th grade results:

> 12th grade students scoring at or above the Proficient achievement level on the 12th grade NAEP Reading or Mathematics Assessment are
>
> - likely to be academically prepared for first year college courses,
> - likely to have a first-year college GPA of B- or better, and
> - not likely to need remedial/developmental courses in reading or mathematics in college.

On the following pages is the draft validity argument in support of these inferences. With the approval of the Committee on Standards, Design and Methodology (COSDAM), the validity argument is based on a model described by Michael Kane, a renowned psychometrician widely recognized for his theoretical work in validity. In summary, the model begins with a score interpretation (i.e., the inferences above), a statement of the propositions or assumptions that underlie that score interpretation, and the presentation of evidence to evaluate those propositions or assumptions. Recognizing that validation is a continuing process, and that validity cannot be established absolutely, Kane's model provides that the criterion that must be met is the **plausibility** of the validity argument. Michael Kane serves as an advisor in the development of the validity argument and has reviewed this draft.

At the May 2013 Governing Board meeting, the draft validity argument will be discussed by the Executive Committee, COSDAM, and the full Board. The purpose is two-fold: to review the Phase I preparedness research and results, and to provide feedback to staff on the draft validity argument in relation to the requirements of the Kane model. Thus, the feedback should address the following questions—are the propositions/assumptions optimally framed, does the evidence appropriately address the propositions/assumptions, are there propositions/assumptions that should be added, and taken as a whole, does the argument seem to meet the plausibility criterion?

Following the May 2013 Board meeting, the draft validity argument will be revised per the feedback provided. The draft then will be subjected to independent external review by technical experts. Two noted psychometricians, Mark Reckase and Gregory Cizek have agreed to review the draft validity argument. The intention is to include their reviews as a part of the final validity argument.

The final validity argument will be presented at the August 2013 meeting for action by the full Board with respect to its use in reporting 12th grade reading and mathematics results.

Finally, please note that the appendices mentioned in the draft validity argument are not included here, but are available upon request.

Update on Evaluation of NAEP Achievement Levels Procurement

Objective To receive a brief update from NCES on the current status of the procurement being planned to evaluate NAEP achievement levels.

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, NCES has been planning a procurement to administer an evaluation of NAEP achievement levels. The last update COSDAM reviewed on this topic was a year ago.

In this brief update, NCES Associate Commissioner Peggy Carr will provide the Committee with a summary of the status of this procurement.

Setting Achievement Levels on the NAEP 2014 Technology and Engineering Literacy (TEL) Assessment

Status: Information and discussion

Objective: To discuss issues that should be addressed in addition to those identified in the recently developed "issues paper," which discusses achievement level setting for TEL.

Attachments: B-1 Issues Paper on achievement level setting for the NAEP TEL assessment

B-2 Abridged NAEP Framework for Technology and Engineering Literacy (National Assessment Governing Board, 2012)

Background

At the March 1, 2013 meeting, the Committee began discussion on setting achievement levels for the 2014 NAEP TEL assessment. For the May 17, 2013 meeting, an issues paper has been developed to support procurement and project planning for developing recommended achievement levels for TEL. Contract award is scheduled for late 2013.

The outcome of both the March 2013 and May 2013 COSDAM discussions will inform the statement of work for the TEL level-setting procurement.

Timeline

The following timeline provides a preliminary list of key dates and activities related to TEL assessment development and achievement level setting.

Date	Activity	Responsibility
2008 - 2010	TEL Framework development	ADC, Board, WestEd (contractor)
2010 - 2012	Assessment development for 2013 pilot test	NCES, NAEP contractors
2010 - 2012	Item review for 2013 pilot test	NCES, NAEP contractors, TEL Standing Committee, ADC
Early 2013	Pilot test – national sample, grade 8	NCES, NAEP contractors
May 2013	TEL ALS issues paper	COSDAM, consultant
Late 2013	ALS procurement and contract award	Board staff, COSDAM
Early 2014	Operational administration – national sample, grade 8	NCES, NAEP contractors
2014 - 2015	Final phase of ALS process and Board action on TEL	COSDAM, ALS contractor, Board
2015	Reporting TEL results	Board, NCES, contractors

TEL Assessment Design

The 2014 Technology and Engineering Literacy (TEL) assessment is based on the Board-adopted Framework and Specifications (see Abridged TEL Framework in Attachment D-2; complete documents are at www.nagb.org, Publications).

The TEL assessment is composed of three major areas:

- Design and Systems
- Information and Communication Technology
- Technology and Society

Another key dimension of the TEL assessment is the three practices, each of which is applicable to the three major areas noted above:

- Understanding Technological Principles
- Developing Solutions and Achieving Goals
- Communicating and Collaborating

The TEL assessment was developed using an evidence-centered design (ECD) approach (see Attachment D-2). From the beginning, all TEL tasks and items were designed using an evidential chain of reasoning that links what is to be measured, the evidence used to make inferences, and the tasks used to collect the desired evidence. In addition to student responses

to complex tasks and discrete items, the computer-based TEL assessment allows NAEP to capture a wide array of data on student performance. For example, NAEP will collect information on how students interact with the TEL simulations and experiments. Such data may include the number of experimental trials run and the number and types of variables controlled. These observable data on "strategies and processes" may also contribute to the scoring of student performance.

TEL Reporting

Based on the ECD approach, TEL reporting will be expanded beyond the traditional NAEP scores. It is expected that data from the complex performance tasks and discrete items will be reported in a number of ways:

- A composite scale score on which the achievement levels will be set
- Subscores for the content areas (Design and Systems; Information Communication Technology; Technology and Society)
- Reporting on the practices (Understanding Technological Principles; Developing Solutions and Achieving Goals; Communicating and Collaborating)
- Information on students' processes and strategies, related to the ECD model, captured as observable data from their work on the TEL scenario-based tasks.

Ongoing Potential Discussion Questions for COSDAM

- Given the emerging field of setting achievement levels on ECD-based complex performance assessments, what additional background materials are needed to inform the COSDAM/Board decision on an appropriate method for ALS on the TEL assessment?

- To what extent should research studies be built into the TEL ALS project?

- Are there examples of ALS exploratory work the Board should undertake in collaboration with the two assessment consortia? (The two assessment consortia are also planning to set performance standards on complex ECD-based tests.)

Mid-Atlantic Psychometric Services

212 Ashton Dr. SW
Leesburg Virginia 20175-2527

703.771.4664 (Voice)
703.771.4686 (FAX)
mlbourque@verizon.net

Issues Paper:
Setting Achievement Levels on the 2014 NAEP TEL Assessment

April 29, 2013

Developed for the National Assessment Governing Board under contract number ED-NAG-13-R-0002 by Mid Atlantic Psychometric Services

Table of Contents

Issues Paper:
Setting Achievement Levels on the
2014 NAEP TEL Assessment

Introduction

In preparation for the development of the achievement levels for the 2014 NAEP Technology and Engineering Literacy (TEL) computer-based assessment at grade 8, the Governing Board wishes to identify those technical and policy issues which could have a substantive impact on the process of setting the achievement levels. Since this assessment is not only a computer-based test (CBT), but has been developed using evidence-centered design (ECD), the Board is particularly interested in having the various standard-setting elements reviewed, as well as gaining insights into other extant standard setting methodologies that are linked to ECD. This White Paper will articulate the various issues where further understanding and research may be helpful in achieving the long-standing NAEP standard-setting goals successfully and efficiently.

This is a new paradigm for NAEP. The TEL assessment is radically different from the NAEP legacy assessments.[1] This paper will assume that no introduction is needed to either the ECD or CBT concepts. Selected references, however, are included for both topics for any reader wishing to pursue ECD and/or CBT in greater detail (cf. Hendrickson, Huff, and Leucht, 2010; Mislevy, Almond, and Lukas, 2003; Parshall, Spray, Kalohn, and Davey, 2002).

[1] Legacy assessment is the term the author uses to refer to earlier NAEP assessments that have not used the ECD design ***and*** are not computer-based testing (CBT). Some earlier NAEP assessments have moved from a paper-and-pencil administration to a computer-administered assessment, e.g., 2011 NAEP writing. However, these are still considered legacy assessments by our definition.

NAEP Standard Setting

NAEP has conducted standard setting on the National Assessment for well over 20 years now. During that period of time great strides have been made in standard-setting methodology and reporting NAEP results. In fact, NAEP methods have been modeled in local, state and federal legislation. NAEP has adjusted its approach as the requirements of different subject areas have demanded, and as greater knowledge and research have been brought to bear on the entire standard-setting initiative.

However, NAEP once more is at a crossroads with the initiation of ECD in the 2014 TEL assessment. How does ECD impact the scoring, scaling, and analysis procedures used in NAEP? How does NAGB adjust its approach to standard setting in this new ECD environment? What policy decisions need to be made to ensure reliability, validity, and usefulness of the NAEP results? How does NAGB design a standard-setting process that is clear and concise for panelists, reasonable to explain to the public, and straightforward to use? This paper will outline some of the salient questions the National Assessment Governing Board needs to consider as it moves into the virtually uncharted waters of ECD and standard setting.

Elements of Standard Setting

There is a consensus in the literature that standard setting is, by and large, a judgmental process which includes some technical aspects such as psychometrics and statistics (AERA, APA, and NCME, 1999, p.54). There is no one right answer. Whether we are dealing with clean water standards, agricultural standards, or student performance standards, the standards are usually a matter of judgment determined ultimately by the legally responsible agency. In K-12 education, up to now, it has been an activity that usually follows *after* test development and

administration. That is because most methods depend either on the performance of examinees in the assessment (examinee-centered methods), or on the nature of the assessment itself (test-centered methods)[2].

Hambleton and Pitoniak (2006) outline the nine typical steps in setting performance standards in legacy assessments[3]. The author will use these steps as a basis for raising the issues involved when setting standards in an ECD/CBT environment.

Step 1: Selecting a Standard-Setting Method

In preparing this paper, a search of the literature was conducted for methodologies in setting standards within the ECD environment, especially outside of K-12 education. Unfortunately, that search yielded few results, in part because the application discussed was highly specialized and only remotely generalizable to a K-12 setting (Behrens, Mislevy, Bauer, Williamson, and Levy, 2009). In the medical field, the on-line literature focused not on what knowledge or skills practitioners should possess, but rather on what hospitals/clinics should do when providing services (BMC, 2005; Kak, Burkhalter, and Cooper, 2001).

That being said, there is standard setting being done in an ECD environment using some traditional approaches or variations thereof that have been around for many years. For example, when queried, Hambleton acknowledged that as far as he was concerned the most important element of standard setting was the development of the achievement levels

[2] This dichotomy has been expanded to subsume ratings of examinees (not just their responses) for example, contrasting groups, the review of score profiles (Jaeger, 1995), and several compromise methods as described by Pitoniak and Hambleton (2006).

[3] Due to its importance, Hambleton and Pitoniak separate collecting panelists' evaluations during the standard-setting process from the other forms of validity evidence that are typically collected, including other forms of procedural evidence, as well as internal and external validity data (nine step schema). However, since ECD views the entire enterprise as gathering validity data along a continuum from test inception to test reporting, the author combined all forms of validity under one umbrella calling it simply "validity evidence" (the eight step schema discussed in this paper).

descriptions (ALDs). After that, probably any method could be used successfully, adapting it, in this case, for the ECD environment.[4]

However, some work is proceeding currently to bridge the gap between old and new ways at Pearson in Austin TX and in Iowa City IA. Beimers, Way, McClarty, and Miles (2012) and McClarty, Way, Porter, Beimers, and Miles (2013) have recently published papers on evidence-based standard setting (which the authors abbreviate as EBSS), as a way of establishing validity evidence for cut scores. Their argument goes all the way back to the early days of NAEP when weaknesses in the NAGB approach were highlighted by various NAEP evaluations. This article links the judgment processes typically employed in standard setting with systematic research data provided to the panelists during the panel meetings. In other words, the standard-setting activity extends the trail of evidence from the elements of ECD (claims, evidence, tasks) up to and including the cut scores.

This approach would obviously require time and resources to collect the research data (not all of which needs to have its origins in NAEP), but which would need to be prepared in a format understandable to panelists.

Step 2: Selecting Standard-Setting Panels

We want to examine now the composition of the panel of subject matter experts (SMEs) for the achievement levels work. The TEL assessment is much like the science assessment in that there is more than one area of subject-specific expertise needed in the mix of participants. There is also cross-over expertise needed, for example individuals who are subject matter experts in two or more areas such as engineering and also information and communication

[4] R.K. Hambleton (personal communication, April 8, 2013)

technology. An additional challenge for selecting SMEs will be the cross-curricular nature of TEL, the content of which could be covered in a range of courses including English, science, U.S. history, engineering, among others. This broad-based expertise is important as the assessment moves forward to craft the initial ALDs based on claims and evidence, and continues the iterative process up to the crafting of the final ALDs. It is also recommended that the facilitator should have broad expertise, or that more than one facilitator be used such that subject-specific content is properly handled during the achievement levels process.

NAGB Achievement Levels Policy guideline #2 speaks to the issue of panel composition, but focuses mostly on securing a "…broadly representative body of teachers, other educators, . . . and non-educators including parents, . . . and specialists in the particular content area. (NAGB, 1995, p.5)" Special attention for the TEL assessment should be paid to just who the specialists are and how well they may fill the needs of the panel and accomplish the panel's work. While demographic background is important from a policy perspective, the skills and content expertise of each and every participant is the primary consideration.

Step 3: Developing Achievement Levels Descriptions (ALDs)

Developing descriptions of the performance categories, (referred to as achievement levels descriptions in the NAEP context), has always been a central element of the standard-setting process. In the ECD environment, this seems to be the most critical step in the process and the one that flows most directly from how ECD was used in developing the TEL assessment. In the past, ALDs were developed as a way of operationalizing the NAGB policy definitions. The ALDs were developed by subject matter experts (SMEs) prior to standard setting, employing the policy definitions, the NAEP assessment framework, test and item specifications, and their own

professional judgment about what students at the Basic, Proficient, and Advanced levels should know and be able to do in a specific content area and at a particular grade[5]. However, the journal, *Applied Measurement in Education,* published a special issue in 2010 on Evidence-Centered Assessment Design in Practice. The article by Plake, Huff, and Reshetar (2010) focused exclusively on ECD and achievement levels descriptors. They too develop ALDs prior to standard setting but use the elements of ECD to do so.

In the ECD environment a test developer will usually articulate a set of knowledge, skills, and abilities (KSAs) that are of measurement interest sometimes called ***claims[6],*** and the subsequent ***evidence[7]*** related to those claims that will be taken as supporting data for the examinee's knowledge of such claims. The ALDs flow directly from these ***claims-evidence pairs***, as found in the TEL framework. Task models for the assessment then flow directly from these pairs as well (Huff, Steinberg, and Matts, 2010).

Through an iterative process Plake et al (2010) judgmentally mapped the claims-evidence pairs to the performance continuum until a full spectrum of claims-evidence pairs was found to be sufficient for reporting examinee performance in all regions of the continuum. Contrary to the legacy NAEP assessments, where the ALDs do not cover necessarily all specific aspects of the assessment, in the ECD environment, the focus was on ensuring that all aspects of examinee performance could be reported on. In some cases that meant going back to the claims-evidence pairs and selecting additional pairs for inclusion.

[5] There are preliminary ALDs crafted during the framework development process (Appendix G). However, these are more appropriately viewed as "working" descriptions, but not the initial ALDs that would be the inputs for the training of SMEs during the standard-setting process. There is no documentation that the preliminary descriptions flowed from the claims, evidence, and student models that are integral to ECD.

[6] In the NAEP TEL documentation this component is identified as the Student Model.

[7] In the NAEP TEL documentation this component is called the Evidence Model.

Note that at this point the student, evidence, and task models cover the full performance continuum. Additionally, in developing the achievement levels, Plake et al. (2010) demonstrates how the components of ECD can be leveraged to produce ALDs that are related to not only the KSAs on the assessment, but to score interpretation and reporting and in the process provide ongoing validity evidence.

A few issues arose during the course of that work that also could become issues for NAEP: (1) a rather large number of claims at each of the performance levels (in the Advanced Placement (AP) science context), they were working with several subject-specific science areas and with score levels labeled as 3, 4, and 5); and (2) the lack of specific content expertise (within the panels) across all subject-specific science areas for developing generalized discipline-specific ALDs[8]. They addressed the first issue by informal selected sampling of claims. The second issue was resolved in part by the expertise of the workshop facilitator who was skilled across disciplines. However, ensuring that kind of expertise within the SME group may also be an acceptable solution as well.

One issue not yet mentioned is that of "what students ***should*** know and be able to do," versus "what students ***do*** know and are able to do." The author believes that in the ECD environment there is a shift: there *is* a claim, there *is* evidence, and therefore, students *do know* and *are able to do*. If that is the case, and the Board agrees, then NAGB policy definitions would need to be adjusted to reflect this new approach. On the other hand, an argument could still be made for the fact that standards are expectations and, therefore, the "should"

[8] Plake et al. makes the distinction between subject-specific ALDs, that is, ALDs that focus on a specific subject area within the natural sciences, e.g., chemistry, biology, physics versus discipline-specific ALDs, that is, ALDs which focus on the areas common across all the natural sciences, e.g., measurement, observation, hypothesizing.

terminology is still appropriate. In other words, claims and evidence are what NAGB expects of examinees, and the Nation's Report Card reports that performance.

