

National Assessment Governing Board

Committee on Standards, Design and Methodology

March 2, 2012 9:45 a.m. – 12:15 p.m.

AGENDA

9:45 -9:50 am	Welcome, Introductions, and Agenda Overview <i>Lou Fabrizio, COSDAM Chair</i>	
9:50 –11:15 am	Update on 12 th Grade Preparedness Research Program	Attachment A
9:50 –10:05 am	<ul style="list-style-type: none"> ▪ Statistical Relationship Analyses: Performance of Florida Students on Preparedness Indicators and in First Year of College Relative to Performance on NAEP <i>(Informational Item for Discussion)</i> 	Attachment A-1
10:05 - 10:20 am	<ul style="list-style-type: none"> ▪ Research on Job Training Curriculum Comparisons to NAEP <i>Susan Loomis, Governing Board Staff</i> 	Attachment A-2
10:20 - 10:50 am	<ul style="list-style-type: none"> ▪ Survey of Placement Assessments and Policies of Post-Secondary Institutions <i>Ray Fields, Governing Board Staff</i> 	
10:50 – 11:25 am	<ul style="list-style-type: none"> ▪ Reporting Grade 12 Preparedness <i>Ray Fields, Governing Board Staff</i> <i>Susan Loomis</i> 	
11:25 - 11:30 am	Recommendations of Future COSDAM Agenda Topics	
CLOSED SESSION		
11:30– 12:15 pm	Writing Achievement Levels-Setting Update <ul style="list-style-type: none"> ▪ Revised Writing Achievement Levels Descriptions ▪ Field Trial with Revised Achievement Levels Descriptions: January 27, 2012 ▪ Writing Achievement Levels Setting Study: February 7-11, 2011 <i>Luz Bay, Measured Progress Project Director</i> <i>Susan Loomis</i> 	Attachment B

Program of Preparedness Research Updates and Final Steps for Reporting

OVERVIEW OF PROGRAM OF PREPAREDNESS RESEARCH

Based on the Program of Preparedness Research adopted by the Board in March 2009, four categories of research studies are recommended to produce results for reporting 12th grade preparedness for grade 12 NAEP 2009 reading and mathematics. The following four categories are included.

- 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 has been implemented. More information is provided below.

As part of the ongoing updates to COSDAM, this document includes an overview of each study type, followed by an update on new developments since the December 2011 Board meeting. A timeline is included to show only the remaining steps in the process to be completed for reporting results in 2012.

Updates for March 2012

For each type of study, there is a description of the study including background information that has already been shared with COSDAM, as well as an update including information about the current status of the research. Briefly, the *updates for March 2012* are as follows:

1. **Content alignment studies:** completed; no updates
2. **Statistical relationship studies:** final briefing on analysis of data for state representative sample of 2009 NAEP examinees in Florida, including statistical relationships of NAEP performance with ACT, SAT, and ACCUPLACER and data for Florida students' first year in college
3. **Judgmental standard setting studies:** staff recommendation on reporting; information about additional research about curriculum requirements for job training programs
4. **Survey of higher education:** 85% response rate attained; preliminary data reported to COSDAM

TIMELINES

Studies listed in the chart below are for both reading and mathematics. This timeline is provided to help focus attention on the final steps to be completed before reporting results from the preparedness studies. Many interim steps are excluded from the chart.

For the content alignment, judgmental standard setting, and higher education survey, a full federal procurement process was required which involved many steps and compliance with specified timelines.