Step 4: Training Panelists in Standard-Setting Methodology

The McClarty et al. (2013) paper describes in one section an implementation procedure that could be used in evidence-based standard setting. These include: (1) identifying the intended interpretation of the assessment results; (2) assembling research, data collection, and analysis plans; (3) synthesizing the results of step (2) in a way that is clear, focused, and readily understandable to standard-setting panelists; (4) implementing the standard-setting activity; and (5) continuing to gather data that supports the validity argument for the standards.

Step (1) is already well underway for NAGB since the intended interpretations for the grade 8 TEL assessment are the policy definitions, further operationalized by the claims-evidence pairs. But how good is good enough for *Proficient*? For solid academic performance? For demonstrated competency over challenging subject matter, including knowledge . . . application . . . and analytical skills? What claims-evidence pairs provide substantiation for these claims?

Steps (2) and (3) would be somewhat more challenging for the standard-setting contractor (to be selected by the Board) since the data are apt to be scattered across a number of possible sources. Appendix C (listing domestic source documents); Appendix D (listing international source documents); and Appendix E (listing professional association source documents) of the NAEP TEL Framework identify a number of sources that have been used in developing the framework (NAGB, n.d.), and should be reviewed and updated for EBSS.

At Step (4) the primary concern would be sufficient time and resources for the standard-setting contractor to prepare all the documentation necessary to implement the procedures smoothly and effectively. This is no small task, so sufficient lead time is critical. Secondly, panelists need to be willing and able to spend sufficient time to prepare for this kind of meeting. It would be unacceptable for participants to plan on reading the briefing book(s) on the plane ride to the standard-setting location. Commitments would need to secured well ahead from all participants that they are willing to do their homework. Further, it is quite possible that the time commitments could be more than a single meeting. All this needs to be thought through at the front end, not after it is too late in the process.

There are other considerations as well. For example, some thought needs to be given to computer platforms, security issues, timing issues (some panelists will be slower than others), and adjudication of disagreements (lack of consensus) during the meetings.

Step 5: Collecting Panelists' Ratings

In the McClarty et al. (2013) paper, they used a traditional standard-setting method. Working with two cut scores (not three as in NAEP), panelists reviewed the evidence and made recommendations for the placement of the cut scores on a raw score scale, over three rounds of judgments. Aggregated data was used as feedback to the group along with inter-rater agreement statistics.

It would be at this point in the process where the rating and scoring of the TEL items would become important for consideration by the panels. Panels need to know what evidence examinees are being scored on, or what enters into the examinee performance record. How this is handled for the different types of items on the assessment is quite important in order for

the panelists to be able to make valid and reliable judgments about performance. If NAEP is collecting data, such as how many times examinees correct errors of solution and this is not being reported/embedded on the NAEP scale, then panelists may or may not find it helpful to know that. In making this call, the rule of thumb should be to provide panelists with any data that will have or could have an impact on examinees' performance on the items and thus, on the panelists work at the standard-setting meeting(s). If it has an impact, tell them about it; if it does not, it is advisable to refrain from sharing this information.

Step 6: Providing Feedback to Panelists

Many different types of feedback have been used in the traditional standard-setting process including, but not limited to, panelists' discussions, p-values, cut-scores (by Round) and the associated standard deviation, rater-location data, intra-rater agreement estimates, Reckase charts[9], impact data and/or consequences data. These data have been displayed for panelists numerically, graphically, and interactively. The key in presenting feedback to panelists is to ensure that such data are user-friendly and understandable to the non-mathematician.

In the ECD context there may be other formats that are equally or more compelling to accomplish the purposes of feedback, which is, to provide information that allows the panelists to make more informed judgments. For example, the judgment about a particular task or set of tasks will have been based on the ALDs, which were based on the claims-evidence pairs. If the initial claims-evidence pair was inaccurate in the ALDs (either through a weak claims-evidence pair to begin with or through an inaccurate assignment of a claims-evidence pair to a particular

[9] Reckase charts are a graphical display of the conditional probabilities of a correct response for each item at each score level on the reporting scale. Each column contains data for a single item from the lowest scale score to the highest; each row contains data across all items at a single scale score point. Readers are referred to Loomis and Bourque (2001) for additional information.

level), then an adjustment would need to be made. Those links (or lack thereof) would become important feedback for panelists.

It becomes likely, as we describe this process of setting standards using feedback, that this is not the usual process with one pilot and then one subsequent operational meeting. It is an iterative process, probably requiring at least one or more pilots, and multiple meetings spanning a longer period of time than has been the case in the past.

Step 7: Compiling Ratings into Performance Standards

This stage is relatively straightforward. Mapping the results of EBSS onto the NAEP scale or scales would be accomplished in the usual way, ensuring that the integrity of the scaling technology is upheld. The NAEP TEL Framework indicates that three subscales have been recommended, as well as a composite NAEP scale. The final determination will be impacted by several factors, including the fact that this is a single-grade assessment (8) with a limited range task pool. Although interesting, NAEP is not a diagnostic instrument reporting on individual examinee performance. Also, standard setting in a multi-scale environment would require more work of the part of panelists. Plake et al. (2010) addressed this issue by having panelists develop standards on the subscales first, and then examining across subscales for "commonalities" to develop composite standards for the overall AP science scale. If NAEP found that helpful a similar procedure could be employed. However, if the ultimate decision is to report achievement levels only on the composite NAEP scale, then there is no need to develop standards on the individual subscales.

Step 8: Compiling Validity Evidence

At this point in the process there should be a very long trail of validity evidence available that could and should be compiled to support validity evidence called for by Kane (2001), Messick (1989), and others. Pitoniak and Hambleton outline a dozen kinds of procedural, internal, and external evidence that is customarily used to support validation efforts. Almost all of these approaches have been touched on in the course of this paper, and assembly of such data should not present a serious impediment to the full documentation of the process.

Summary

The following summary of the issues raised in this paper may be helpful in planning future agendas for the Board, seeking further advice from stakeholders and advisors, laying out the sequence of events in future Board contracts, and developing a research agenda to meet the needs of the TEL standard-setting meetings. They are not in priority order, and are presented as questions for consideration rather than recommendations.

1. What standard-setting methodology is best used to develop performance standards on the TEL assessment? Would it be best to use a legacy method and simply adapt it to a new context? Or, since this is a new assessment with no trend line to uphold, would it be best to start fresh? What risks are involved in using a new approach?
2. What should the Board be looking for in content experts identified for standard-setting panels? Can the selection criteria be operationalized in terms of both knowledge and skills background and demographics background?

3. What level of resources can be committed to the development of the ALDs? Is there documentation for the claims-evidence pairs that entered into the development of the item pool? How complete are those claims and evidence models? Can the pairs be mapped to a range of performances expected from grade 8 examinees? What is the Board's position on the "should" versus "can" issue? All things considered, is it advisable to change policy on this issue?
4. What resources can be committed to preparing all the documentation (e.g., internal and external research evidence) for the standard-setting contractor to implement the procedures smoothly, and for the SMEs to be trained efficiently? What approach will the Board require the standard-setting contractor to implement in order to ensure full participation by those selected for the panels?
5. To ensure feedback to panelists that is user-friendly and understandable to all panelists irrespective of background knowledge, will there be an opportunity for small pilot studies to test clarity of the feedback provided to panelists during the process, in addition to the field testing of the chosen method?
6. At what point in the process will the scaling be done? Will the field test results be scaled? If so, are the data representative enough to use as an indicator of what the final scaling might look like?
7. Will the trail of evidence be the sole responsibility of the standard-setting contractor? Or will there be an inter-contractor agreement for both the standard-setting contractor and the operations contractor to be mutually supportive of collecting and documenting such evidence?

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Behrens, J.T., Mislevy, R.J., Bauer, M., Williamson, D.M., & Levy, R. (2004). Introduction to evidence centered design and lessons learned from its application in a global e-learning program. *International Journal of Testing, 4,* 295-301.

Beimers, J.N., Way, W.D., McClarty, K.L., & Miles, J.A. (2012). *Evidence based standard setting: establishing cut scores by integrating research evidence with expert content judgments.* (Bulletin Jan 2012, Issue 21) Pearson Education. Retrieved April19, 2013 from the Pearson website http://www.pearsonassessments.com/hai/images/tmrs/Bulletin21_Evidence_Based_Standard_Setting.pdf

BMC Medical Education (2005). *Sicily statement on evidence-based practice.* Retrieved April 24, 2013 from the BMC website http://www.biomedcentral.com1472-6920/5/1

Cizek, G.J. (Ed.) (2001). *Setting performance standards: Concepts, methods, and perspectives.* Mahweh, NJ: Lawrence Earlbaum Associates, Publishers

Ewing, M., Packman, S., Hamen, C., & Thurber, A.C. (2010). Representing targets of measurement within evidence-centered design. *Applied Measurement in Education, 23*, 325-341.

Hambleton, R.K., & Pitoniak, M.J. (2006). Setting performance standards. In R.L. Brennan (Ed), *Educational measurement, 4th ed.* (pp. 433 – 470). Westport, CT: Praeger Publishers.

Hendrickson, A. , Huff, K., & Luecht, R. (2010). Claims, evidence, and achievement level descriptors as a foundation for item design and test specifications. *Applied Measurement in Education, 23*, 358-377.

Huff, K., Steinberg, L., & Matts, T. (2010). The promises and challenges of implementing evidence-centered design in large-scale assessment. *Applied Measurement in Education, 23*, 310-324.

Jaeger, R.M. (1995). Setting standards for complex performances: An iterative, judgmental, policy-capturing strategy. *Educational Measurement: Issues and Practices, 14 (4),* 16-20.

Kane, M. (2001). So much remains the same: Conception and status of validation in setting standards. In G. Cizek (Ed.), *Standard setting: Concepts, methods, and perspectives* (pp. 53 -88). Mahweh, NJ: Lawrence Earlbaum Associates, Publishers

Kak, N. & Cooper, M.A. (2001). *Measuring the competence of healthcare providers.* (Quality Assurance Project Issues Paper, Vol. 2, No. 1) Center for Human Services, U.S. Agency for International Development. Retrieved April 21, 2013 from the USAID website

Loomis, S. & Bourque, M.L. (2001). From tradition to innovation: Standard setting on the National Assessment of Educational Progress. In G. Cizek (Ed.), *Standard setting: Concepts, methods, and perspectives* (pp. 175-217). Mahweh, NJ: Lawrence Earlbaum Associates, Publishers

McClarty, K.L., Way, W.D., Porter, A.C., Beimers, J.N., & Miles, J.A. (2013). Evidence-based standard setting: Establishing a validity framework for cut scores. *Educational Researcher, 42,* 78-88.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement, 3rd ed.* (pp. 13-103). New York, NY: Macmillan Publishing

Mislevy, R.J., Almond, R.G., & Lukas, J.F. (2003). *A brief introduction to evidence-centered design* (Educational Testing Service Res. Rep. No. R-R-03-16). Princeton, NJ: Educational Testing Service.

National Assessment Governing Board. (March 4, 1995). *Developing student performance levels for the National Assessment of Educational Progress: Policy statement.* Retrieved April 17, 2013, from the NAGB web site: http://www.nagb.gov/policies.

National Assessment Governing Board. (Pre-publication edition, n.d.). *Technology and Engineering Literacy Framework for the 2014 National Assessment of Educational Progress.* (Available from the National Assessment Governing Board, 800 North Capitol Street NW, Suite 825, Washington, DC 20002)

Parshall, C.G., Spray, J.A., Kalohn, J.C., & Davey, T. (2002). *Practical considerations in computer-based testing.* New York, NY: Springer Publishing

Plake, B.S., Huff, K., & Reshetar, R. (2010). Evidence-centered assessment design as a foundation for achievement levels descriptor development and for standard setting. *Applied Measurement in Education, 23*, 307-309.

Prospective NCES Studies o NAEP as an Indicator of College and Career Readiness

Using Virginia Data to Examine the Relationship between NAEP Scores and Student Success in College and in the Labor Market

AIR proposes to use existing administrative records to examine the relationship between NAEP scores and set of concrete measures of student success in college and in the labor market. In this approach, "readiness" is reflected in the measurable success of students in college and in the workforce. The study will use detailed student level administrative records database maintained by the State Commission on Higher Education in Virginia (SCHEV) that tracks students' success while enrolled in Virginia's colleges and universities and their economic success in the workforce.

Each student record contains detailed information on student academic success (institution enrolled, remedial courses, GPA, major field of study, time to graduation) and student demographics (race, gender, ethnicity, Pell grant status). The wage data report annual earnings and industry of employment. For the purposes of examining the link between NAEP and college/career ready, Virginia's student records also contain student SAT scores. Since NAGB and NCES have created a table of correspondence between SAT and NAEP scores, each student in the SCHEV database can be assigned a NAEP score based o their recorded SAT test score. Given this linkage, AIR will be able to explore the relationship between student NAEP scores and a variety of college and labor market outcomes: Developmental Education, Retention/Duration of enrollment, Graduation, Student Labor Market Success (Earnings).

NAEP Reading – Lexile Linking Study

This special study is designed to link the NAEP Grade Reading scale to the MetaMetrics' Lexile® scale. Its goal is to use the relationship between the NAEP and Lexile® measures to evaluate whether grade 8 students are on track in terms of reaching preparedness for job training and/or college following graduation from grade 12. One advantage of the Lexile scale is that it is a continuous vertical scale that spans grades to 1 and into college and post-secondary career training. It has established benchmarks (made from studies conducted by MetaMetrics) that assess the complexity of reading required for university, community college, military, citizenship, and workforce preparedness.

In 2013 data were collected for a special study in which some students were administered both operational NAEP Reading items and Lexile®-calibrated items. Another group of students were administered Lexile®-only booklets. Data obtained in this study will allow us to estimate the correlation between NAEP and the Lexile® scale and provide a basis for projecting 8th grade NAEP Reading scores onto the Lexile® scale.

The linking study will empirically test the relationship between NAEP Reading and the Lexile® scale to determine if it is invariant across gender, race/ethnicity, and socio-economic status. For states in which

Lexile® data are available, the study should also investigate the correlation of state assessment scores in Reading in grade with those in grade 12. Additionally, for states in which Lexile® data are available, the study should estimate the correlation between NAEP reading (and the special Lexile® scaling test) and the Lexile® scores from the state assessment in reading for each grade available.

NAEP – High School Longitudinal Study

The High School Longitudinal Study of 200 (HSLS:09) is nationally representative, longitudinal study of more than 21,000 9th graders in 944 schools who will be followed throughout their secondary and postsecondary years. The study focuses o understanding students' trajectories from the beginning of high school into postsecondary education, the workforce, and beyond. What students decide to pursue, when, why, and how, are crucial questions for HSLS:09, especially, but not solely, in regards to science, technology, engineering, and math (STEM) courses, majors, and careers.

The NAEP-HSLS study ensured an overlap of NAEP grade 1 mathematics and HSLS samples such that student and parent background variable information from HSLS can be used to provide additional context for NAEP scores in the reporting of postsecondary outcomes. Students participating in the HSLS were sampled to take a grade 12 NAEP mathematics assessment in the main NAEP assessment testing window. The final numbers are not yet in but we can reasonably project the following:

1. Number of HSLS Schools that participated in NAEP : 340
2. Number of HSLS cohort kids in those schools: 5,200
3. Number of HSLS cohort kids that took NAEP: 4,400 (3,700 math; 700 reading).

AIR/ESSIN Research, Analysis, and Psychometric Support Study

In this project, ESSIN will estimate the NAEP equivalent scores for the SAT and ACT benchmarks for college readiness as selected by NAGB. This will be re-analysis of 201 study conducted by AIR/NESSI which had estimated the NAEP scores corresponding to the SAT benchmark for college readiness established by the College Board. For analyses involving statistical linking for the grade 1 mathematics assessments, data come from the 200 and 200 grade 1 NAEP High School Transcript Study. NAEP scale scores and SAT and/or ACT scores are included in the data for a portion of the sample.

The HSTS dataset includes only NAEP Mathematics and Science; therefore the linking between ACT Reading and grade 12 NAEP Reading assessments is accomplished by statistical moderation using national statistics (i.e., mean, variance, sample size) from 2005 and 2009 ACT and grade 12 NAEP assessments. However, the linking between ACT and SAT Mathematics and NAEP Mathematics can be accomplished using the regression method.

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

Ray Fields – May 9. 2013

Introduction

Rationale for NAEP Reporting on 12th Grade Academic Preparedness

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

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

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

Recognizing the importance of 12th grade as the gateway to postsecondary education and training, and viewing NAEP as a "truth teller" about student achievement, this distinguished panel of state and national leaders recommended unanimously that "NAEP should report 12th grade students' readiness for college-credit coursework, training for employment, and entrance into the military." (National Commission on NAEP 12th Grade Assessment and Reporting; p. 6.). They stated that "America needs to know how well prepared its high school seniors are…[only NAEP] can provide this information…and it is necessary for our nation's well-being that it be provided." (Ibid.; p. 2.).