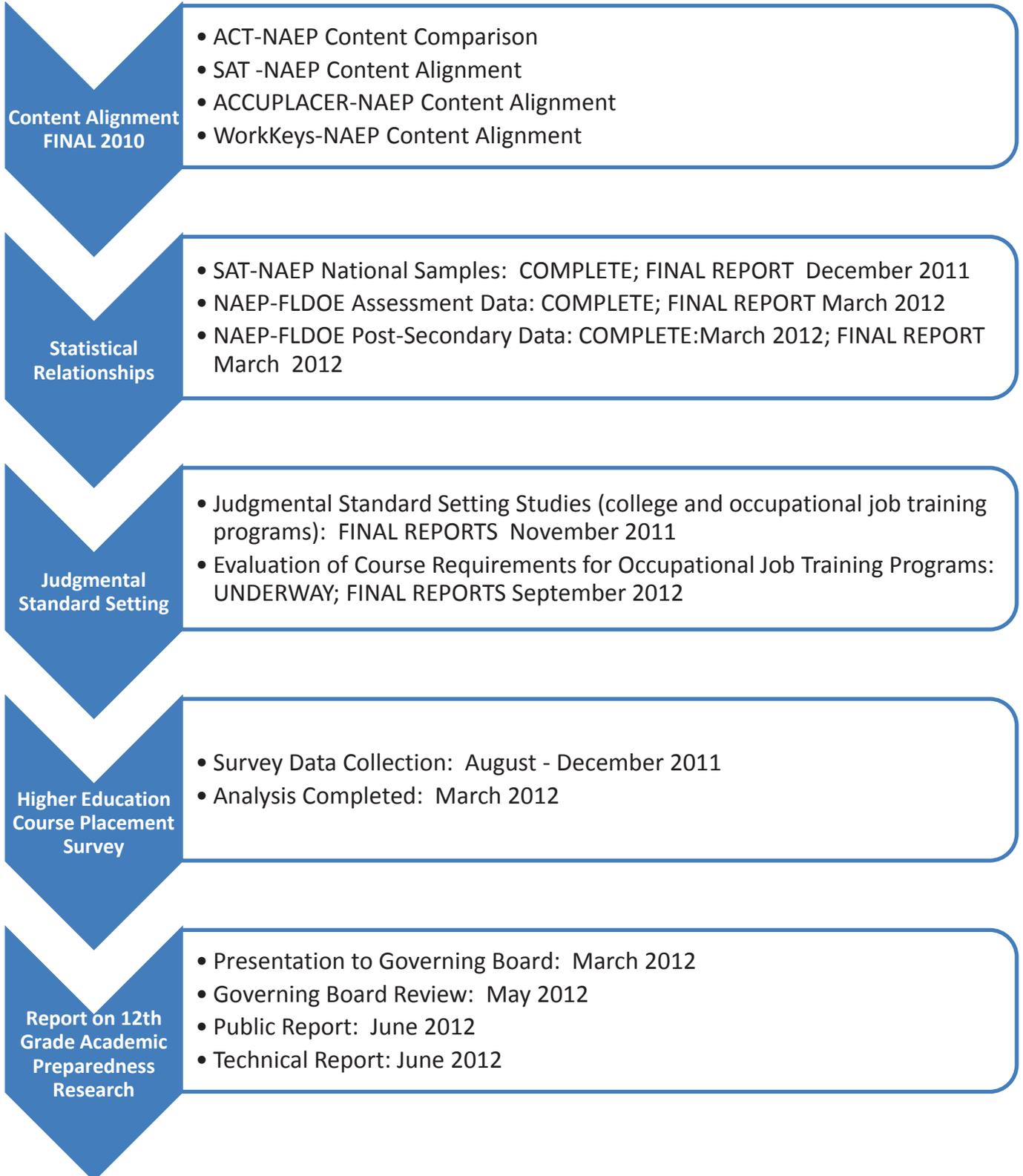
In addition, a pilot study was required for each type of study, except for the analysis of statistical relationships. The analyses of statistical relationships were reviewed by the Technical Advisors for 12th Grade Preparedness Research at several stages, and additional analyses were subsequently produced and reviewed. Board staff worked in collaboration with The College Board (SAT) and Florida Department of Education (FLDOE) staff to develop data sharing agreements for the studies of NAEP statistical relationships with SAT and FLDOE data. Board staff worked in collaboration with NCES staff, College Board staff, FLDOE staff, ETS staff, and Westat staff to develop procedures for maintaining the confidentiality of student data, matching records, and producing statistical results.

Design documents were developed by Board staff working with consultants for the content alignment and judgmental standard setting studies. Development of the higher education survey was a lengthy process of collaboration between Board staff and contractors, first at ACT and finally at Westat, and between Board staff and the Office of Management and Budget.

For the content alignment and judgmental standard setting studies, there was a lengthy process of identifying programs and panelists appropriate for the studies. Panelist recruitment is always a labor and time intensive process, and the effort was even greater for the judgmental standard setting studies involving panelists in areas entirely new to the Governing Board's work.

The following chart shows the types of studies and indicates the current status and the date when studies will be complete and final reports submitted.

**TIMELINES FOR REPORTING RESULTS OF STUDIES OF
12th GRADE PREPARED RESEARCH FOR
2009 NAEP READING AND MATHEMATICS**



Content Alignment Studies

Background: 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. Content alignment studies for grade 12 NAEP in reading and mathematics will be implemented for the assessments for which data and materials are available. We have not yet succeeded in negotiating the use of The Armed Services Vocational Aptitude Battery (ASVAB) in our research, but Governing Board staff continue to pursue this potential.

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. A brief summary of those studies was shared with COSDAM at the November 2010 meeting.

Update: Final reports have been received and a summary version has been shared with COSDAM and posted at www.nagb.org. The complete reports for each study are available upon request. This work is final. There is no additional information for the March 2012 briefing.

Studies to Establish Statistical Relationships

Background: Highest priority is generally placed on these studies. Currently, two main sets of studies have been conducted under this heading. One is a study to relate SAT scores in reading and in mathematics to the national sample of NAEP scores for grade 12. The goal is 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 11 states that participated in the pilot state-NAEP for grade 12 included some of the more populous states such as Massachusetts and New Jersey that are largely “SAT states,” but the 11 included more “ACT states.”

The correlation between NAEP and SAT reading was found to be lower than that for mathematics. The correlation between the NAEP reading scores and Florida state assessment

scores was also lower than for mathematics. Research into those relationships, as suggested by the Technical Advisors for 12th Grade Preparedness Research, included: (1) inclusion of SAT scores in the NAEP conditioning model, (2) identification and removal of outliers, (3) evaluation of demographics of outliers, i.e., sensitivity analysis, (4) evaluation of alternative SAT scores (e.g. highest, most recent, composite driven), and (5) disaggregation of NAEP reading students' scores based on block content of their assessment booklet.

The final sample used for linking the NAEP reading and SAT critical reading included approximately 16,200 students. The correlation between the two reading scales was 0.74. For NAEP and SAT mathematics, the linking sample included approximately 15,300 students, and the correlation between the math scales was 0.91.