The Governing Board approved this recommendation, with a minor modification. The term "readiness" was changed to "academic preparedness" and "entrance into the military" was subsumed by "job training." "Readiness" was changed to "academic preparedness" because "readiness" is broadly understood to include academic preparedness and other characteristics needed for success in postsecondary education and training, such as habits of mind, time management, and persistence (Conley). NAEP does not measure such characteristics. Rather, NAEP is designed to measure academic knowledge and skills. "Entrance into the military" was subsumed by "job training" with the intention of identifying occupations with civilian and military counterparts and utilizing both the military's experience as the world's largest occupational training organization and its extensive research on the relationship between performance on the Armed Service Vocational Aptitude Battery (ASVAB) and job training outcomes.

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

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

The Governing Board is not alone in recognizing the importance of 12th grade academic preparedness for the nation. Since the acceptance of the blue-ribbon panel report in 2004, the focus on ensuring that 12th grade students graduate "college and career ready" has been widely embraced as a policy goal by state and national leaders. These include the National Governors Association (NGA), the Council of Chief State School Officers (CCSSO), the Business Roundtable (BRT), the U.S. Chamber of Commerce (the Chamber), and the Obama Administration. The impetus for this attention to academic preparedness for college and job training is well summarized by a statement of the Business Coalition for Student Achievement, an organization coordinated by BRT and the Chamber:

> "Ensuring that all students graduate academically prepared for college, citizenship and the 21st century workplace…is necessary to provide a strong foundation for both U.S. competitiveness and for individuals to succeed in our rapidly changing world."

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

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

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

State tests and college admission tests do not provide a measure of what students know and can do at the very end of K-12 education. Using state tests and college admission tests for this purpose would be like performing final quality control on a product while it was still on the assembly line. Even if these state tests and college admission tests were administered at the 12th grade, they could not be combined to produce nationally representative results.

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

Defining Academic Preparedness for College

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

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

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

In order to design the NAEP 12th grade preparedness research, a working definition of preparedness was needed. The Governing Board's Technical Panel on 12th Grade Preparedness Research recommended use of the following definition with respect to academic preparedness for college.

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

This definition was intended to apply to the "typical" college, not to highly selective institutions, and thus, to the vast majority of prospective students, or about 80% of the college freshmen who enrolled in 2-year and 4-year institutions within 2 years following high school graduation (Ross, Kena, Rathbun, KewalRamani, Zhang, Kristapovich, and Manning, p 175). To make this clear, the definition is further elaborated as follows.

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

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

The Central Issue: Validity

Having made the decision to determine the feasibility of NAEP reporting on 12th grade academic preparedness, the Governing Board recognized that the central concern would be establishing the validity of inferences about 12th grade academic preparedness for use in NAEP reports. The Governing Board would need to ensure that the content of NAEP 12th grade reading and mathematics assessments was appropriate for measuring academic preparedness and that research was conducted to collect evidence by which the validity of proposed inferences could be evaluated. Finally, a formal validity argument would need to be developed, specifying the proposed inference(s) for NAEP reporting, the underlying assumptions or propositions, and the evidence related to the assumptions or propositions.

Accordingly, the Governing Board

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

The first phase of the Governing Board's program of preparedness research is completed. The studies were conducted in connection with the 2009 NAEP 12th grade assessments in reading and mathematics. More than 30 studies of five distinct types have been conducted. Study results are available and the complete studies are posted at http://www.nagb.org/what-we-do/preparedness-research.html. The National Center for Education Statistics (NCES) has provide additional data drawn from analyses of the 2005 and 2009 High School Transcript Studies conducted in connection with the NAEP 12th grade assessments in those years.

From this research, Governing Board staff developed a set of proposed inferences related to 12th grade academic preparedness for college. Following below is the validity argument for these proposed inferences. The validity argument begins with a statement of the proposed inferences. This is followed by a discussion of the limitations on interpretation and other caveats. An outline is then presented of the propositions and assumptions on which the inferences are based and the evidence related to the propositions and assumptions. The outline is followed by the text of the validity argument.

Validity Argument

Proposed Inferences

12th grade students scoring at or above the Proficient achievement level on the 12th grade NAEP Reading or Mathematics Assessment are

- likely to be academically prepared for first year college courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading or mathematics in college.

Limitations on Interpretation and Other Caveats

False Negatives

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

- likely to be academically prepared for first-year college courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading or mathematics in college,

but with a lower probability than those at or above Proficient. In mathematics, much more so than in reading, the research results suggest that the point on the NAEP scale indicating academic preparedness for college (i.e., "just academically prepared") is below Proficient, somewhere in the middle of the range between the Basic and Proficient achievement level cut scores.

Not a Preparedness Standard

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

Academically Prepared for College

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

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

The proposed inferences are not intended to reflect academic requirements for highly selective postsecondary institutions; to the additional academic requirements for specific majors or pre-professional programs, such as mathematics, engineering, or medicine; or to academic

requirements applicable to entry into certificate or diploma programs for job training or professional development in postsecondary institutions.

The proposed inferences are focused on the first year of college; they do not support conclusions about college persistence beyond the first year or completion of a degree.

GPA of B- or Better

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

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

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

Data Limitations

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

Preparedness for Job Training

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

Validity Argument Outline

Proposition/Assumption	Evidence
1. The content, test questions, and scoring criteria of the NAEP 12th grade reading and mathematics assessments cover academic knowledge and skills needed for college freshmen to be placed into entry-level, credit bearing courses.	1a. The documentation of the content and changes to the content of the 12th grade NAEP reading and mathematics frameworks, as revised in response to recommendations by Achieve, based on their American Diploma Project research 1b. Content alignment studies (ACT, SAT, ACCUPLACER)
2. The NAEP sampling, scaling and statistical procedures yield accurate estimates of the percentage of students scoring at or above a selected cut-score.	2. NAEP technical documentation of sampling, scaling and statistical procedures.
3. Scores on 12th grade NAEP reading and mathematics assessments provide accurate estimates of academic preparedness for entry level credit-bearing college courses.	Performance on the 12th grade NAEP reading and mathematics assessments is related to other indicators or criteria of academic preparedness for placement into entry-level credit-bearing college courses. 3a. Linking studies with ACT and SAT 3b. Cut-scores on SAT/ACT from higher education survey 3c. College Readiness Standards/Benchmarks for the ACT and the SAT 3d. Average NAEP scores of Florida students in/not in remedial and with GPA of B- or better Empirical indicators of student engagement do not support the assertion that NAEP 12th grade test-takers are not motivated. 3e. Percentage of test items attempted, including constructed response test questions. 3f. Correlations between performance on SAT and NAEP
4. The proposed test uses are appropriate and consequences are commensurate with intended uses.	5. Intended audience for the results is clearly stated; intended use is clearly described and disseminated to intended audience, along with caveats about potential over- or misinterpretation; the definition of preparedness is clearly defined and qualified; and materials are developed and disseminated consistent with the preceding requirements.

1. The content, test questions, and scoring criteria of the NAEP 12th grade reading and mathematics assessments cover academic knowledge and skills needed for college freshmen to be placed into entry-level, credit bearing courses.

NAEP Assessment Frameworks Were Revised to Measure Academic Preparedness

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

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

For NAEP to report on 12th grade academic preparedness for college, it must measure relevant content at the 12th grade. The content of each assessment is determined by the NAEP assessment frameworks, which the Governing Board is responsible for developing and approving. Accordingly, the Governing Board decided that the extant NAEP frameworks intended for the 2009 for reading and mathematics at the 12th grade would be reviewed. The review would identify changes needed to measure 12th grade academic preparedness for college.[2] Assessments at the 12th grade in reading and mathematics are conducted at least once every 4 years. In 2004, when the Board decided to proceed with the 12th grade academic preparedness initiative, 2009 was the next assessment year in which the 12th grade reading and mathematics assessments could be affected by framework changes.

In September 2004, the Governing Board contracted with Achieve, Inc. (Achieve) to review the NAEP 12th grade reading and mathematics assessment frameworks and identify where changes, if any, would be needed. Achieve had established the American Diploma Project (ADP) "...to improve postsecondary preparation by aligning high school standards, graduation requirements and assessment and accountability systems with the demands of college and careers (see www.achieve.org/adp-network)." The ADP had conducted research to identify key competencies in English and mathematics needed for high school graduates who aspire to higher education. They refer to these as the "ADP benchmarks."

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

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

Reading

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

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

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

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

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

> In May 2005, the Governing Board adopted a policy statement regarding NAEP and 12th-grade preparedness. The policy states that NAEP will pursue assessment and reporting on 12th-grade student achievement as it relates to preparedness for post-secondary education and training. This policy resulted from recommendations of the Board's National Commission on NAEP 12th Grade Assessment and Reporting in March 2004. Subsequent studies and deliberations by the Board took place during 2004 and 2005.
>
> In reading, the Board adopted minor modifications to the 2009 NAEP Reading Framework at grade 12 based on a comprehensive analysis of the framework conducted by Achieve, Inc. The current version of the reading framework incorporates these modifications at grade 12 to enable NAEP to measure and report on preparedness for postsecondary endeavors (National Assessment Governing Board, 2008, *Reading Framework*, p. v).

Mathematics

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

The Governing Board convened a panel of mathematicians and mathematics educators to review and revise the objectives in relation to the objectives for grades 4 and 8. The panel conducted focus groups with various NAEP constituents, using repeated rounds of reviews. The Governing Board approved the final set of grade 12 objectives on August 5, 2006. The changes to the framework were sufficiently modest to permit the continuation of the 12th grade trend line begun with the 2005 12th grade mathematics assessment under the previous 12th grade framework. Like the reading framework, the 2009/2013 mathematics framework for grade 12 states the Board's intention to measure 12th grade academic preparedness (National Assessment Governing Board, 2008, *Mathematics Framework*, pp. 2-3).

Examples of Objectives added to the 2009 Grade 12 Mathematics Framework

Number properties and operations

b) * Analyze or interpret a proof by mathematical induction of a simple numerical relationship.

Measurement

d) Interpret and use the identity $\sin^2\theta + \cos^2\theta = 1$ for angles θ between 0° and 90°; recognize this identity as a special representation of the Pythagorean theorem.

e) * Determine the radian measure of an angle and explain how radian measurement is related to a circle of radius 1.

f) * Use trigonometric formulas such as addition and double angle formulas.

g) * Use the law of cosines and the law of sines to find unknown sides and angles of a triangle.

Geometry

e) * Use vectors to represent velocity and direction; multiply a vector by a scalar and add vectors both algebraically and graphically.

g) * Graph ellipses and hyperbolas whose axes are parallel to the coordinate axes and demonstrate understanding of the relationship between their standard algebraic form and their graphical characteristics.

h) * Represent situations and solve problems involving polar coordinates.

Data Analysis, Statistics, and Probability
c) * Draw inferences from samples, such as estimates of proportions in a population, estimates of population means, or decisions about differences in means for two “”treatments”.”

e) * Recognize the differences in design and in conclusions between randomized experiments and observational studies.

k) * Use the binomial theorem to solve problems.

e) * Recognize and explain the potential errors caused by extrapolating from data.

Algebra
e) Identify or analyze distinguishing properties of linear, quadratic, rational, exponential, or *trigonometric functions from tables, graphs, or equations.

j) * Given a function, determine its inverse if it exists and explain the contextual meaning of the inverse for a given situation.

h) *Analyze properties of exponential, logarithmic, and rational functions.

g) * Determine the sum of finite and infinite arithmetic and geometric series.

Conclusion

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

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

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

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

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

The SAT is developed "to ensure that the topics measured on the SAT…reflect what is being taught in the nation's high schools and what college professors consider to be required for college success." (Kim, Wiley, and Packman, p.1)

The ACCUPLACER has the purpose of "… [determining] which course placements are appropriate for [incoming college] students and whether or not remedial work is needed." (ACCUPLACER, p. A-2)

The SAT, ACT and ACCUPLACER in reading and mathematics are widely used for these purposes by admissions and placement professionals in postsecondary education institutions. These testing programs regularly conduct curriculum surveys, validity studies and other research to support their claims that the content measured is directly related to the reading and mathematics knowledge and skills needed to qualify for entry-level credit-bearing courses.

Therefore, with the assumption that the SAT, ACT, and ACCUPLACER do measure the content needed for college level work, significant content overlap between NAEP and these other assessments would support the conclusion that what NAEP measures covers the knowledge and skills needed by college freshmen to be placed into entry-level credit bearing courses. That is,

- If A (the college admissions and placement tests) = B (the knowledge and skills needed for placement into entry-level credit-bearing courses without remediation); and
- C (NAEP) = A;
- Then C = B).

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

The Content Alignment Studies: Overview

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

The general design for the content alignment studies was to compare the 12th grade NAEP frameworks in reading and mathematics with the analogous document for the other test, and then to compare the test items from one test to the framework/analogous document of the other test. The reviews were performed by subject specific (i.e., mathematics, reading) panels, composed of experts in mathematics or reading and English instruction at the high school and college levels.

Alignment studies that compare an assessment to the content standards on which it is based are relatively common and have well-established methodologies. However, this is not true for the types of alignment studies the Governing Board planned to conduct: content alignment studies comparing different assessment programs. Different assessment programs have different purposes, different approaches to describing the domain being measured, and, possibly, different "grain size" in the level of detail in describing the domain. The Governing Board contracted with Norman Webb, a noted expert in content alignment studies, to prepare a design document for conducting the assessment to assessment alignment studies. The purpose was to put in place a methodology that considered the special challenges of assessment to assessment alignment studies and to foster comparability in the conduct of the studies and the reporting metrics across studies and contractors. The link to the Webb design document is at (http://www.nagb.org/content/nagb/assets/documents/publications/design-document-final.pdf). The Webb design was developed after the ACT alignment studies were completed. It was used in conducting the SAT and ACCUPLACER content alignment studies.

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

The Content Alignment Studies: Reading Results

Reading: ACT

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

The reading panel was composed of 7 members, with expertise in reading and/or English instruction at the high school and college levels. The panel was about evenly divided in terms of prior familiarity with either the ACT or NAEP reading domains.

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

> "All of the skills highlighted in the ACT [reading] domain and in the [ACT] College Readiness Standards [for reading] were identified within the NAEP Reading framework. In performing the comparison in the other direction—NAEP to ACT—it was the sense of the panel that the ACT measured primarily those skills that NAEP identifies as *Locate/Recall* and *Integrate/Interpret* skills, those that pertain primarily to finding explicit information in text (what the ACT would call Referring skills) and to making inferences, drawing conclusions, and making generalizations from information within text (what the ACT would call Reasoning skills). The panel saw less evidence of the higher-level analytical and evaluative *Critique/Evaluate* skills in the ACT domain, and attributed that to the multiple-choice format of the ACT [whereas NAEP includes constructed response items as well as multiple choice]. Another difference is that NAEP includes items and texts measuring how well an examinee can apply reading skills across texts, whereas the paired passage format is not a feature of the ACT. So, while the NAEP Reading framework and the ACT Reading domain, test specifications, and College Readiness Standards share similarities, important differences in what and how the assessments measure suggest caution when drawing comparisons between the assessments." (p.17)

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

> "A total of 152 Reading items (comprising 17 blocks) were classified in [the reading] study. Of these, 97 were multiple-choice (MC). Nine were dichotomously-scored ("incorrect" or "correct") short constructed-response (DSCR) items. Thirty-three were polytomously-scored short constructed-response (PSCR) items, each scored using a three-point scoring rubric. Thirteen were extended constructed-response (ECR) items, each scored using a four-point rubric. Each DSCR had one creditable score category, each PSCR had two, and each ECR had three. Each Reading panelist, therefore, assigned a total of 211 classifications to the NAEP Reading items [and rubric scoring categories]." (p.54)

An item or score category was deemed "classified" if there was majority agreement (at least 4 of the 7 panel members) or supermajority agreement (5 or more panel members) about the score band to which an item (or creditable score category under an item rubric) was assigned.

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

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

items involving multiple texts in NAEP but not the ACT, and the greater presence of "Critique/Evaluate" type items in NAEP than the ACT.

Of the 137 classifications, 24 were in the score bands from 13-19; 113 of the classifications were in the score bands from 20-36. This is noted because the ACT College Readiness Benchmark for reading is 21. The ACT College Readiness Benchmark signifies the score at which a student has a 50% chance of attaining a grade of B or better in a relevant subject and a 75% change of a C or better. In addition, the Governing Board conducted a survey of postsecondary institutions' use of tests in making entry-level decisions about placement into remedial or regular credit-bearing courses. With respect to the ACT, 18 was the mean reading score below which students were deemed to need remedial course work (Fields and Parsad, P. 19). While this provides a context for the study results, it must be kept in mind that in making their judgments about assessment content, the panelists did not have data about NAEP item difficulty or data on how performance on NAEP compares with performance on the ACT.

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

Reading: SAT

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

Overall, the study found similar content in the NAEP 12th grade reading assessment and the SAT critical reading test. Following below is an excerpt from the Executive Summary of the report (pp. iv-vi).

> ***What is the correspondence between the reading content domain assessed by NAEP and that assessed by SAT?***
> The greatest commonality between the two tests is their shared emphasis on the broad skills of integrating and interpreting both informational and literary texts. This is evident in the majority of items from both tests aligned to NAEP Standard 2, Integrate/Interpret," including many to Goal 2.1, "Make complex inferences within and across *both literary and informational texts*."
>
> Despite the difference in the degree of specificity of the two frameworks (most NAEP objectives are much more finely grained than the SAT objectives), there is also considerable overlap at the level of more specific skills.