Several methods were examined for evaluating the statistical relationship between NAEP and SAT scores for both mathematics and reading. A concordance was possible for mathematics, but the relatively low correlation did not support this method for establishing the SAT reference point for NAEP reading. Each method results in a different way of interpreting the results, and each can be helpful in understanding the overall relationship of NAEP and SAT for reporting preparedness. The preparedness benchmark for college readiness adopted by the College Board is 1550 for the composite score and 500 for writing, mathematics, or critical reading subtests. These benchmarks represent the score on the SAT that students would need in order to have a .65 probability of earning a B- freshman year grade point average. For each of the linking methods examined, the results indicate that the SAT readiness benchmark of 500 is very close to the NAEP Proficient cut score for both reading and mathematics, and the scale alignment varies by the probability of the estimate.

The second set of studies analyzes a series of statistical relationship 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.

Update: Included in Attachment A-1 is a description of the analysis of the statistical relationship between NAEP grade 12 data and the state-representative sample of 12th grade Florida students who took the NAEP in 2009 in either reading or mathematics. The analysis for Florida students includes both statistical linking studies with ACT, SAT, and ACCUPLACER test data and data on the first year in college for students who entered college. This research is on-going, and data will be provided annually to update the post-secondary performance of Florida students who were in the 12th grade NAEP sample in 2009. The complete report will be finalized and shared with COSDAM prior to the March 2, 2012 meeting.

Judgmental Standard Setting Studies

Background: 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. Unlike most other studies of preparedness for post-secondary activities in college or job training programs, the Governing Board has not assumed that prepared for college and the work place are the same. Rather, our studies are aimed at determining whether the level of performance on NAEP is approximately the same or significantly different for entry in job training programs for the occupations included in our research studies and 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

An overview of the pilot study for automotive master technicians and college course placement, conducted April 26-29, was presented to COSDAM on May 13, 2011. Implementation of the operational studies for job training programs in 5 occupations and for college course placement was completed on July 1, 2011. Operational sessions were conducted for two post-secondary areas each and for both reading mathematics. 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.

A comprehensive report on the results for the judgmental standard setting studies was provided to COSDAM in November 2011 via a WebEx virtual meeting. At that time, COSDAM discussed the general approach to the preparedness studies and the need to evaluate results of the judgmental standard setting studies in the context of the overall set of studies. COSDAM recommended that reports emphasize the fact that the studies included job training programs for only five occupations and noted that the results for five other occupations could point to different interpretations regarding the minimal level of academic preparedness.

Additional research is underway to collect more information from the job training programs included in these JSS studies. This work is being conducted under the direction of WestEd as the primary contractor and the Educational Policy Improvement Center (EPIC) as the subcontractor. The study requires collection of course materials for both entry level courses and courses at the completion of job training programs. The materials will be reviewed to determine the knowledge, skills, and abilities (KSAs) that students need to have to enter the job training program, i.e. pre-requisites, and those that are taught in the program. The KSAs represented in course requirements will be compared to the borderline performance descriptions developed in the JSS studies for each job training program in each subject to evaluate the extent to which the descriptions used to set the cut scores on NAEP for each job training program align with the pre-requisite knowledge, skills, and abilities, those taught in the beginning courses, and those taught in the final courses. The information from course materials will also be compared to the NAEP item pool for each subject to evaluate the alignment of knowledge, skills and abilities assessed by NAEP relative to those required as pre-requisites, in courses taught at program entry, and in courses taught just at program completion. Finally, the knowledge, skills, and abilities represented by the course materials will be compared to items in the NAEP item pool near the cut scores set by the panelists to represent minimal preparedness for entry in each training program. This information is being collected to evaluate the reasonableness of the cut score recommendations across the job training programs.

Update: A report on the project is included as Attachment A-2. A timeline for the key activities of the project is included in that report. Information regarding preliminary findings will be provided to COSDAM at the meeting on March 2, 2012.

Higher Education Survey

Background: A survey of two-year and four-year post-secondary institutions was implemented to gather information regarding (1) the placement tests used and (2) the cut scores in reading and mathematics below which remedial reading and mathematics course placement results, and at or above which placement in credit-bearing entry level courses results. The sample was nationally representative of accredited postsecondary education institutions. Results will be analyzed according to several attributes of the institutions, including enrollment size and level of selectivity for admissions.

Several small-scale studies were conducted to refine the survey instrument. A pilot study of 120 postsecondary institutions yielded an overall response rate of 86 percent. Based on the results of the pilot study, minor changes to the questionnaire were made.

The survey was submitted to the Office of Management and Budget (OMB), and clearance was achieved in July, 2011. Distribution of the survey to institutions began in August, 2011.

Update: An 86% response rate was achieved with responses from 1522 institutions. Data tables have been prepared and presented to Governing Board staff for analysis. Ray Fields will join the COSDAM meeting to discuss preliminary findings.

Benchmarking Studies

Background: The Governing Board has discussed the possibility of conducting benchmarking studies as part of the 12th grade preparedness research. This type of study was not included in the recommendations of the Technical Panel for 12th Grade NAEP Preparedness Research for the first cycle of research, largely because of the challenges involved in collecting the data. They suggested that the design would potentially provide useful preparedness information and help to answer questions that were likely to arise from the research studies in the first cycle.

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 study, which was carried out by Westat. . Westat, the NAEP sampling and administration contractor to NCES, conducted interviews with officials in each of the 9 higher education institutions in Texas that volunteered to participate in the pilot study at the invitation of the Texas Commissioner of Higher Education. The institutions included both two-

and four-year institutions with considerable diversity among the enrollments across the institutions. Those interviews were conducted to gather specific information related to logistics, student sampling, and student recruitment prerequisite for the pilot study. The response from each campus was positive and enthusiastic. 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.

High School Transcript Study Data

Background: Board staff plan to use data from the 2009 High School Transcript Study in reporting the grade 12 NAEP preparedness research as a source of contextual information for the results of the Board’s program of preparedness research. The transcript data are available for mathematics and science courses and not for courses related to student performance in reading. Course information is standardized as part of the High School Transcript Study to enable national comparisons. Data were collected from over 700 high schools in the 2009 study and include a variety of information in addition to course titles, grades, and credits earned. The data also include, for example:

- course and program offerings at each high school in the sample;
- graduation requirements at each high school in the sample;
- type of course (e.g., remedial, honors, Advanced Placement, International Baccalaureate);
- location of the course (e.g., on campus, distance learning, vocational education center);
- approximate percentage of graduates who enrolled in 2-year, 4-year, and vocational postsecondary institutions for each high school in the sample; and
- graduation status at the end of grade 12 (e.g., type of diploma earned, still enrolled, dropped out).

Data on course-taking patterns can supplement the Board’s preparedness research in the following ways.

1. Provide contextual indicators on the NAEP scale

Each study in the full compilation of the Board’s Program of Preparedness Research is expected to result in a preparedness “reference point” on the NAEP scale. A reference point could be a score point or a range of scores on the NAEP scale. Findings from the full compilation of preparedness studies will be examined in relation to particular descriptive statistics, such as:

- Average NAEP score for students who took a “rigorous” curriculum;
- Average NAEP score for students who took an Advanced Placement course; and

- Average NAEP score for students who took an International Baccalaureate course. These data can be useful in interpreting the various reference points associated with preparedness and can be used to evaluate the logical relationships and mutually confirmatory evidence of validity.

2. *Identify course-taking patterns associated with reference points*

Selected reference points, or score ranges, can be used to identify course-taking patterns for students whose scores fall within these same score ranges:

- Use score ranges associated with preparedness that are identified in the Judgmental Standard Setting Studies and the Statistical Linking Studies in the Board's Program of Preparedness Research, and identify patterns in course-taking for students whose scores fall within or above these score ranges.
- Use the longitudinal data provided by Florida to identify the average NAEP score for students who placed into college-level courses, and identify patterns in high school course-taking for students who scored at or above this level.

Update: Data from the 2009 high school transcript study will be incorporated into the reports on 12th grade preparedness research.

12th Grade Preparedness Validity Framework

Development of validity framework continues for the 12th Grade NAEP Program of Preparedness Research. Versions of the document have been reviewed by experts in validity and in program evaluation, as well as by COSDAM. Comments from the NAEP Validity Studies Panel have been received, and changes to the document in response to that input are in progress.

Update: Staff will complete work on the validity framework and share the document with Technical Advisors for Preparedness Research, other advisory groups, and with individuals and groups identified as having an interest in the research. The comments and recommendations will be shared with COSDAM and the Governing Board to inform reporting of 2009 preparedness data and research for the 2013 and subsequent grade 12 assessments.

Procurements for Reporting 12th Grade NAEP Preparedness Research

A contract was awarded to Widmeyer Communications to work with the Governing Board staff for production of the “public” report on the overall results of the 2009 NAEP preparedness research to be produced as both a print document and web-based document and addressed to a general audience. Production of completely electronic web-based technical report will be primarily addressed to the research community. Work is underway on developing the overall conceptual design of the reports. Widmeyer staff are attending meetings of the High School Commission on 12th Grade NAEP Preparedness Research, the Technical Advisors for 12th Grade NAEP Preparedness Research, and meetings of Board committees—COSDAM and Reporting and Dissemination to gain greater understanding of the research and related Board policies.

The Reporting and Dissemination Committee will be the primary focus of updates on preparation of the final reports documents, and COSDAM will continue to be the primary focus of updates on the studies in the reports.

The Governing Board will be briefed on the program of research and findings to date at the March 2012 meeting and on the final reports at the May 2012 meeting. The final reports are expected to be ready for distribution by June 2012.

OVERVIEW OF 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 will be analyzed, 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 an assessment designed for use by employers to evaluate the knowledge and skills of a prospective employee relative to a job profile. WorkKeys scores are used more generally to certify workplace readiness. WorkKeys assesses 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.
- 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.