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

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

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

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

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

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

Even with the minimal coverage of NAEP "Locate/Recall" and "Critique/Evaluate" standards by SAT items, all NAEP items found a match in the SAT framework. However, the broad language of the SAT framework can encompass the range of the NAEP items. For example, SAT Goal B.2, "Literal Comprehension," refers to items that "ask what is being said" in a "small but significant portion of a reading passage," a description that can easily accommodate most NAEP "Locate/Recall" items and objectives. In fact, nearly all items on the NAEP short version that were coded to

"Locate/Recall" objectives in the NAEP framework were matched to SAT Goal B.2 in the SAT framework.

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

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

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

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

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

Reading: ACCUPLACER

The Governing Board contracted with WestEd, an independent third party, to conduct the content alignment study comparing the NAEP 12th grade reading assessment and the ACCUPLACER reading test. The ACCUPLACER is used specifically to determine whether entry-level students have the reading skills necessary for college level work or require remedial reading courses. WestEd conducted the content alignment study using the design developed for the Governing Board by Norman Webb. The full report of the content alignment study can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACCUPLACER-NAEP_Reading_Content_Comparison.pdf.

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

> ***What is the correspondence between the reading content domain assessed by NAEP and that assessed by ACCUPLACER?***
> The greatest commonality between the two tests is in their shared emphasis on the broad skills of comprehending and interpreting informational text, primarily through inferential reasoning. This is evident in the majority of items on both tests (two-thirds to three-fourths) matched to the NAEP standard "Integrate/Interpret: Make complex inferences within and across texts." On both tests, the majority of alignments to "Integrate/Interpret" were to objectives that apply to informational text only or across both informational and literary texts.
>
> The shared emphasis on the comprehension and interpretation of informational text can also be seen in the alignments on both tests to the ACCUPLACER framework. Although the ACCUPLACER standards do not explicitly refer to text type, they focus almost exclusively on elements typical of informational text. A majority of both NAEP and ACCUPLACER items were matched to the ACCUPLACER standard "Inferences," and both tests had notable percentages of alignments to "Direct statements and secondary ideas" and "Applications." A smaller percentage of items on both tests were aligned to "Identifying main ideas."
>
> ***To what extent is the emphasis of reading content on NAEP proportionally equal to that on ACCUPLACER?***
> As previously discussed, the alignments both within and across frameworks show that both tests emphasize the comprehension and interpretation of informational text, particularly through the use of inference. Within this broad area of convergence, however, there are differences in emphasis revealed in the alignments to specific objectives within both frameworks. In relation to the NAEP framework, the NAEP short-version items showed a far greater emphasis on the comprehension of vocabulary in context (Objective 4.a) and on the analysis of an author's use of language (Objective 1.d). In relation to the ACCUPLACER framework, NAEP items showed more emphasis on the use of inference to interpret text ("Inferences"). The higher percentage of NAEP items aligned to "Applications" also reflects the greater emphasis in NAEP on understanding authors' use of language.
>
> In relation to the ACCUPLACER framework, the ACCUPLACER items showed a greater emphasis than the NAEP items on the identification of main ideas. In relation to the NAEP framework, the ACCUPLACER items showed more emphasis on the recall of specific details, facts, and information (NAEP 1.1.a).
>
> In general, in the cross-framework alignments, the matches found in each test to the other's framework (NAEP to ACCUPLACER and ACCUPLACER to NAEP) tended to be for the most general objectives within that framework. For example, the great majority

of hits for ACCUPLACER items to NAEP objectives for "Integrate/Interpret" were to two of the most broadly stated NAEP objectives, "Draw conclusions" (2.3.b) and "Compare or connect ideas" (2.1.b). Many of the more specific NAEP objectives for "Integrate/Interpret," such as "Find evidence in support of an argument" (2.2.c), received far fewer or no hits from ACCUPLACER items. Compared to ACCUPLACER, the NAEP items were more evenly distributed among NAEP objectives.

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

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

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

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

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

Although the majority of the ACCUPLACER item alignments were to objectives for NAEP "Integrate/Interpret," over half of these items were found to have a DOK level below that of the standard. In addition, the use of very short reading passages and exclusively multiple-choice items in ACCUPLACER may be less conducive to the more

in-depth reasoning required by DOK Level 3. NAEP, by contrast, includes much longer reading passages and both multiple-choice and constructed-response items.

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

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

The Content Alignment Studies: Summary Discussion for Reading

Three content alignment studies were conducted to examine the extent to which

- The content of the NAEP 12th grade reading assessment covers the knowledge and skills needed for college freshmen to be placed into entry-level credit bearing courses. and
- NAEP 12th grade reading test items and scoring criteria are appropriate for obtaining evidence of test takers' possession of knowledge and skills needed for college freshmen to be placed into entry-level credit-bearing courses requiring college level reading.

For short-hand, this will be referred to as "academic preparedness for college."

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

assumption is that if the content of the 12th grade NAEP reading assessment is similar to the content of these reading tests, then the NAEP content is directly related to "academic preparedness for college."

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

The SAT study found that "Overall, the alignment results across the two sets of items and frameworks show a strong area of overlap in their coverage of SAT "Passage-Based Reading" objectives and NAEP "Integrate/Interpret" objectives, as well as some important differences." With respect to the differences, "…SAT items had limited coverage of the knowledge and skills described by the NAEP standards "Locate/Recall" and "Critique/Evaluate." This difference is also reflected in test format, with the use of longer reading passages and both constructed-response and multiple-choice items in NAEP. In comparison, all SAT items are multiple-choice. With regard to differences in content as described in the SAT framework, NAEP does not include sentence-completion items."

The ACCUPLACER study found that "The greatest commonality between the two tests is in their shared emphasis on the broad skills of comprehending and interpreting informational text, primarily through inferential reasoning. This is evident in the majority of items on both tests (two-thirds to three-fourths) matched to the NAEP standard 'Integrate/Interpret: Make complex inferences within and across texts.' On both tests, the majority of alignments to 'Integrate/ Interpret' were to objectives that apply to informational text only or across both informational and literary texts…Overall, the NAEP [frameworks and] items covered a broader range of cognitive complexity than the ACCUPLACER items…The three NAEP standards, defined in terms of three different "cognitive targets" ("Locate/Recall," "Integrate/Interpret," and "Critique/Evaluate"), cover a broader range of cognitive complexity supported by the use of longer reading passages and the inclusion of both short and extended constructed-response items."

The results across the three studies are consistent. In general, the content of the ACT, SAT, and ACCUPLACER reading tests are present in NAEP, but NAEP is generally broader. Alignment between NAEP and the other three respective assessments is substantial, but not perfect; perfect alignment is not expected. A component of the SAT critical reading assessment not present in NAEP is sentence completion, measuring vocabulary knowledge in a different way than NAEP does.

These results support the conclusion that

- The content of the NAEP 12th grade reading assessment covers the knowledge and skills needed for college freshmen to be placed into entry-level credit bearing courses. and
- NAEP 12th grade reading test items and scoring criteria are appropriate for obtaining evidence of test takers' possession of knowledge and skills needed for college freshmen to be placed into entry-level credit-bearing courses requiring college level reading.

The Content Alignment Studies: Mathematics Results

Mathematics: ACT

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

The mathematics panel was composed of 7 members, with expertise in mathematics instruction at the high school and college levels. The panel was about evenly divided in terms of prior familiarity with either the ACT or NAEP mathematics domains.

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

> "… the two assessments have much of their content domains in common. However, in the NAEP-to-ACT comparison, the difference in specificity with which the domains are articulated in the assessment documents left the panel uncertain as to whether a number of NAEP content topics—those pertaining to transformations, probability, statistics, and data analysis—are assessed by the ACT. In addition, there was some uncertainty within the panel on the degree to which higher-order analytic skills were assessed, and it was the sense of the panel that the ACT Mathematics Test contained few items involving high mathematical complexity, at least as the NAEP defines it. With regard to the ACT to-NAEP comparison, the Mathematics panel found nearly all of the ACT Mathematics domain and College Readiness Standards reflected in the NAEP Mathematics domain, but determined that a number of the lower-level topics in the ACT Pre-Algebra subdomain were more consistent with Grade 8 NAEP topics. All of these points suggest that while there may be substantial overlap in what the two assessments measure and how they measure it, there are areas of difference, as well. (p. 17)

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

An item or score category was deemed "classified" if there was majority agreement (at least 4 of the 7 panel members) or supermajority agreement (5 or more panel members) about the score band to which an item (or creditable score category under an item rubric) was assigned.

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

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

Of the 190 classifications, 10 were in the score bands from 13-19; 180 of the classifications were in the score bands from 20-36. This is noted because the ACT College Readiness Benchmark for mathematics is 22. The ACT College Readiness Benchmark signifies the score at which a student has a 50% chance of attaining a grade of B or better in a relevant subject and a 75% change of a C or better. In addition, the Governing Board conducted a survey of postsecondary institutions' use of tests in making entry-level decisions about placement into remedial or regular credit-bearing courses. With respect to the ACT, 19 was the mean mathematics score below which students were deemed to need remedial course work in mathematics (Fields and Parsad, p. 13). While this provides a context for the study results, it must be kept in mind that in making their judgments about content, the panelists did not have data about NAEP item difficulty or data on how performance on NAEP compares with performance on the ACT.

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

Mathematics: SAT

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

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

> ***"What is the correspondence between the mathematics content domain assessed by NAEP and that assessed by SAT?***
> At the standard level, the wording of the standards in the two frameworks is very similar. Both the NAEP and SAT frameworks include virtually the same five broad content categories, with SAT combining geometry and measurement into one standard. Each framework contains both general and specific objectives, although the SAT objectives, which are presented as content topics without indication of the cognitive level at which that content would be assessed, may be interpreted as more general than the NAEP objectives.
>
> Although the structures of the two frameworks differ greatly beyond the standard level (including the NAEP framework having three levels while SAT has two), the mathematics areas typically expected of grade 12 students—number and operations, geometry and measurement, data analysis and probability, and algebra—are addressed in somewhat similar proportions.
>
> ***To what extent is the emphasis of mathematics content on NAEP proportionally equal to that on SAT?***
> The greatest commonality between the two tests is their emphasis at the standard level. This is evident in the distribution of percentages of total hits from both assessments matched to each set of standards. Although there are some differences of emphasis, such as the full NAEP item pool's greater proportion of alignment to SAT "Data analysis, statistics, and probability," and the SAT short-version's greater proportion of alignment to SAT "Geometry and measurement," the proportions of alignments to "Algebra and functions" and "Number and operations" are comparable. There is also considerable overlap among some specific skills, with both assessments addressing many of the same NAEP "Number properties and operations" objectives and SAT objectives…
>
> Despite the difference in the degree of specificity of the two frameworks (most NAEP objectives are much more finely grained than the SAT objectives), it is clear that both assessments emphasize a number of the same or closely related skills. These include properties, equivalence, and operations on rational numbers (included in NAEP Goals 1.1 and 1.3 and included in SAT Objective N.2) and properties of two-dimensional shapes (included in NAEP Goals 3.1 and 3.3 and included in SAT Objective G.6).
>
> ***Are there systematic differences in content and complexity between NAEP and SAT assessments in their alignment to the NAEP framework and between NAEP and SAT assessments in their alignment to the SAT framework? Are these differences such that entire mathematics subdomains are missing or not aligned?***
> While there is considerable overlap between the two assessments, primarily in the intersection of the NAEP "Algebra" and SAT "Algebra and functions" standards, there

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

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

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

Mathematics: ACCUPLACER

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

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

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

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

The NAEP and ACCUPLACER assessments both cover certain content traditionally expected of grade 12 students, namely the two content subdomains of number or number operations and algebra (included in NAEP's "Number properties and operations" and "Algebra" standards and in ACCUPLACER's "Arithmetic," "Elementary algebra," and "College level math" standards), although their respective degrees of alignment and focus in these subdomains vary. Whereas the NAEP items focus primarily on number or number operations and algebra content at the grade 12 level, with an emphasis on problem solving and application of concepts at that grade level, the ACCUPLACER items span a wider developmental and grade-level range (from basic to more advanced). This difference in focus is consistent with the purposes of the two assessments and their frameworks. The NAEP objectives are written to describe assessable content for grade 12 mathematics; thus, the 130 objectives tend to address the skills and concepts specific to that grade. The purpose of ACCUPLACER is to help determine appropriate placement for an individual student, and so the 87 ACCUPLACER objectives are spread more broadly across grade levels and are intended to be more general.

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

Regarding alignment to the NAEP framework, within the "Number properties and operations" and "Algebra" standards, NAEP items had broader overall coverage of the NAEP objectives than did ACCUPLACER. The 42 NAEP items (the short version used for within-framework alignment) aligned to 72 NAEP objectives, whereas the 105 ACCUPLACER items (one complete form of each of the three ACCUPLACER Mathematics Core tests) aligned to only 56 NAEP objectives, with 44% of the ACCUPLACER item alignments aligning to only three NAEP objectives (all in "Number properties and operations" and "Algebra"). These differences in breadth and emphasis between the two assessments were evident across all NAEP standards. For example, in each assessment, items were aligned to four NAEP "Algebra" objectives for which the other assessment had no alignments, reflecting differences in emphasis within that standard.

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

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

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

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

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

The Content Alignment Studies: Summary Discussion for Mathematics

Three content alignment studies were conducted to examine the extent to which

- The content of the NAEP 12th grade mathematics assessment covers the knowledge and skills needed for college freshmen to be placed into entry-level credit bearing mathematics courses. and
- NAEP 12th grade mathematics test items and scoring criteria are appropriate for obtaining evidence of test takers' possession of knowledge and skills needed for college freshmen to be placed into entry-level credit-bearing mathematics courses.

For short-hand, this will be referred to as "academic preparedness for college."

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

The ACT study found that "With regard to the ACT to-NAEP comparison…nearly all of the ACT Mathematics domain and College Readiness Standards [are] reflected in the NAEP Mathematics domain, but…a number of the lower-level topics in the ACT Pre-Algebra subdomain were more consistent with Grade 8 NAEP topics." In the NAEP-to ACT comparison, there was uncertainty about "…whether a number of NAEP content topics—those pertaining to transformations, probability, statistics, and data analysis—are assessed by the ACT….and the degree to which higher-order analytic skills were assessed…and it was the sense of the panel that the ACT Mathematics Test contained few items involving high mathematical complexity, at least as the NAEP defines it."

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

The ACCUPLACER study found that "The NAEP and ACCUPLACER assessments both cover certain content traditionally expected of grade 12 students, namely the two content subdomains of number or number operations and algebra…although their respective degrees of alignment and focus in these subdomains vary… the 105 ACCUPLACER items aligned primarily (90%) to a total of seven out of 24 NAEP goals: three of the six goals from "Number properties and operations" in the NAEP framework, and four of the five goals in "Algebra." Conversely, the NAEP items from the full pool of 164 items that aligned to the ACCUPLACER framework were distributed fairly evenly across the three ACCUPLACER standards and found to align to 75

ACCUPLACER objectives…Regarding differences in alignment of content, ACCUPLACER items had very limited coverage of measurement, geometry, and data analysis, content that is not included in the ACCUPLACER framework but that is included in the NAEP framework. Many NAEP items assessing these subdomains were found to be uncodable to the ACCUPLACER objectives…"

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

These results support the conclusion that

- The content of the NAEP 12th grade mathematics assessment covers the knowledge and skills needed for college freshmen to be placed into entry-level credit bearing mathematics courses. and
- NAEP 12th grade mathematics test items and scoring criteria are appropriate for obtaining evidence of test takers' possession of knowledge and skills needed for college freshmen to be placed into entry-level credit-bearing mathematics courses.

2. The NAEP sampling, scaling and statistical procedures yield accurate estimates of the percentage of students scoring at or above a selected cut-score..

The NAEP sampling, scaling, and statistical procedures are widely accepted, well documented (for example, see National Center for Education Statistics, pp. 70-71) and have been periodically evaluated over two decades (for example, see complete list of research conducted by the NAEP Validity Studies Panel at http://www.air.org/reports-products/index.cfm?fa=viewContent&content_id=890 and "Evaluation of the National Assessment of Educational Progress: Study Reports" at http://www2.ed.gov/rschstat/eval/other/naep/naep-complete.pdf). Other than issues relating to the comparability among the state-level NAEP samples of inclusion rates of students with disabilities and students who are English language learners (about which the Governing Board and NAEP have taken and continue to take significant action), there is little dispute about the appropriateness of the NAEP sampling, scaling and statistical procedures for estimating the percentage of students scoring at or above a selected cut-score.

This is relevant because the proposed inferences that are the subject of this validity argument are interpretations to add meaning to the Proficient achievement levels for NAEP 12th grade reading and mathematics. The percentages of students at or above each of the NAEP achievement levels (Basic, Proficient, and Advanced) have been estimated and reported regularly, beginning with assessments in 1992. The added meaning being given to the Proficient achievement levels will not affect in any way the accuracy of the estimates of the percentages of students scoring at or above the Proficient cut-score.