NAEP 12th Grade Preparedness Research: Overview of Analyses Relating Florida Students' Performance on NAEP to Preparedness Indicators and Postsecondary Performance

Rebecca Moran, David Freund, and Andreas Oranje, ETS

As part of the National Assessment Governing Board's efforts to enable NAEP to report on the preparedness of U.S. twelfth graders for postsecondary education or entry into job training programs, studies were conducted to statistically relate performance on NAEP with results from other assessments that serve as indicators of preparedness for college entry, course placement, and entry into the workforce (National Assessment Governing Board, 2009). Both nationally-representative data, such as those used to establish a statistical link between NAEP and SAT, and data representative of students in individual states were of interest for such statistical relationship studies. The 2009 12th grade NAEP Reading and Math assessments included a first-time pilot state assessment for 11 states; Florida was one of the participants. The Florida Department of Education (DOE) maintains a longitudinal data base (K-20 Education Data Warehouse) that includes college entrance and placement test scores and first-year college performance data for those students who attended public colleges in the state of Florida during the 2009-2010 academic year.

This document describes the data and procedures used to evaluate Florida students' performance in high school and first year of college relative to scores on the NAEP 12th grade reading and mathematics assessments and other test scores. This will be followed by a description of the analyses that serve as a follow-up to the statistical relationships established between NAEP and SAT at the national level. Of particular note are the analyses of postsecondary data to provide validity evidence for the potential preparedness reference points on the NAEP scales identified in the national statistical relationship studies.

Data

This study used data from Florida public school students who participated in the 2009 NAEP 12th grade reading or mathematics assessments, approximately 3,200 in math and 3,400 in reading. Analyses were conducted with the use of NAEP sampling weights to appropriately represent 12th grade public school students in Florida in 2009.

Matching NAEP and Florida DOE Data

The Governing Board entered into an agreement with the Florida DOE to obtain longitudinal data for public school students selected to participate in the 2009 NAEP 12th grade assessment. The process of matching data between the Florida database and NAEP participants was carried out in coordination with NAEP contractors, Westat and ETS, and the Florida DOE.

A critical requirement of the matching of student records was to protect students' identity and maintain confidentiality. This was assured through the assignment of a unique pseudo ID for students sampled to

participate in NAEP. At the time of sample selection of students for operational NAEP, Florida DOE staff appended the pseudo ID to files within the Florida DOE and transmitted the pseudo ID to Westat with other administration data. On all subsequent data files containing Florida data (e.g., ACT scores), only the pseudo ID was included on the files. The pseudo ID was used by Westat to match files from Florida back to the NAEP data files. Westat in turn provided files to ETS with the additional Florida data appended to NAEP student records. Throughout the process, ETS had no access to any Personally Identifiable Information (PII), such as names, birthdates, or social security numbers. This process was essentially identical to the matching process conducted with the NAEP-SAT national linking study.

Data Elements evaluated for use in the Florida Preparedness Research

Of the variables available in the rich Florida longitudinal dataset for the 2009 12th grade cohort, those examined for use in this research are described briefly below. Some of these data elements lacked sufficient power (i.e., small sample sizes in the linked set) and, therefore, value for extensive use in the current research.

1. **Florida Comprehensive Assessment Test (FCAT)** is Florida's K-12 state assessment. Scores on the reading and mathematics tests from 3rd through 10th grades were available in the longitudinal dataset. Match rates were very high, with 10th grade scores in reading and math matched to approximately 94% of the NAEP test takers. However, concerns about the relevance of relating students' 12th grade NAEP performance to FCAT scores earned two years earlier, while in 10th grade, precluded further analyses being pursued with the FCAT data.
2. **WorkKeys® is a job skills assessment** system that helps employers select, hire, train, develop, and retain a high-performance workforce. WorkKeys® includes three relevant tests: Applied Mathematics, Locating information, and Reading for Information. Matched sample sizes were about 300 students (about 10% of the NAEP sample in each subject) for each WorkKeys® test and therefore were inadequate for further analysis.
3. **Advanced Placement (AP)** college-level exams enable students to earn college credit and advanced placement in college courses. Approximately 36% of students in the matched NAEP-Florida dataset took one or more AP tests. However, only 16% of the NAEP reading sample took a relevant reading AP exam (English or English Literature) and only 8% of the NAEP sample took a relevant math AP exam (calculus). The small sample sizes limited the efficacy of these data for further analysis.
4. **High School Program:** One of the background questions asked of students on the 12th grade NAEP assessment was "Which of the following best describes your high school program?" Response options included (1) General, (2) Academic or college preparatory, and (3) Vocational or technical school. For the linked NAEP/Florida sample, approximately 47% of students indicate their program was "general", 43% indicated "academic or college preparatory", and 9% indicated "vocational or technical school". These data were examined in much greater detail; the technical report to be shared at the March Governing Board COSDAM meeting will provide more results.

5. **SAT and ACT College Entry Exams and ACCUPLACER College Placement Exam**. Approximately 43% of Florida’s NAEP sample took the SAT test; 47% took the ACT test, and 18% took the ACCUPLACER test.
6. **College Enrollment Status, First-Year Course-taking and Grade-point Average**: Data were obtained for students attending public colleges and universities in Florida for the 2009-2010 academic year. Approximately 54% of the students in Florida’s 2009 NAEP 12th grade sample attended a public postsecondary institution in Florida, with 36% attending community colleges and 17% attending four-year colleges and universities.