3. Scores on 12th grade NAEP reading and mathematics assessments provide accurate estimates of academic preparedness for entry level credit-bearing college courses.

- **Performance on the 12th grade NAEP reading and mathematics assessments is related to other indicators or criteria of academic preparedness for placement into entry-level credit-bearing college courses.**

In addition to examining the overlap in test content between NAEP and the tests for college admission and placement, the Governing Board determined that it would be relevant and important to examine how performance on NAEP relates to performance on the SAT and ACT, including the college readiness benchmarks associated with these testing programs. There are several data sources for the analyses: the NAEP/SAT linking studies (see report at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/SAT-NAEP_Linking_Study.pdf), the Florida longitudinal study (see report at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/Florida_Statistical_Study.pdf), the 2005 and 2009 NAEP High School Transcript Studies, and the Governing Board's survey of postsecondary education institutions' use of tests and the cut-scores on those tests for determining whether incoming students need remedial instruction in reading and mathematics (Fields and Parsad).

Indicators: College Board and ACT College Readiness Benchmarks

The College Board and ACT, Inc. have established college readiness benchmarks for the SAT and the ACT in a number of subjects tested, including reading and mathematics. The SAT College Readiness Benchmark for critical reading and mathematics is a score of 500 on the respective tests. According to the College Board's research, a score of 500 predicts, with a .65 probability, a first-year GPA of B- or better. The ACT College Readiness Benchmark for reading is a score of 21. According to ACT's research, a score of 21 predicts, with a .50 probability, a grade of B or better (or .75 probability of a C or better) in first year courses requiring college reading, such as history and the social sciences. A score of 22 on the ACT mathematics tests predicts a .50 probability of a grade of B or better in a first-year mathematics course, or a .75 probability of a grade of C or better. The College Board and ACT research is based on the first-year outcomes of their respective test takers.

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

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

Analysis of Results for Mathematics

The Governing Board's program of preparedness research included a statistical linking study between the NAEP 12th grade mathematics assessment and the SAT mathematics test. Through a partnership with the College Board, the mathematics SAT scores of students who took the 12th grade NAEP mathematics assessment in 2009 were obtained and analyzed.

A correlation of .91 was found for performance on the NAEP 12th grade mathematics assessment and the SAT mathematics test. This high correlation, together with the substantial overlap in test content found in the content alignment studies between the NAEP and SAT mathematics tests, supports inferences about NAEP performance in relation to SAT performance. Of particular interest, is how performance on NAEP relates to the SAT College Readiness Benchmark for mathematics (i.e., a score on the SAT mathematics test of 500 or better). The SAT benchmark provides "an indication of college readiness at a typical college (College Board)." This is consistent with the Governing Board's definition of academic preparedness.

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

The analysis of the mathematics indicators is displayed in Figure 1. A consistent pattern is evident across studies and across time. This consistent pattern supports the inferences that 12th grade students scoring at or above Proficient on the 12th grade NAEP Mathematics Assessment are

- likely to be academically prepared for first year college mathematics courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in mathematics in college.

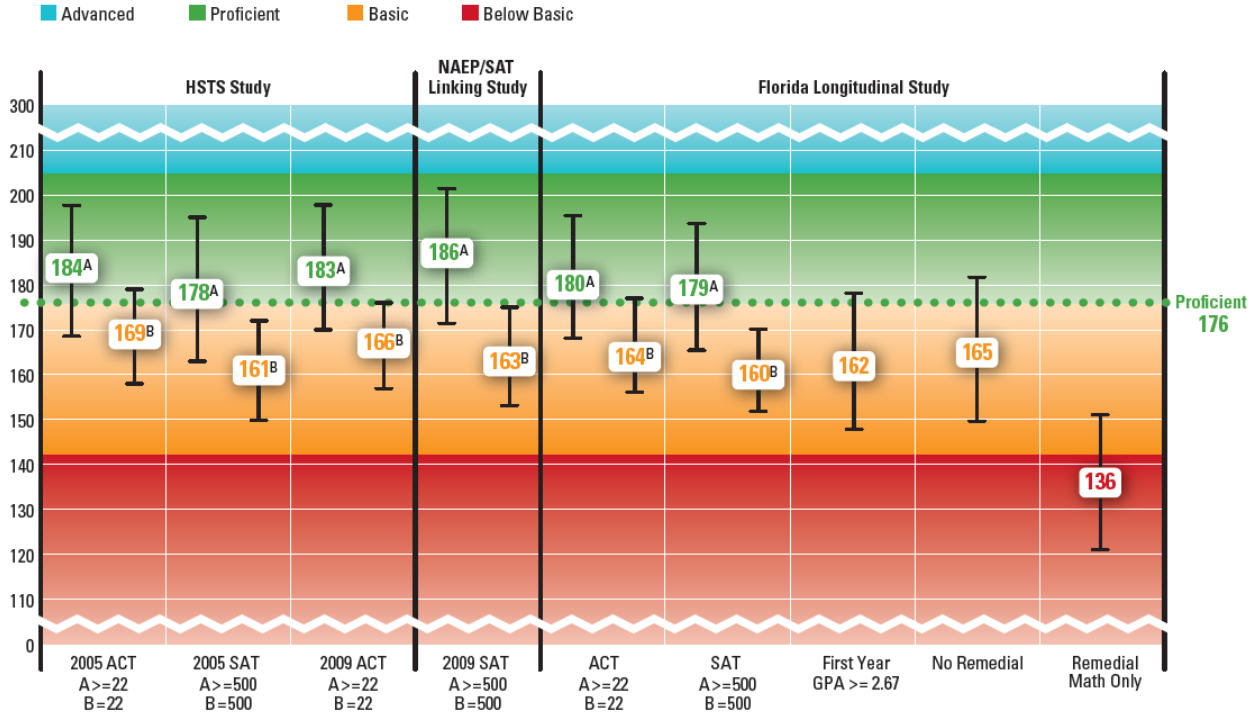
The average NAEP mathematics scores for 12th grade students scoring **at** the SAT College Readiness Benchmark for mathematics are compared first for the two national studies: the 2005 High School Transcript Study (HSTS) and the 2009 NAEP/SAT Linking Study (NSLS). The average scores are 161 and 163 respectively. These scores are somewhat below the cut-score for Proficient, which is 176 on the NAEP 12th grade mathematics scale. If the interpretation is made that students scoring at 163 on the NAEP 12th grade mathematics scale have a .65 chance of attaining a first-year GPA of B- or better, and this score is below Proficient, then it follows that students scoring at or above Proficient have increasingly higher probabilities of attaining a first-year GPA of B- or better.

It also means that many students who score in the mid-range and above of the Basic achievement level on the 12th-grade NAEP mathematics assessment may be academically prepared for college. For example, considering the NSLS results, students with 2009 SAT scores of 500 have an average NAEP score of 163, with an interquartile range of 153 to 175. A substantial percentage of the NAEP scores for these students are in this range. Similar results are observed for other measures in Table 1.

Figure 1

NAEP 12th Grade Preparedness Research: Mathematics

Average Scores and Inter-quartile Ranges For Selected Variables, 2005 High School Transcript Study, 2009 High School Transcript Study, 2009 NAEP/SAT Linking Study, 2009 Florida Longitudinal Study



Thus, while getting a NAEP score in the Proficient category provides a very strong indication that a student is academically prepared for college, students in the upper region of the Basic range are also likely to be academically prepared for college, but with a lower probability.

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

Another analysis examines the average scores and interquartile ranges for students scoring **at or above** the SAT College Readiness Benchmark for mathematics from the 2005 HSTS, 2009 NSLS, and 2009 FLS. In all three cases, the interquartile ranges fall around Proficient and overlap to a high degree.

As discussed previously, the ACT College Readiness Benchmark for mathematics is defined differently than the SAT College Readiness Benchmark for mathematics. However, it is noteworthy that even with this different definition, the results from the 2005 HSTS, 2009 HSTS, and 2009 FLS analyses are very similar to the results for the SAT.

Taken together, these results support the inference that students scoring at or above Proficient on the NAEP 12th grade mathematics scale are likely to be academically prepared for entry-level credit-bearing mathematics courses and to attain a first-year GPA of B- or better.

To answer the question, what is the relationship between performance on NAEP and student outcomes, we look to the Florida longitudinal study results. First we examine the average NAEP mathematics score for the 12th grade Florida NAEP test takers who actually attained a first-year GPA of B- or better. The average NAEP score for these students was 162, somewhat below the Proficient cut point. This is consistent with the SAT College Readiness Benchmark analyses and further supports the inference that students at or above Proficient are likely to be academically prepared and attain a first-year GPA of B- or better. It follows, of course, that students who are academically prepared will not require remedial courses.

Thus, another outcome of interest is placement of entry-level students into remedial college courses versus non-remedial credit-bearing courses. Here again, we look to the FLS as a data source. The average NAEP mathematics score for the Florida NAEP test-takers not placed into remedial courses was 165, somewhat below the NAEP Proficient cut-score of 176. The average score for Florida students placed into remedial mathematic was 136, which is in the range below Basic. These results lend support, together with the SAT and ACT analyses, to the inference that students scoring at or above Proficient are not likely to need remedial courses in mathematics.

Analysis of Results for Reading

The Governing Board's program of preparedness research included a statistical linking study between the NAEP 12th grade reading assessment and the SAT critical reading test. Through a partnership with the College Board, the SAT critical reading scores of students who took the 12th grade NAEP reading assessment in 2009 were obtained and analyzed.

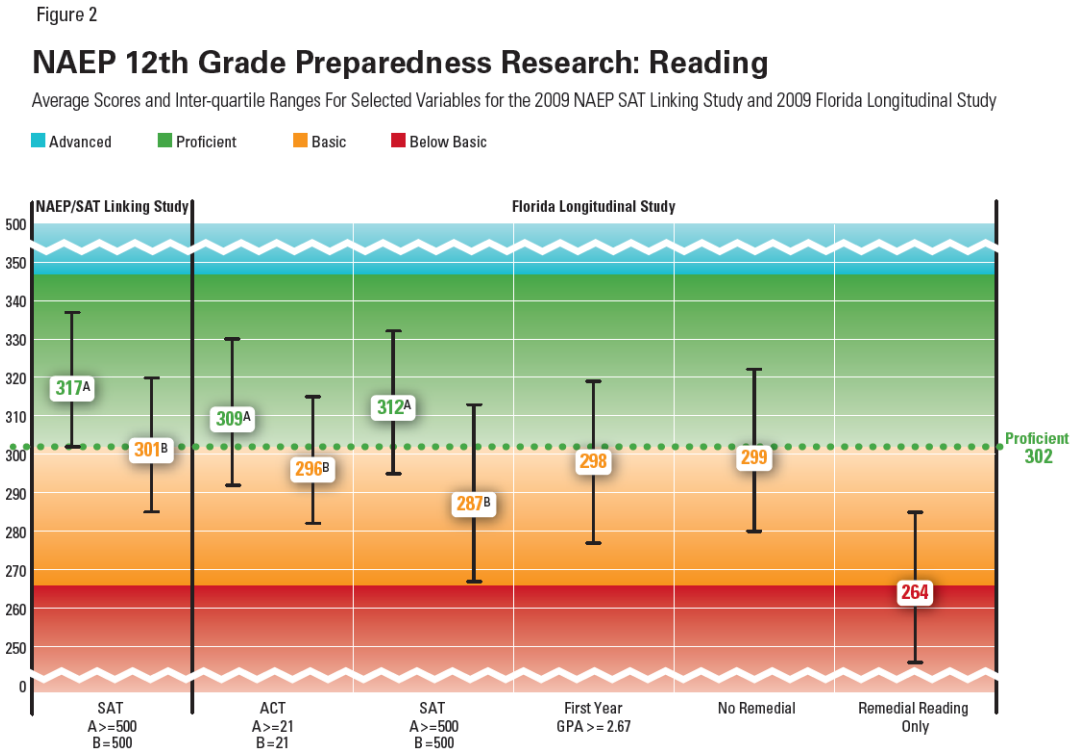
A correlation of .74 was found for performance on the NAEP 12th grade reading assessment and the SAT critical reading test. This is a substantial correlation. While it may not be high enough to predict the performance of individual students from one test to another, it is sufficient to support the group-level inferences reported by NAEP. This, together with the substantial overlap in test content found in the content alignment studies between the NAEP and SAT reading tests, supports inferences about NAEP performance in relation to SAT performance. Of particular interest, is how performance on NAEP relates to the SAT College Readiness Benchmark for reading (i.e., a score on the SAT mathematics test of 500 or better). The SAT benchmark provides "an indication of college readiness at a typical college (College Board)." This is consistent with the Governing Board's definition of academic preparedness.

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

The analysis of the reading indicators is displayed in Figure 2. A consistent pattern is evident across studies. This consistent pattern supports the inferences that 12th grade students scoring at or above Proficient on the 12th grade NAEP Reading Assessment are

- likely to be academically prepared for first year courses requiring college level reading,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading in college.

Figure 2



The average NAEP reading score for 12th grade students scoring at the SAT College Readiness Benchmark for reading is examined first for the national 2009 NSLS. The average score is 301, just below the cut score for Proficient, which is 302 on the NAEP 12th grade reading scale. If the interpretation is made that students scoring at 301 on the NAEP 12th grade reading scale have a .65 probability of attaining a first-year GPA of B- or better, and this score is below Proficient, then it follows that students scoring at or above Proficient have increasingly higher probabilities of attaining a first-year GPA of B- or better.

These results are confirmed by the Florida longitudinal study results (FLS). The average NAEP reading score for the 12th grade Florida NAEP test takers who scored at the SAT College Readiness Benchmark of 500 was 287, somewhat below the Proficient cut score, like the 2009 NSLS results.

Another analysis examines the average scores and interquartile ranges for students scoring at or above the SAT College Readiness Benchmark for reading from the 2009 NSLS and 2009 FLS. In both cases, the interquartile ranges fall around Proficient and overlap to a high degree.

As discussed previously, the ACT College Readiness Benchmark for mathematics is defined differently than the SAT College Readiness Benchmark for mathematics. However, it is noteworthy that even with this different definition, the results from the 2009 FLS analysis is very similar to the results for the SAT.

Taken together, these results support the inference that students scoring at or above Proficient on the NAEP 12th grade reading scale are likely to be academically prepared for entry-level credit-bearing courses requiring college level reading and to attain a first-year GPA of B- or better.

To answer the question, what is the relationship between performance on NAEP and student outcomes, we look to the Florida longitudinal study results. First we examine the average NAEP reading score for the 12th grade Florida NAEP test takers who actually attained a first-year GPA of B- or better. The average NAEP score for these students was 298, just below the Proficient cut point. This is consistent with the SAT College Readiness Benchmark analysis and further supports the inference that students at or above Proficient are likely to be academically prepared and attain a first-year GPA of B- or better. It follows, of course, that students who are academically prepared will not require remedial courses.

Thus, another outcome of interest is placement of entry-level students into remedial versus non-remedial credit-bearing courses. Here again, we look to the FLS as a data source. The average NAEP reading score for the Florida NAEP test-takers not placed into remedial courses was 299, again, just below the NAEP Proficient cut-score of 302. This lends support, together with the SAT and ACT analyses, to the inference that students scoring at or above Proficient are not likely to need remedial courses in reading.

- **Empirical indicators of student engagement do not support the assertion that NAEP 12th grade test-takers are not motivated.**

A recurring question about NAEP in general is whether student achievement is underestimated because the test-takers receive no test results back and bear no consequences for their performance. It is a relevant and legitimate question potentially affecting the accuracy of NAEP estimates.

The question is asked with special skepticism about 12th grade NAEP test-takers: will "test-wise" high school seniors in the last semester of their K-12 experience, knowing that the results will not affect their grades or future opportunities, apply the same effort that they would to tests that do come with high stakes for them? Will they be "motivated" when they sit for NAEP?

Associated with this question is the assertion that 12th grade NAEP test takers are not motivated. This assertion has been supported by anecdote, the logic of the apparent incentives inherent in the NAEP 12th grade testing situation, or common wisdom, but it has not been supported by empirical evidence. Research on this topic has been inconclusive. Similarly, information from NAEP background questions has been inconclusive. For example, NAEP background questions asking 12th graders whether they tried hard when taking NAEP are consistently associated with

higher average scores for the students who say they didn't try hard and lower average scores for students who say they did try hard. Perhaps student proficiency and effort are being conflated in the analysis of the responses, but the data provide no evidence that achievement is underestimated by NAEP because of a lack of student motivation.

In 2009, the 12th grade students who took the NAEP reading and mathematics assessments answered 95% of the test questions, including the constructed response items that require students to do much more than merely fill in a bubble on a multiple choice answer sheet. With respect to the multiple choice questions, there was little if any evidence of "Christmas tree" or random responses to the questions, which would have been a sign that students were not seriously engaged in the test-taking task.

The correlations between performance on the high stakes SAT and the low stakes NAEP are additional evidence to consider. The correlation was .91 in comparing mathematics performance and .74 for reading. While these substantial correlations do not prove that the 12th grade NAEP test takers were motivated, they do not support the assertion that they are not motivated. Although it is logical to assume that the 12th grade students sitting for both tests may not have taken low stakes NAEP as seriously as the high stakes SAT to some degree, it is not possible to determine if this is true. And of course, a decrease of all of the NAEP scores by any consistent number of points would still yield the same correlations. However, the correlations do suggest that any diminution in motivation that might be present is not diminishing the effectiveness of the NAEP scores as indicators of academic preparedness for college.