Analyses Conducted

The purpose of this research activity was to explore the relationships between Florida students’ performance on the 12th grade NAEP assessments and other indicators of postsecondary preparedness to provide validity evidence for the potential preparedness reference points on the NAEP scales that were identified by the national NAEP-SAT statistical relationship study. Those potential reference points are as follows:

<u>Statistical Projection</u>		
Percentage at or above 500 on SAT	Math	Reading
50%	164	302
67%	169	313
80%	175	325
<u>Concordance</u>		
SAT Subscore = 500	165	303
<u>NAEP "Proficient"</u>		
	174	302

Thus, NAEP scale score distributions were examined in relation to:

- SAT performance, defined in terms of whether students met the SAT college readiness benchmark in each subject area (determined by College Board to be a score of 500 on the SAT critical reading and mathematics measures)
- ACT performance; the benchmarks for college readiness established by ACT are 22 for math and 21 for reading
- ACCUPLACER performance, where the cut scores used in Florida for placement into credit-bearing courses are 72 for math and 83 for reading

- Students' self-reported program of study in high school, whether vocational/technical, general, or academic/college preparatory. This information was collected as part of the NAEP student background questionnaire.
- College enrollment: community college, four-year college/university, or unknown
- First-year college course-taking: remedial or credit-bearing
- First-year college grade point average: above or below a B-

Summary of Results

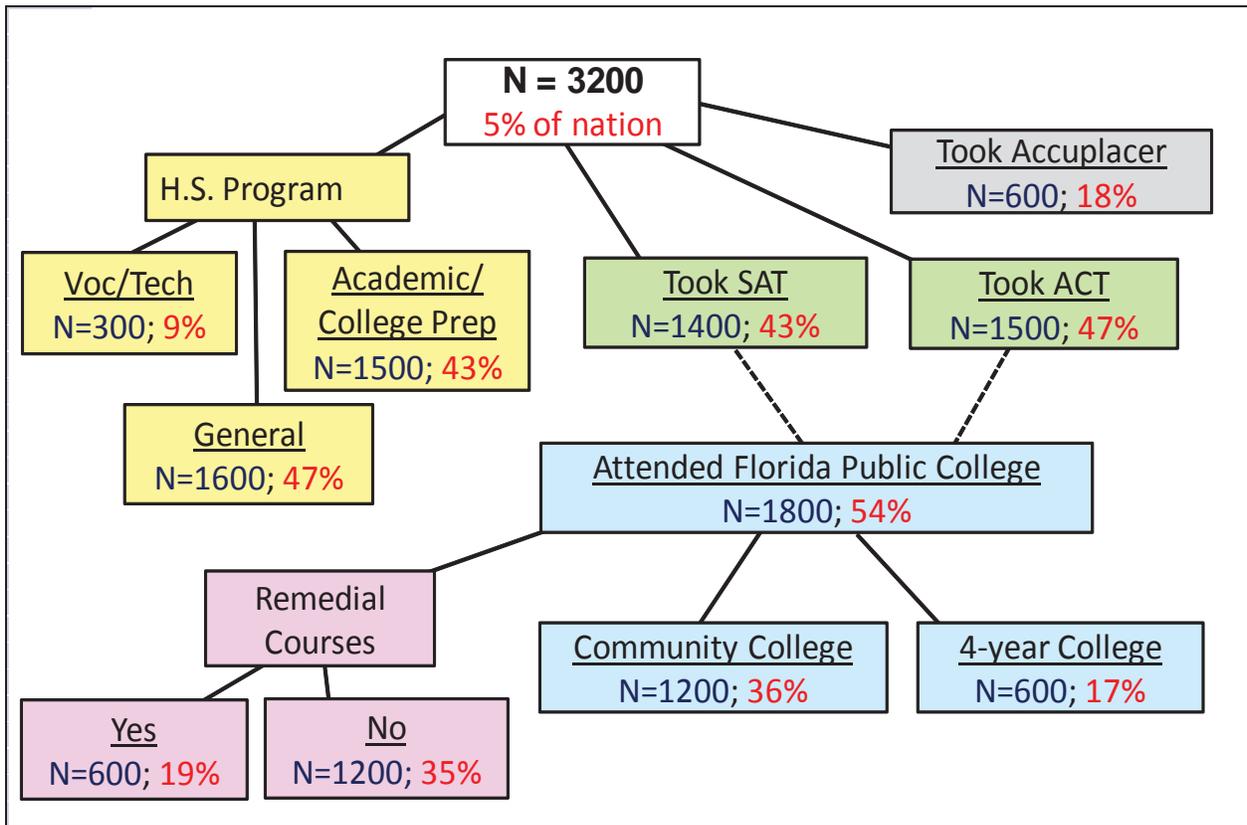
Overall, the patterns of results observed for this cohort of Florida students did not contradict the potential preparedness reference points on the NAEP reading and mathematics scales identified through the national NAEP-SAT linking study. For instance,

- Average NAEP scores for students who met the SAT preparedness benchmarks were near the NAEP *Proficient* cut scores and fell in or above the range of possible NAEP preparedness reference points. The average scores for these students were roughly one standard deviation higher than average NAEP scores for their peers who did not meet the SAT preparedness benchmark.
- Performance on NAEP was similarly differentiated for students who met vs. did not meet the ACT preparedness benchmarks.
- Average NAEP scores for students attending four-year colleges fell within or exceeded the potential NAEP preparedness reference points, whereas average scores for their peers attending community colleges were below the reference points.
- A greater percentage of students taking no remedial courses during the first year of college scored at or above the NAEP preparedness reference points compared to students who required one or more remedial courses.

However, the limitations of the Florida data, namely the availability of data only for students enrolled in Florida public postsecondary institutions, must be taken into consideration when interpreting these results.

Figure 1 provides sample sizes and percentages for the 2009 NAEP 12th grade Florida sample disaggregated by high school program, test-taking, college attendance, and remedial course-taking.

Figure 1: Sample sizes and Percentages for 2009 NAEP/Florida Grade 12 Math



Job Training Program Curriculum Study for NAEP Preparedness Research

Submitted by WestEd

The National Assessment Governing Board (Governing Board) adopted a Program of Preparedness Research in March 2009 that included judgmental standard-setting (JSS) studies for the 12th grade National Assessment of Educational Progress (NAEP). These studies produced preparedness reference points on the NAEP scale for entry into job-training programs and for placement in college credit-bearing courses, representing the academic knowledge and skills required for postsecondary course and training program placement. A total of 180 job training programs were represented in the judgmental standard setting studies focusing on five occupations:

Occupation	Programs
Automotive master technician	41
Computer support specialist	31
Heating, ventilation, and air conditioning technician	31
Licensed practical nurse	40
Pharmacy technician	37

The Governing Board requested additional research to examine the validity of findings obtained from the JSS studies and to better understand the knowledge, skills, and abilities in reading and mathematics required for these occupational training programs. This additional research is intended to provide a clearer understanding of the knowledge, skills, and abilities (KSAs) required for entry- and exit-level coursework in designated job training programs within these occupations. This study will help to determine if the KSAs required of students in the training programs are appropriately represented by the borderline preparedness descriptions (BPDs) and by the NAEP items near the reference points developed in the JSS studies to represent the minimal level of academic knowledge and skills in the subject matter necessary for a student to be prepared to enter the job training course.

METHODOLOGY

This study addresses the following research questions:

1. What mathematics and reading KSAs are prerequisite to the introductory-level courses, and what mathematics and reading KSAs are taught in the introductory courses for the job-training programs for each occupation?
2. What mathematics and reading KSAs are students expected to have attained at the conclusion of the job-training programs for each occupation?

3. How do the prerequisites (KSA expectations for entry) for job training programs in each occupation relate to descriptions of minimal academic preparedness on NAEP (as described by the BPDs from the JSS studies)?
4. How do prerequisites (KSA expectations for entry) for job training programs in each occupation relate to the content assessed by NAEP (as determined by NAEP items representing minimal academic preparedness)?

This study comprises three primary phases:

1. Identification and collection of course artifacts
2. Review of course artifacts by Review Teams
3. Review of resulting KSAs by NAEP Content Expert Teams

Identification and Collection of Course Artifacts

Programs from the five occupations used in the JSS studies will comprise the population of programs for this study; from this population, a minimum of 20 programs per occupation will be recruited from the 180 programs represented on the JSS panels.

Occupational job-training instructors who served on the JSS panels are being recruited to participate in this study. These job training instructors are being asked to identify courses that best address the objectives of this study and to submit artifacts for those courses. These instructors also have the option of nominating colleagues who teach one or more courses selected for the study to participate in this activity. Course artifacts will be collected for all programs in each occupational area that agree to participate, with course submission remaining open until either materials are obtained from a minimum of 20 programs or the population of programs has been exhausted.

Each participating program instructor is being asked to (1) identify foundational textbooks for her/his program; (2) verify program information collected by EPIC (e.g., accreditation status, course sequencing, school and department admission requirements, degree accreditation, and credit requirements); and (3) submit course artifacts for two introductory courses and two concluding courses. Course artifacts may be submitted via a web-based upload tool, email, facsimile, and physical mail.

Introductory courses

Introductory courses differ across programs within an occupation, and across occupations, in terms of standardization and sequencing. As such, “entry-level” courses could embody one or more of numerous definitions, including (1) those that occur lowest in the course sequence for a program, regardless of course title; (2) those that are core “Introduction to...” or “Foundations of...” courses that occur across the majority of programs, and (3) those that are identified by instructors as being most representative of the mathematics and reading expectations for entry-

level students in the program. Because the focus of the study is on identifying the mathematics and reading skills expected upon entry into introductory-level courses in the job-training programs for each occupation, courses are being selected for inclusion using the third definition.

Concluding courses

Concluding, or exit-level, courses also differ in level of standardization, and multiple options for identifying such courses also exist. For consistency, the same approach is being used to identify the exit-level courses for inclusion in the study: instructors are asked to identify those courses that best represent the mathematics and reading knowledge and skills that students are expected to know upon program completion.

For each training program, a set of course materials will be collected for introductory courses and a set for concluding courses. The following types of artifacts will be submitted and assembled into a course packet (with at least one of each type of artifact required):

1. Course syllabus
2. Textbook title(s) (with author and ISBN)
3. Textbook table of contents (instructor will copy and upload or EPIC will download from publisher website)
4. Course exam (one or more), preferably the mid-term or earlier for introductory courses and the final exam for concluding courses
5. Text-based assignment (one or more), with corresponding passage, that best illustrates mathematics and reading KSAs needed by students—one or more for introductory courses and one or more for concluding courses
6. Stand-alone assignment (one or more) such as a lab, worksheet, problem sheet, essay, or group project that best represents mathematics and reading KSAs needed for students—one or more for introductory courses and one or more for concluding courses

Instructors representing institutions that offer more than one program within an occupational area are asked to complete a submission for one program and to complete submissions for additional degree programs if selected courses are different than those already submitted.