4. The proposed test uses are appropriate and consequences are commensurate with intended uses.

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

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

Currently, there are either no or very low stakes consequences associated with the use of NAEP results. NAEP is not used as a basis for evaluating or diagnosing individual students, classroom or school performance, the effectiveness of individual teachers or administrators, or for any other accountability purpose. This will not change with the added meaning to the NAEP 12th grade

Proficient achievement levels for reading and mathematics afforded by the inferences proposed in this validity argument.

While the uses and consequences of NAEP will not change, the added meaning to NAEP Proficient at the 12th grade brings with it the potential for misinterpretation. These were discussed in detail on pages 5-6 above, and will be summarized here. NAEP reports should include text explaining the limitations on interpretation and other caveats that follow.

False Negatives

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

- likely to be academically prepared for first-year college courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading or mathematics in college,

but with a lower probability than those at or above Proficient.

Not a Preparedness Standard

The proposed inferences are not intended to represent or be used as standards for minimal academic preparedness for college.

Academically Prepared for College

The proposed inferences are intended to apply to placement policies affecting the typical degree-seeking entry-level college student at the typical college, not the admission policies. Thus, "academically prepared for first year college courses" refers to the reading and mathematics knowledge and skills needed for placement into entry-level, credit-bearing, non-remedial courses in broad access 4-year institutions and, for 2-year institutions, the general policies for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

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

Data Limitations

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

Preparedness for Job Training

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

Summary and Conclusion

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

Through a partnership with the College Board, performance on 12th grade NAEP was compared with performance on the SAT mathematics and critical reading assessments, with correlations of .91 and .74 respectively. Analyses of these data examined the average NAEP scores and inter-quartile ranges for students scoring "at" and "at or above" the College Board College Readiness Benchmarks for reading and mathematics. Similar analyses were conducted using data from the 2005 and 2009 NAEP High School Transcript Studies, using the college readiness benchmarks developed by ACT and by the College Board. A longitudinal study was conducted in partnership with the Florida Department of Education, following the 12th grade students in the state NAEP sample into postsecondary employing Florida's longitudinal data base. The average NAEP scores and interquartile ranges were calculated for the Florida students in relation to the ACT or SAT college readiness benchmarks, whether they achieved a first-year GPA of B- or better, and whether they were placed into a remedial course in their first year of college. The results of these analyses were consistent across studies and across years. In addition, indicators of the engagement in the NAEP test taking of 12th grade students in 2009 do not lend support to the assertion that NAEP 12th grade results are underestimates due to a lack of student motivation.

That the NAEP sampling, scaling and statistical procedures yield accurate estimates of the percentage of students scoring at or above a selected cut-score (i.e., NAEP achievement level) is well established as a result of numerous validity studies and evaluations.

Thus, the NAEP 12th grade preparedness research results support the inferences that students scoring at or above the Proficient achievement level on the 12th grade NAEP Reading or Mathematics Assessment are

- likely to be academically prepared for first year college courses
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading in college.

A substantial percentage of students whose scores are in the range between Basic and Proficient are likely to be academically prepared for college, but with a lower probability.

Including these inferences in NAEP 12th grade reports will add meaning to the interpretation of the NAEP achievement levels. However, steps must be taken to avoid potential misinterpretation. NAEP reports using these inferences must also include the limitations on interpretation and caveats described previously in this validity argument. In addition, the reports should explain the rationale for NAEP reporting on academic preparedness and describe appropriate and inappropriate uses of the results.

References

ACT Technical Manual; http://www.act.org/aap/pdf/ACT_Technical_Manual.pdf

ACCUPLACER on-line technical manual;
http://isp.southtexascollege.edu/ras/research/pdf/ACCUPLACER_OnLine_Technical_Manual.pdf

AERA/APA/NCME *Standards for Educational and Psychological Testing (1999)*

Allen, Jeff, Sconing, Jim (2005). *Using ACT Assessment Scores to Set Benchmarks for College Readiness (ACT Research Series 2005-3).* Iowa City, IA: ACT, Inc,

Bozick, R., and Lauff, E. (2007). *Education Longitudinal Study of 2002 (ELS:2002): A First Look at the Initial Postsecondary Experiences of the Sophomore Class of 2002* (NCES 2008-308). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

College Board. *The SAT® College and Career Readiness Benchmark User Guidelines; http://media.collegeboard.com/digitalServices/pdf/sat/12b_6661_SAT_Benchmarks_PR_120914.pdf*

Conley, D.T. (2007). *Toward a More Comprehensive Conception of College Readiness*. Eugene, OR: Educational Policy Improvement Center.

Fields, R. & Parsad, B. (2012). *Tests and Cut Scores Used for Student Placement in Postsecondary Education: Fall 2011.* Washington, DC: National Assessment Governing Board.

Kane, Michael T. (2013, Spring). *Validating the Interpretations and Uses of Test Scores.* Journal of Educational Measurement, Vol. 50, No. 1, pp. 1-73.

Kim, Y.K., Wiley, A., and Packman, S., *National Curriculum Survey on English and Mathematics*, College Board Research Report 2011-13 (New York: The College Board, 2011) http://professionals.collegeboard.com/data-reports-research/cb/RR2011-13

National Assessment Governing Board. (2008, September). *Mathematics Framework for the 2009 National Assessment of Educational Progress.* Washington, DC: Author.

National Assessment Governing Board. (2008, September). *Reading Framework for the 2009 National Assessment of Educational Progress.* Washington, DC: Author

National Assessment Governing Board. (2009, June). *Making New Links: 12th Grade and Beyond.* Washington, DC: Technical Panel on 12th Grade Preparedness Research.

National Center for Education Statistics (2010).*The Nation's Report Card: Grade 12 Reading and Mathematics 2009 National and Pilot State Results* (NCES 2011–455). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Commission on NAEP 12th Grade Assessment and Reporting. (2004, March 5). *12th Grade Student Achievement in America: A New Vision for NAEP.* Washington, DC: Author.

Ross, T., Kena, G., Rathbun, A., KewalRamani, A., Zhang, J., Kristapovich, P., and Manning, E. (2012). *Higher Education: Gaps in Access and Persistence Study* (NCES 2012-046). U.S. Department of Education, National Center for Education Statistics. Washington, DC: Government Printing Office.

WestEd ACCUPLACER Mathematics Report; http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACCUPLACER-NAEP_Mathematics_Content_Comparison.pdf.

WestEd ACCUPLACER Reading Report; http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACCUPLACER-NAEP_Reading_Content_Comparison.pdf.

WestEd SAT Mathematics Report; http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/SAT-NAEP_Mathematics_Content_Comparison.pd.f

WestEd SAT Reading Report; http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/SAT-NAEP_Reading_Content_Comparison.pdf.

Wyatt, Jeffrey, Kobrin, Jennifer, Wiley, Andrew, Camara, Wayne J., and Proestler, Nina (2011). *SAT Benchmarks: Development of a College Readiness Benchmark and its Relationship to Secondary and Postsecondary School Performance (Research Report 2011-5).* New York, NY: College Board. http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-5-sat-college-readiness-benchmark-secondary-performance.pdf

NAEP 12th Grade Preparedness Research

Based on the Program of Preparedness Research adopted by the Governing Board in March 2009, four categories of research studies were conducted to produce evidence to develop and support the validity of statements for NAEP reporting on the academic preparedness in reading and mathematics of 12th grade students for college and job training.

- content alignment studies;
- statistical relationship studies;
- judgmental standard setting studies; and
- surveys

Additionally, the Texas Commissioner of Higher Education offered the opportunity to conduct a benchmarking study with Texas higher education institutions, and a pilot study to examine the feasibility was conducted.

The research studies completed to date are available in an online technical report. In addition, the NAEP 12th Grade Preparedness Commission is planning a symposium in Washington, DC for Summer 2013 focused on the Board's preparedness research results and the Phase 2 research plans.

The following informational attachments are provided:

- Updates related to the Board's Course Content Analysis Research:
  - College Course Content Analysis Progress Update (Attachment F-1) .....Page F2
  - Job Training Program Final Report: Executive Summary (Attachment F-2)Page F16

Additionally, the following attachments are provided for reference:

- Proposed research projects for phase 2 of the Board's preparedness research program (Attachment F-3).....Page F22
  - National and State Partnerships
  - Research with Frameworks
- Background materials describing each study category (Attachment F-4) .....Page F24

Attachment F-1
College Course Content Analysis Progress Update

In September 2012, the Governing Board awarded a contract to the Education Policy Improvement Center (EPIC) to conduct research on entry level non-remedial college course content in order to (1) identify the prerequisite knowledge and skills in reading and mathematics for entry-level college courses and (2) determine the extent to which there is a match with the content of grade 12 NAEP reading and mathematics assessments. This project addresses academic preparedness for college only—a separate parallel research project addresses preparedness for job training (described below).

In this project, EPIC will determine the entry-level (introductory) credit-bearing courses most frequently taken by entering students that are reflective of college-level reading and mathematics demands and that satisfy general education requirements. These introductory courses should have no college-level prerequisite course requirements, and only non-remedial courses that satisfy general education requirements should be included in the analysis. Further, in cases where multiple versions of a course are offered for majors and non-majors, only the course for non-majors should be included.

Using course artifacts for a generally representative sample of institutions, EPIC will analyze the introductory course artifacts for commonalities and differences in the reading and mathematics prerequisites needed to qualify for placement into the course. From these analyses, EPIC will develop descriptions of the knowledge, skills, and abilities (i.e., the prerequisite KSAs) needed for students to qualify for placement into the introductory course, based on an analysis of the course artifacts. And as part of a set of comparative analyses, EPIC will then use these descriptions to review:

- the description of minimal requirements for placement into college-level coursework as developed in the NAEP preparedness judgmental standard setting (JSS) research
- KSAs represented by 2009 grade 12 items that map to the NAEP scale with a response probability of .67 and fall within the range of cut scores set by the two replicate panels in the JSS research
- 2009 and 2013 grade 12 NAEP items
- the KSAs represented by 2009 items that map in the range of the NAEP score scale from the the Basic level through the Proficient level; and
- the NAEP achievement level descriptions.

A progress report is attached with more details on the project and a description of work completed to date.

Attachment F-2
Job Training Program Content Analysis Final Report

In October 2011, the Governing Board began work with WestEd and its subcontractor, the Education Policy Improvement Center (EPIC), to conduct follow-up research relative to the NAEP preparedness judgmental standard setting (JSS) research, wherein panelists reviewed NAEP questions and made judgments about the content knowledge needed by minimally prepared students. The research results from this project are intended to supplement the JSS research findings by providing a clearer understanding of the knowledge and skills required for entry- and exit-level coursework in designated occupational programs. By reviewing course artifacts such as syllabi, text books, and assignments, this study will help to determine if the knowledge, skills, and abilities (KSAs) required of students in the training programs are appropriately represented by the borderline preparedness descriptions (developed in the JSS research), by all the items on the 2009 NAEP, and by the 2009 NAEP items in the scale score ranges identified by panelists in the JSS research project.

Attached is the executive summary for the final report detailing the results of this research project.

Attachment F-3
Phase 2 Academic Preparedness Research Plans

Continued research plans call for NAEP-SAT, NAEP-ACT, and NAEP-EXPLORE statistical linking studies, more research partnerships with states, analysis of course content prerequisites for job training programs and freshman college courses, and efforts to partner with experts in military occupational training. A summary of each proposed research study follows. At the November 2012 Board meeting, COSDAM began discussion on these research plans.

National and State Statistical Linking Studies with the SAT and with the ACT

In 2013, the Governing Board will partner again with the College Board, as it did in 2009, 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 2013 will be matched, and through this match, the linking will be performed. A similar study at the national level is planned in partnership with ACT, Inc.

In addition, the state-level studies, begun in 2009 with Florida, will be expanded in 2013. Again using a procedure that protects student confidentiality, the postsecondary activities of NAEP 12th grade test takers in the state samples in partner states will be followed for up to five years using the state longitudinal data bases. Five states will be partners in these studies: Florida, Illinois, Massachusetts, Michigan, and Tennessee. These studies will examine the relationship between 12th grade NAEP scores and GPA, placement into remedial versus credit-bearing courses, and scores on admissions and placement tests.

May 2013 Update: Data sharing agreements are in development for each state partner.

Statistical Linking of Grade 8 NAEP and 8th Grade 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 two states, KY and TN. The objective is to determine the feasibility of identifying the point on the NAEP scales that indicate students are "on track" for being academically prepared for college and job training by 12th grade. As a foundation for the linking study, content alignment studies between 8th grade NAEP reading and mathematics and 8th grade EXPLORE would also be conducted as a part of the planned partnership with Act, Inc.

May 2013 Update: No updates at this time.

Evaluation of NAEP Frameworks and Item Pools

The Governing Board is conducting a procurement (1) to design a comprehensive and multi-method evaluation of the grade 12 NAEP frameworks and item pools in both reading and mathematics as measures of academic preparedness for college and job training; and (2) based on the evaluation, to produce specific recommendations for changes that may be

needed to further refine 12th grade NAEP in reading and mathematics as a measure of academic preparedness for college and to determine the extent to which changes would be needed to make 12th grade NAEP in reading and mathematics a valid measure of academic preparedness for entry into job training programs that require at least three months of post-secondary training, but not a bachelor's degree in college.

The review of the 12th grade reading and mathematics frameworks by Achieve, Inc. in 2005 and 2006 led to changes in the frameworks for the 2009 assessments intended to measure 12th grade academic preparedness for college and job training. The content alignment studies between 12th grade NAEP reading and mathematics and the SAT and ACT college admissions tests in reading and mathematics tests found a high degree of overlap in content widely recognized as representing academic preparedness for college. The content alignment study with WorkKeys, as well as the Judgmental Standard Setting studies for job training, surfaced questions about the capacity of the current 12th grade NAEP to measure academic preparedness for job training. The planned evaluation is part of the continuing program of preparedness validity research.

In this procurement, the Board seeks innovative, practicable design proposals for evaluations that will provide the foundation needed to make valid statements about academic preparedness.

May 2013 Update: The procurement process is ongoing.

Research Design Proposals for NAEP and Academic Preparedness for Job Training

Reporting on academic preparedness for college and job training is a challenging and important new direction for NAEP. Hence, the Governing Board is also conducting a procurement to seek proposals for research designs and studies that are feasible. The objective of the research is to advance the Governing Board's efforts to identify locations on the 12th grade NAEP reading and mathematics scales that represent the knowledge and skills to qualify for training in various occupations.

May 2013 Update: The procurement process is ongoing.

Attachment F-4
Overview of the Types of NAEP Preparedness Research

As part of the ongoing updates to COSDAM, the following is a summary of each research study category from phase 1 of the Board's program of research for reporting academic preparedness.

Content Alignment Studies

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.

A design document was developed by Dr. Norman Webb for the NAEP preparedness research alignment studies, and this design was implemented for the studies of the 2009 NAEP with the SAT and ACUPLACER in reading and mathematics. This design, with minor modifications, has also been used for the alignment of the 2009 NAEP with WorkKeys tests in these subject areas.

Content alignment studies for the first phase of the Board's Program of Preparedness Research have been completed for NAEP in reading and in mathematics with WorkKeys, the SAT, and ACCUPLACER. In addition, a content alignment study was designed and conducted by ACT for the ACT and NAEP in reading and mathematics before the content alignment design document was developed.

Studies to Establish Statistical Relationships

Highest priority has generally been placed on these studies. Currently, two main sets of studies have been conducted under this heading. One set addresses ***statistical linking*** of NAEP with other assessments, and the other set examines ***longitudinal data*** for NAEP examinees.

For statistical linking, there has been a study to relate SAT scores in reading and in mathematics to the national sample of NAEP scores for grade 12. The objective was to provide a statistical linking of SAT and NAEP scores for all students in the 2009 grade 12 NAEP who had taken the SAT by June 2009. ETS staff reported that the match rate of approximately 33% of NAEP scores to SAT scores compares favorably to the national SAT participation rate of approximately 36% of public school students. The final sample used for linking the NAEP reading and SAT critical reading included approximately 16,200 students. For NAEP and SAT mathematics, the linking sample included approximately 15,300 students.

For longitudinal data, a series of analyses were conducted to examine statistical relationships for Florida's NAEP examinees. NAEP's 2009 state-representative sample of Florida 12th graders was used to match NAEP scores for reading and mathematics to student scores on several tests

collected by the Florida Department of Education (FLDOE). The data sharing agreement with FLDOE provides access to scores for the SAT, ACCUPLACER, and WorkKeys. Additionally, ACT, Inc. has given permission to the Florida Department of Education to share ACT scores with the Governing Board for purposes of conducting the grade 12 preparedness research. We also plan to obtain employment data and salary data for Florida examinees, but access to those data was not included under the current data sharing agreement. A plan to allow for electronic transfer of data was developed to keep secure the identity of students, consistent with the NAEP legislation, FLDOE requirements, and requirements of each assessment program.

Records for roughly half of the Florida grade 12 NAEP examinees in 2009 could be matched to an ACT score and half to an SAT score. This match rate is consistent with other data for Florida students. The match of WorkKeys scores to the total 2009 state NAEP sample of 12th graders was only about 6%. FLDOE reported that around 89,300 Florida 12th graders were enrolled in vocational-technical programs in school year 2008-09. The match of WorkKeys examinees to NAEP examinees was not sufficient to warrant additional analyses for the 2009 cycle. The state of Florida has only recently implemented the testing of high school students in vocational programs with the WorkKeys exam, and we anticipate that the number of examinees will increase in subsequent years.

Judgmental Standard Setting Studies

A series of judgmental standard setting studies was planned to produce preparedness reference points on the NAEP scale for entry into job training programs and for placement in college credit-bearing courses. Within this category of studies, the Technical Panel for 12th Grade Preparedness Research placed highest priority on the judgmental studies related to preparedness for job training programs in 5-7 exemplar jobs. This priority is largely related to the paucity of national data available for statistical studies in these areas. The Governing Board has not assumed that academic preparedness for college and for job training are the same. Rather, our studies are aimed at determining the level of performance on NAEP that represents the reading and mathematics knowledge and skills needed to qualify for job training programs for each of the occupations included in our research studies and for placement in credit-bearing college courses that fulfill general education requirements for a bachelor's degree.

In order to maximize the standardization of judgmental standard setting (JSS) studies within and across post-secondary areas, a design document was developed to specify the number of panelists, the eligibility criteria for panelists, the procedures for drafting and finalizing borderline performance descriptions, the methodology to be implemented, feedback to be provided, key aspects to be evaluated, and reports to be produced. The methodology and basic procedures specified for the design of these studies were those implemented for the achievement levels-setting process for the 2006 grade 12 economics NAEP and for the 2009 science NAEP for grades 4, 8, and 12.

The five exemplar jobs approved by COSDAM for inclusion in these studies are as follows:

1. automotive master technicians
2. computer support specialists
3. heating, ventilation, and air conditioning technicians

4. licensed practical nurses
5. pharmacy technicians

A pair of replicate panels with 10 panelists each was convened for each subject and post-secondary area for a total of 24 operational panels.

Higher Education Survey

A survey of two-year and four-year post-secondary institutions was conducted in Fall 2011 to gather information regarding (1) the placement tests used and (2) the cut scores on those tests in reading and mathematics below which need was indicated for remedial/developmental courses in reading and mathematics, and at or above which placement in credit-bearing entry level courses was indicated. The sample of accredited postsecondary education institutions was nationally representative. A weighted response rate of 81% was achieved.

Benchmarking Studies

Benchmarking studies in the preparedness research context are studies in which NAEP is administered to groups of interest, e.g., college freshmen enrolled in credit-bearing college level courses that fulfill general education requirements for a four-year degree without the need for remediation. Determining the average NAEP performance of this group would then provide a "benchmark" score that can be considered as one of the reference points on the NAEP scale. A benchmarking study in combination with reference points from other studies in the Program of Preparedness Research can assist the Board in determining the areas of the NAEP scale that indicate preparedness. A benchmarking study of Texas college freshmen was planned, and it had the support of the Texas Commissioner of Higher Education and the cooperation of nine Texas higher education institutions. A small scale pilot study to evaluate the feasibility of the study design was implemented.

The Governing Board and the National Center for Education Statistics (NCES) collaborated on the implementation of this small scale pilot study, which was carried out by Westat, the NAEP sampling and administration contractor to NCES. The data collection phase for the pilot ended on October 15, 2010. Of the eligible sample of 1,234 students, 255 actually attended a NAEP session, for an overall response rate of 20.7 percent. As announced at the November 2010 meeting of COSDAM, NCES, Westat, and Governing Board staff met to discuss alternatives. Board staff decided that we will not proceed to the operational phase of this study due to low participation rates and the lack of feasible alternatives to increase participation.

No additional benchmarking studies are planned for the 2009 NAEP preparedness research.

Overview of Referenced Assessments

For additional background information, the following list presents a brief description of the assessments that the Technical Panel on 12th Grade Preparedness Research recommended for analysis in NAEP preparedness research. Many of these assessments are the primary focus of the proposed content alignment studies and statistical relationship 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.

- ACCUPLACER – ACCUPLACER is a computer adaptive test used for college course placement decisions in two-year and four-year institutions. It is produced by the College Board and includes assessments of sentence skills, reading comprehension, arithmetic, elementary algebra, college level math, and written essays.
- 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 WorkKeys –WorkKeys is a workplace focused set of tests that assess knowledge and skills in communication (business writing, listening, reading for information, writing) as well as problem solving (applied technology, applied mathematics, locating information, observation). There is also an interpersonal skills section of WorkKeys.
- 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 College Board has provided SAT score data to be used in research studies to establish a statistical relationship between the SAT and NAEP.

College Course Content Analysis Study for NAEP Preparedness Research
Progress Update

Submitted by
Educational Policy Improvement Center (EPIC)

INTRODUCTION AND BACKGROUND

The College Course Content Analysis (CCCA) study is one of a series of studies contributing to 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 will answer 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 (i.e., planning and analysis efforts relative to the 2013 grade 12 NAEP reading and mathematics assessments)?

The final report is due May 2014, and until then COSDAM will receive detailed reports at each Board meeting.

---

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).

METHODOLOGY

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

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

The study is well into the first phase of the identification and collection of course artifacts and the development of instruments to collect data. Plans for the second phase and third phases of the study are in the design document, which includes independent and group reviews of the course artifacts and related analyses.

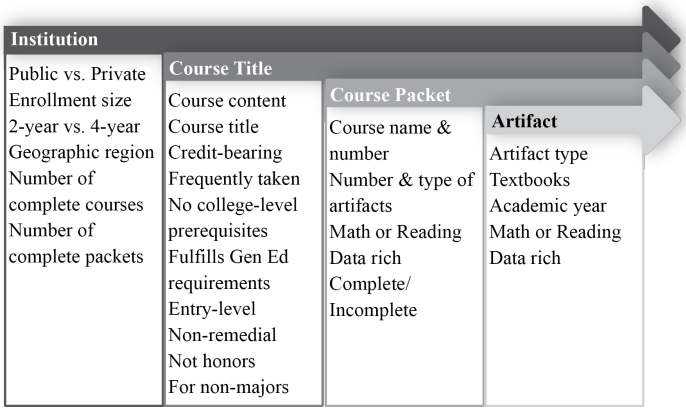
Phase 1: Identification and collection of course artifacts

In the CCCA study, a *course artifact* is defined as a syllabus, assignment or assessment, and textbook excerpt. 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 extant artifacts compiled during previous research on entry-level curricula at postsecondary educational institutions.

In the first phase, project staff identified extant artifacts from EPIC's repository of course data that met the requirements of the CCCA study and evaluated the representativeness and sufficiency of the artifacts, course titles, and institutions represented in the extant artifact bank. Project staff will also solicit new course artifacts as needed to create a complete and 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 will comprise a set of courses and artifacts that will be used as the basis for the content review. The Artifact Inventory, a report of extant artifacts contributing to the CCCA Artifact Bank, and the Sampling Gap Analysis, a report of the discrepancy between applicable extant artifacts and the set of required artifacts necessary to meet study objectives, have been completed. Illustration 1 represents the coding scheme that will be applied to all courses and artifacts prior to inclusion into the CCCA Artifact Bank.

Illustration 1: CCCA Inclusion Criteria (Artifact Metadata Coding Scheme)



At the conclusion of artifact collection, the CCCA Artifact Bank will include all relevant artifacts compiled into course packets to be reviewed by mathematics and reading content review teams in the next phase of the study.

Phase 2: Review of course artifacts by Review Teams

The CCCA study's application of convergent consensus combines independent individual judgments with panel consensus processes. This two-part approach allows the capture and integration of responses from two types of participants: those with a conceptual understanding of the mathematics or reading knowledge and skills required in entry-level college courses, experience in teaching these types of courses, and training and experience in the EPIC methodology of coding artifacts; and those with content expertise, experience in college-level teaching, and extensive experience in the development of NAEP frameworks, assessments, and preparedness research projects for the Governing Board.

In Phase 2, preparatory work for the content reviews will be conducted by project staff including recruitment of the content reviewers, convening of a NAEP advisory panel, and finalizing the coding scheme and initial decision rules. The training materials and an initial set of content review decision rules will be reviewed by an advisory panel of experts on the NAEP frameworks. Guidance from this NAEP Advisory Panel will be integrated into the decision rules prior to training.

Qualified content reviewers will review the course packets and identify prerequisite mathematics and reading KSAs. The NAEP frameworks for grade 12 reading and math will be used as a set of foundational KSAs. If additional KSAs are identified during reviews, the new KSAs will be documented and included in all successive reviews, comparisons and data analyses. EPIC facilitators, trained by CCCA project staff, will facilitate the group coding review process.

The decision rules will be applied in a consistent structured manner throughout the study and it is anticipated that there will be very few, if any changes. Content reviewers will have the opportunity to comment on, and suggest changes to the rules during training, after training, during independent review and after independent review. After the independent reviews are completed, no changes to decision rules will be made. The content review will include both individual and group coding of course packets. The end product of this review process will be a comprehensive list of prerequisite KSAs.

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 and metrics necessary will be employed to monitor and demonstrate validity and reliability, and both conceptual (information processing/document analysis) and technical (quantitative) analyses will be conducted.

In this phase of the study, NAEP experts will review and summarize the prerequisite KSAs and conduct several comparisons with respect to NAEP. The experts will compare and contrast the college prerequisite KSAs with the NAEP framework and the results of other NAEP preparedness research. Project staff will work to support the NAEP experts in their review process by orienting them to the CCCA project goals and providing data necessary for their work. The results will be narrative summaries of the prerequisite knowledge, skills, and abilities in mathematics and reading. Project staff will share these summaries with the content reviewers to collect their feedback on whether the summaries appropriately capture their judgments. Project staff will conduct summary analyses and write a final report on the CCCA study.

Illustration 2 describes the project design. As project elements are completed, they are shaded in dark gray. Project elements that have begun and are in progress are shaded in a lighter gray. Those project elements in the future have no shading in the diagram.

Illustration 2: Project Design

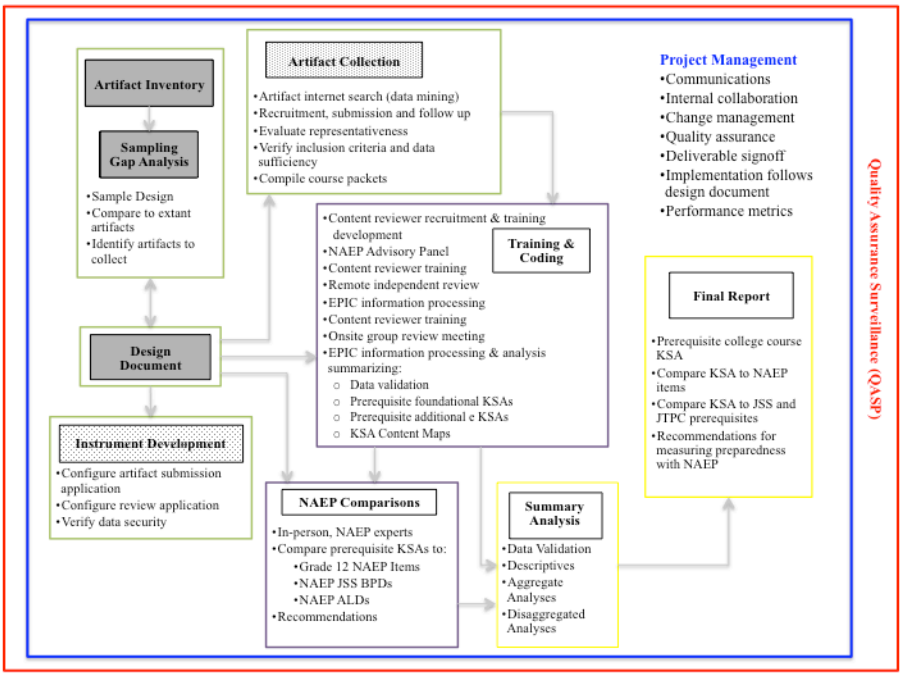


Illustration 3 displays a draft schedule of the study.

Illustration 3: CCCA Study Gantt chart

	Start Date	End Date	Duration	QUARTER 2013			QUARTER 3 2013			QUARTER 4 2013			QUARTER 1 2014			FINAL
				APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR
NAEP Technical Panel Meeting	28-May	31-May	3 days													
Facilitator Training	10-Jun	14-Jun	4 days													
Math Content Training/Orientation	17-Jun	21-Jun	4 days													
Read Content Training/Orientation	17-Jun	21-Jun	4 days													
Independent Content Reviews	24-Jun	19-Jul	3+ weeks													
EPIC Data Analysis Period 1	22-Jul	2-Aug	1.5 week													
Math Content Review Meeting	3-Aug	11-Aug	1+ week													
Read Content Review Meeting	3-Aug	11-Aug	1+ week													
EPIC Data Analysis Period 2	12-Aug	11-Oct	2.5 months													
NAEP Expert Math Review Meeting	14-Oct	18-Oct	4 days													
NAEP Expert Read Review Meeting	11-Nov	15-Nov	4 days													
EPIC Data Analysis Period 3	18-Nov	31-Jan	10.5 weeks													
Final Report Writing and Review	3-Feb	28-Mar	7.5 weeks													
Board Review and Presentation	1-Apr	30-Apr	1 month													
Final Report Deliverable Due	30-Apr	30-Apr	Final Deliver													

PROGRESS UPDATE

Identification and Collection of Course Artifacts (Phase 1)

EPIC drew on previous research to select course titles that meet the study's inclusion criteria and were likely to contain sufficient mathematics and reading content to identify the prerequisite KSAs. Table 1 contains the finalized list of entry-level courses to be included in the CCCA study.

Table 1: Course Titles Included in the CCCA Study

Mathematics	Reading
College algebra	English literature
Finite math	Introduction to psychology
Introduction to calculus/Precalculus	U.S. government/Introduction to political science
Statistics	U.S. history

The content analysis of artifacts from a sample of 20 courses per course title is sufficient to reach redundancy of prerequisite knowledge, skills and abilities within a course title. In total, 160 course packets will be the basis of findings for prerequisite KSAs (half of these for reading and the other half for mathematics). Additional course packets will be collected to serve as validation packets to assess coding consistency within each course title, and for reviewer training and qualification.

Table 2 and Table 3 are summaries of the characteristics of institutions from which project staff will recruit to complete partial packets and submit new packets for each course title.

Table 2: Institutional Characteristics of Initial Sample Targeted in Sampling Plan for Mathematics

Characteristic	College algebra (*N* 19)	Finite math (*N* 20)	Introduction to calculus/precalculus (*N* 19)	Statistics (*N* 19)	Mathematics Overall (*N* 76)
Program type					
2-Year	37%	35%	39%	37%	37%
4-Year	63%	65%	61%	63%	63%
Size					
Small	68%	64%	61%	63%	65%
Medium	21%	17%	22%	16%	18%
Large	11%	19%	17%	21%	17%
Control					
Public	53%	53%	56%	58%	53%
Private not-for-profit	47%	47%	44%	42%	47%
Geographic Region					
West	16%	15%	17%	16%	17%
Midwest	26%	25%	22%	26%	26%
East	26%	24%	28%	26%	25%
Southeast	26%	25%	22%	26%	25%
Southwest	6%	8%	11%	6%	7%

Table 3: Institutional Characteristics of Initial Sample Targeted in Sampling Plan for Reading

Characteristic	English literature (*N* 14)	Introduction to psychology (*N* 1)	U.S. government (*N* 6)	U.S. history (*N* 8)	Reading Overall (*N* 29)
Program type					
2-Year	43%	100%	66%	75%	46%
4-Year	57%	*0%*	34%	25%	54%
Size				6	
Small	57%	100%	67%	63%	66%
Medium	14%	0%	33%	0%	15%
Large	29%	0%	0%	37%	19%
Control				50%	
Public	64%	100%	100%	50%	63%
Private not-for-profit	36%	0%	0%	50%	37%
Geographic Region				52	
West	43%	0%	17%	25%	17%
Midwest	14%	0%	0%	25%	3%
East	14%	100%	33%	25%	48%
Southeast	22%	0%	33%	25%	24%
Southwest	7%	0%	17%	0%	8%

Design Document and Review of the Course Artifacts (Phase 2)

A major project milestone completed since the March 2013 Governing Board meeting was the finalization of the design document, which will guide the project in order to address the core research questions of the CCCA study. Project staff will use standard statistical methods and metrics to monitor and demonstrate reliability and support validity. The CCCA design adapts the basic expert-judgment model by developing and employing explicit decision-making criteria that inform the judgments experts make throughout the process of identifying KSAs within the course artifacts. All judgments are justified in relation to a set of established decision rules.

Reviewers work independently during the first round of review. A second round of review is conducted at a face-to-face meeting, reviewing those course packets where universal agreement was not achieved. At the group review, the course artifacts are compared and discussed to provide an opportunity for reviewers to reach agreement. The reviewers refer to decision rules and evidence in the artifacts being reviewed to justify their decisions. EPIC refers to this process as *convergent consensus*.

This approach is built upon the principles of evidence-centered design. Evidence-centered approaches are particularly useful in situations where the validity of the final product is a prime consideration, as in the case of this study. As adapted here, the evidentiary process requires expert reviewers to adhere to explicit decision criteria and to be able to justify all decisions with reference to specific evidence.

Tables 4 and 5 illustrate the review processes for the Mathematics and Reading independent content reviews, while Tables 6 and 7 illustrate the processes for the group content reviews.

Table 4: Independent Review—Mathematics Course Packet Assignments

Content Review Team*	College algebra	Finite math	Introduction to calculus/precalculus	Statistics	# of Packets Rated
Team M1 **3 Reviewers** **+ 1 Alternate**	Packets 1–5 Validation Packets 1-2	Packets 1–5 Validation Packets 3-4	Packets 1–5 Validation Packets 5-6	Packets 1–5 Validation Packets 7-8	28
Team M2 **3 Reviewers** **+ 1 Alternate**	Packets 6–10 Validation Packets 1-2	Packets 6–10 Validation Packets 3-4	Packets 6–10 Validation Packets 5-6	Packets 6–10 Validation Packets 7-8	28
Team M3 **3 Reviewers** **+ 1 Alternate**	Packets 11–15 Validation Packets 1-2	Packets 11–15 Validation Packets 3-4	Packets 11–15 Validation Packets 5-6	Packets 11–15 Validation Packets 7-8	28
Team M4 **3 Reviewers** **+ 1 Alternate**	Packets 16–20 Validation Packets 1-2	Packets 16–20 Validation Packets 3-4	Packets 16–20 Validation Packets 5-6	Packets 16–20 Validation Packets 7-8	28
	22	22	22	22	88*

***** Total is 88, not 104, because validation packets are repeated across teams.

Table 5: Independent Review—Reading Course Packet Assignment

Content Review Team*	English literature	Psychology	U.S. government	U.S. history	# of Packets Rated
Team R1 **3 Reviewers** **+ 1 Alternate**	Packets 1–5 Validation Packets 1-2	Packets 1–5 Validation Packets 3-4	Packets 1–5 Validation Packets 5-6	Packets 1–5 Validation Packets 7-8	28
Team R2 **3 Reviewers** **+ 1 Alternate**	Packets 6–10 Validation Packets 1-2	Packets 6–10 Validation Packets 3-4	Packets 6–10 Validation Packets 5-6	Packets 6–10 Validation Packets 7-8	28
Team R3 **3 Reviewers** **+ 1 Alternate**	Packets 11–15 Validation Packets 1-2	Packets 11–15 Validation Packets 3-4	Packets 11–15 Validation Packets 5-6	Packets 11–15 Validation Packets 7-8	28
Team R4 **3 Reviewers** **+ 1 Alternate**	Packets 16–20 Validation Packets 1-2	Packets 16–20 Validation Packets 3-4	Packets 16–20 Validation Packets 5-6	Packets 16–20 Validation Packets 7-8	28
	22	22	22	22	88*

***** Total is 88, not 104, because validation packets are repeated across teams.

Table 6: Mathematics Group Review—Participant and Course Packet Assignments

Participants	Day 1	Day 2	Day 3	Packets Reviewed (Total Count)
Team M1 (3) **EPIC Facilitator (1)** **EPIC Scribe (1)**	Orientation (a.m.) Statistics (p.m.)	Finite math (a.m.) Intro calculus/precalculus (p.m.)	College Algebra (a.m.) Debriefing (p.m.)	Course packets 1–5 Validation packets 1–8 (28)
Team M2 (3) **EPIC Facilitator (1)** **EPIC Scribe (1)**	Orientation (a.m.) Intro calculus/precalculus (p.m.)	Statistics (a.m.) College Algebra (p.m.)	Finite math (a.m.) Debriefing (p.m.)	Course packets 6–10 Validation packets 1–8 (28)
Team M3 (3) **EPIC Facilitator (1)** **EPIC Scribe (1)**	Orientation (a.m.) College algebra (p.m.)	Intro calculus/precalculus (a.m.) Finite math (p.m.)	Statistics (a.m.) Debriefing (p.m.)	Course packets 11–15 Validation packets 1–8 (28)
Team M4 (3) **EPIC Facilitator (1)** **EPIC Scribe (1)**	Orientation (a.m.) Finite math (p.m.)	College algebra (a.m.) Statistics (p.m.)	Intro calculus/precalculus (a.m.) Debriefing (p.m.)	Course packets 16–20 Validation packets 1–8 (28)
Reviewers = 12 **EPIC Staff = 8** **NAEP Expert = 1** **Meeting Spaces = 4**	Packets/Team = 7	Packets/Team = 14	Packets/Team = 7	Total Packets (88)

Table 7: Reading Group Review—Participant and Course Packet Assignments

Participants	Day 1	Day 2	Day 3	Packets Reviewed (Total Count)
Team R1 (3) **EPIC Facilitator (1)** **EPIC Scribe (1)**	Orientation (a.m.) Literature (p.m.)	Psychology (a.m.) U.S. history (p.m.)	U.S. government (a.m.) Debriefing (p.m.)	Course packets 1–5 Validation packets 1–8 (28)
Team R2 (3) **EPIC Facilitator (1)** **EPIC Scribe (1)**	Orientation (a.m.) U.S. history (p.m.)	U.S. government (a.m.) Psychology (p.m.)	Literature (a.m.) Debriefing (p.m.)	Course packets 6–10 Validation packets 1–8 (28)
Team R3 (3) **EPIC Facilitator (1)** **EPIC Scribe (1)**	Orientation (a.m.) U.S. government (p.m.)	U.S. history (a.m.) Literature (p.m.)	Psychology (a.m.) Debriefing (p.m.)	Course packets 11–15 Validation packets 1–8 (28)
Team R4 (3) **EPIC Facilitator (1)** **EPIC Scribe (1)**	Orientation (a.m.) Psychology (p.m.)	Literature (a.m.) U.S. government (p.m.)	U.S. history (a.m.) Debriefing (p.m.)	Course packets 16–20 Validation packets 1–8 (28)
Content Reviewers = 12 **EPIC Staff = 8** **NAEP Expert = 1** **Meeting Spaces = 4**	Packets/Team = 7	Packets/Team = 14	Packets/Team = 7	Total Packets (88)

The CCCA design has planned for the NAEP experts to consult on the preparation of the training and qualifying packets. During training, reviewers will code two training packets and must accurately code a third qualifying packet. If they are inconsistent with the NAEP experts in the application of the coding scheme, they will be retrained by project staff and receive an additional qualifying packet to review.

The CCCA design has embedded validity checks within the process to evaluate the reliability of the review team coding. Project staff will create two validation packets for each of the four course titles. These validation packets will look like any other course packet and will be mixed in with the others during the independent and group reviews. The content reviewers will not know which packets are the validation packets. The NAEP experts will code those packets and their coding will serve as a reference. Project staff will report the percent agreement between the four review teams' group consensus coding on the validation packets and the reference coding as reliability evidence. Project staff will also report the agreement of group consensus coding by the four review teams within each course title. The agreement statistic will be calculated using the same method.

The CCCA design mixes the course title packets across each review team. Each team will review five course packets from each of the course titles plus the two validation packets for that course title, which will be reviewed by every reviewer. Project staff will calculate a one-way analysis of variance on the KSAs and a non-significant intraclass correlation coefficient for the course title factor will indicate that the coding within courses are not sufficiently different to prevent interpreting the results across course titles.

Design Document and Analysis and Reporting (Phase 3)

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.

Project staff will examine the consensus coding of each of the validation packets across review teams to see if there is excessive error variation. The CCCA study will use percent agreement as an index of coding consistency because the statistic is directly interpretable; i.e., consistency near 100% is the goal. If review team coding consistency is high on the validation packets and there is no evidence that coding within course titles is significantly divergent, then project staff will conclude that the accumulated list of prerequisite KSAs is a reliable indication of the mathematics and reading that students need to be prepared for entry-level college courses.

Reliability and the Study Design (Phases 1, 2, and 3)

To summarize, the evidence of reliability in this study will begin with the representativeness of the colleges sampled and the completeness of the course packets collected. Project staff will

evaluate the results of the convergent consensus process first by confirming that the qualifications of the reviewers who complete the KSA coding meet the requirements set for the study. Evidence for the reliability of the independent review process will include information from the reviewers on their experience with the review and confirmation that they all completed the review tasks. Project staff will collect similar information during the group review to check that all the reviewers are sufficiently engaged and contributing to the convergent consensus process. When reviewers have completed their work and there is evidence that the convergent consensus process has successfully engaged the reviewers, project staff will conclude that the consensus coding of the KSAs are reliable judgments on the course packets.

National Assessment Governing Board

National Assessment of Educational Progress Grade 12 Preparedness Research Project Job Training Programs Curriculum Study

Submitted: March 29, 2013

Executive Summary

Submitted to:
National Assessment Governing Board
80 North Capitol Street, NW, Suite 825
Washington, DC 20002-4233
Phone: 202.357.6938

This study was funded by the
National Assessment Governing Board
under Contract ED-NAG-10-C-0004.

Submitted by:
WestEd
73 Harrison Street
San Francisco, CA 94107
Phone: 415.615.3400

Prepared by:
The Educational Policy Improvement Center
170 Millrace Drive
Eugene, OR 97403
Phone: 877.766.2279

WestEd

Executive Summary

Overview

This report describes the Job Training Programs Curriculum Study, which was commissioned by the National Assessment Governing Board (Governing Board) to analyze the content of course materials from five job training occupational areas and is part of a larger program of preparedness research projects that are being conducted for the Governing Board.

This study was designed to identify the knowledge, skills, and abilities (KSAs) that are prerequisite to and taught in entry-level job training programs; to describe the KSAs expected at the conclusion of job training programs; and to compare the prerequisite KSAs identified through analysis of job training course materials to KSAs identified as part o the Judgmental Standard Setting (JSS) study (WestEd & Measured Progress, 2011; Measured Progress & WestEd, 2012).

Artifacts from such courses included syllabi, assignments, examinations, reading/textbook passages, and textbook tables of contents. Teams of mathematics and reading content-area experts and occupational course instructors employed a convergent consensus model to analyze the artifacts in order to identify the minimal knowledge and skills required of students entering the courses. The KSAs identified were analyzed for job training programs within five occupational areas. These areas are: Automotive Master Technician; Computer Support Specialist; Heating, Ventilation, and Air Conditioning (HVAC); Licensed Practical and Licensed Vocational Nurse (LPN), and Pharmacy Technician (entry and concluding course level). Reviewers completed artifact analyses independently, and then were brought together in small groups to discuss the codings where they disagreed in order to resolve differences and reach consensus on the KSAs that are prerequisite for each of the job training programs.

The comprehensive mathematics and reading frameworks provided by the National Assessment of Educational Progress (NAEP) outline the structure for defining the KSAs. However, because the goal of the study was to identify all KSAs required of students entering job training programs, the analysis was not limited to the NAEP objectives; reviewers also recorded non-NAEP KSAs and identified parts of the existing NAEP frameworks that did not apply. This more inclusive process meets an important goal of the study: to develop rich text descriptions of what students need to know and be able to do overall, based on the evidence from course materials. After these group results were summarized across courses within programs, teams of NAEP framework experts compared them to the NAEP items that are associated with the borderline performance descriptions (BPDs) and cut scores from the Judgmental Standard Setting (JSS) study.

This study addresses the following research questions:

1. What mathematics and reading KSAs are prerequisite to the entry-level courses for the job training programs in each occupation, and what mathematics and reading KSAs are taught in these entry-level courses?
2. What mathematics and reading KSAs are students expected to have attained at the conclusion of the job training programs in each occupation?
3. How do the prerequisites for job training programs (KSA expectations for entry) in each occupation relate to descriptions of minimal academic preparedness on NAEP (as described by the BPDs from the JSS studies)?
4. How do the prerequisites for job training programs (KSA expectations for entry) in each occupation relate to the content assessed by NAEP (as determined by NAEP items representing minimal academic preparedness)?

Studies were conducted by the Educational Policy Improvement Center (EPIC), as subcontractors to WestEd and under the guidance of the Governing Board.

Summary of Findings

This study analyzed artifacts from 85 mathematics courses and 80 reading courses from 122 institutions. As noted, the courses were for job training programs in five occupational areas: Automotive Master Technician; Computer Support Specialist; Heating, Ventilation, and Air Conditioning (HVAC); Licensed Practical and Licensed Vocational Nurse (LPN), and Pharmacy Technician (entry and concluding level).

Key Findings Describing the Prerequisite KSAs

- The prerequisites are largely included in the grade 12 NAEP frameworks, but the full content of NAEP frameworks is much larger and broader. Course artifacts provided evidence of relatively few prerequisites that were not measured by NAEP.
- The job training programs studied have few prerequisite expectations in mathematics. The largest number of prerequisites across all occupationa training programs are found in the Number Properties and Operations domain and the "Systems of measurement," "Variables, expressions, and operations," and "Equations and inequalities" standards. No programs had prerequisites in the Data Analysis, Statistics, and Probability domain, and few had prerequisites in the Geometry domain. The artifacts included no evidence of irrational numbers, exponents and logarithms, or absolute value as prerequisites.
- Across all programs, only the NAEP objectives identified as prerequisites for entry-level courses in all five areas were those related to reading informational texts. NAEP objectives in the areas of literary text and literary devices were not found to be present in any programs. Any part of an objective that was not relevant to any program was labeled as an "exclusion." Specific reading skills that are prerequisite to all five job training programs include "Locate or recall causal relations" and "Locate or recall organizing

structures of texts, such as comparison/contrast, problem/solution, enumeration, etc."

- The mathematics exclusions removed much of the complex mathematics knowledge and skills that differentiate the grades 8 and 12 frameworks. As a result, some prerequisite KSAs appear to be better described by the grade 8 objectives.
- Mathematics prerequisites found in a small subset of courses but not assessed by NAEP include the following. The course in which they were found is in parentheses:
    - Boolean algebra, other number bases, and solution-driven algorithm design (Computer Support Specialist);
    - Interpreting mathematics symbols (LPN); and
    - Converting temperature and business mathematics (to understand profits and losses; entry-level Pharmacy Technician).
- Reading prerequisites evident in the course materials for specific courses, but not assessed by NAEP, include:
    - Comprehending and following written instructions, and writing documentation (Computer Support Specialist);
    - Comprehending and following written instructions, reading charts, graphs and diagrams, and conceptual understanding sufficient to apply scientific concepts (HVAC);
    - Identifying, recalling, and discussing information; applying knowledge; demonstrating evidence of and reflecting o one's knowledge; and conceptualizing and integrating (LPN); and
    - Reading materials on a computer screen rather than on paper, and deciphering text that includes spelling/grammatical errors in a context-appropriate way and without difficulty (entry-level Pharmacy Technician).
- Many grade 12 NAEP items were deemed irrelevant to determining academic preparedness for job training programs.
    - The number of reading objectives not evident as prerequisite in any course within the five occupations ranged between 6 and 25 of the 37.
    - Between 83 and 101 of the 130 mathematics objectives were not evident as prerequisite in any course within the five occupations.

Key Findings in Relation to the JSS Study

- The prerequisites evident in the course artifacts do not match findings from the JSS study and are generally less rigorous than the BPDs.
- The objectives for which evidence was found were heavily concentrated in the Number Properties and Operations content domain of mathematics. This domain is generally considered to be easier and less challenging than the other content domains. Only 10% of the grade 12 NAEP mathematics item pool includes this category of items (National Assessment Governing Board, 2008a).

- Mathematics prerequisites correspond to KSAs assessed by items falling below the Proficient level on the NAEP scale. All reading prerequisites correspond to KSAs assessed by items just above and just below the Proficient level on the NAEP scale (Measured Progress & WestEd, 2012).

Key Findings o Concluding-Level Course Prerequisites

- The Pharmacy Technician occupational area was the only one for which concluding-level courses were also analyzed in addition to entry-level courses. For mathematics, similar KSAs in entry-level and concluding-level Pharmacy Technician courses were identified as new material that would be taught in both courses. The most-taught KSAs include:
    - "Solve problems involving rates such as speed, density, population density, or flow rates" (evident as new material in 45% of the entry-level courses and 73% of the concluding-level courses);
    - "Solve problems involving conversions within or between measurement systems, given the relationship between the units" (evident as new material in 40% of the entry-level courses and 60% of the concluding-level courses);
    - "Write algebraic expressions, equations, or inequalities to represent a situation" (evident as new material in 30% of the entry-level courses and 47% of the concluding-level courses);
    - "Solve problems involving special formulas such as: $A = P(I + r)^t$, $A = Pe^{rt}$" (evident as new material in 40% of the entry-level courses and 47% of the concluding-level courses); and
    - "Use proportions to solve problems, including rates of change" (evident as new material in 40% of the entry-level courses and 47% of the concluding-level courses).
- No evidence of grade 12 NAEP reading objectives taught in the concluding-level Pharmacy Technician courses was found in the course artifacts.
- Slightly more NAEP mathematics objectives are prerequisite to entry-level Pharmacy Technician courses (9 objectives rated as prerequisite in at least 20% of courses) than to concluding-level courses (8 objectives rated as prerequisite in at least 20% of courses).