Review of Course Artifacts

Once course artifacts are gathered, a team of content experts will be trained to consistently and reliably apply a coding scheme to the course artifacts to identify prerequisite and taught content for each of the occupational training programs. Two Review Teams, one for mathematics and one for reading, have been recruited. Each consists of two mathematics or two reading experts and one occupational area (e.g., automotive master technician) expert. The content experts in mathematics and reading were recruited from a pool of trained analysts who have substantial experience in this type of work. The occupational-area experts recruited for each of the Review Teams (one for mathematics and one for reading) were drawn from pools of mathematics and reading JSS panelists who were nominated by the JSS studies' content and process facilitators as being well qualified for this type of work.

Review Teams will independently code the course packets for their content area. In order to maximize the efficiency of the Review Teams, an initial set of foundational KSAs is used to analyze course materials. These foundational KSAs include the NAEP frameworks and additional KSAs derived from the National Career Clusters™ Essential Knowledge and Skill Statements, synthesized to reduce redundancy and to present only those KSAs relevant to mathematics and reading.

Once the Review Teams' review of course materials is complete, EPIC staff will aggregate the individual ratings for each course within each program to summarize the mathematics and reading KSAs that are prerequisite to and taught in introductory-level courses and that students are expected to have attained at program completion. Responses will be aggregated to create overall content maps describing the relationship between frameworks and prerequisite KSAs for each occupation. In addition to tabular data displays, the data will be displayed using color shading, as well as summary statistics, to show the extent of overlap in content between standards and programs. Content maps, grouped by key characteristics, will also be created for programs, to show the impact of key program characteristics that impact findings. EPIC staff will review the content maps to identify similarities and differences across program types within occupations and will note differences in findings due to program characteristics. Final results will be provided both overall and by key program characteristics. EPIC staff will also compute descriptive statistics to summarize the Review Teams' demand ratings overall (by occupation) and by program type, should program characteristics have an impact on the demand of occupational courses.

Review of Knowledge, Skills, and Abilities Required for Training Courses

Two NAEP Expert Teams, one team for mathematics and one for reading, each consisting of two experts, will review the prerequisite and taught KSAs (as identified by the Review Teams) in the context of NAEP. They will describe the relationships between the prerequisite content and both the BPDs and the content on the 2009 NAEP, evaluating the results of the Review Team analyses to describe

KSAs assessed by NAEP that are not included in the job-training programs and KSAs included in the job-training programs that are not part of the NAEP frameworks or assessments.

Comparison to BPDs

Using the Review Teams' determination of KSA requirements and course artifacts, the NAEP Expert Teams will synthesize and describe the relationship between the content that is prerequisite to and taught in occupational programs and the content described in the BPDs for that program. Conclusions will be provided overall for each occupation, identifying differences related to program characteristics.

Comparison to NAEP items

Each NAEP Expert Team will also compare KSAs identified for each program's introductory courses (drawing upon the content maps and BPD comparisons) to the NAEP item pools. Starting with a set of items near the cut scores identified in the JSS studies, they will judge the correspondence between the course prerequisite KSAs and the KSAs needed to correctly respond to the items with a .67 probability. They will be asked to identify the items in the range of the cut score plus one standard deviation that are prerequisite to or required in the courses. They will also be asked to examine items below the cut score and above the range in the first analysis to determine if the KSAs represented in the curricular requirements are largely above or below this range.

PILOT STUDY

In order to address unanticipated challenges that may arise when implementing the proposed design, materials, and/or logistics, a pilot—or feasibility—study is being implemented. The automotive master technician occupation has been selected for the pilot study. Lessons learned through the pilot study will be used to refine the study design as needed for the subsequent four occupations.

However, to expedite the collection of and reporting on introductory course data—which is of particular and immediate importance to the Governing Board—coding and analysis of introductory course data, which begins with the automotive master technician occupation, will continue for the remaining occupational areas before the analysis of concluding courses commences. Similarly, the NAEP Expert Teams will conduct their analyses for each occupation's introductory courses prior to conducting their analyses of the concluding courses; in addition to providing introductory and concluding course analyses, the Expert Teams will also provide an overall review of each program and occupational area as well as a summary report of KSAs identified as pre-requisites to automotive master technician programs.

Decision points identified during the pilot study will be discussed with the Governing Board, and decisions will be documented and implemented for data collection and analyses within the remaining occupations.

KEY ACTIVITIES IN PROPOSED PROJECT SCHEDULE

PLANNING ACTIVITIES	
Document Submission Tool released to participants	1/12/12
PILOT STUDY ACTIVITIES (INTRODUCTORY COURSES)	
Course artifact collection ¹	1/12/12–2/14/12
Review Teams course packet reviews	2/3/12–2/24/12
NAEP Expert Teams reviews	3/1/12–3/12/12
REMAINING OCCUPATIONS ACTIVITIES (INTRODUCTORY COURSES)	
Course artifact collection ¹	2/13/12–3/9/12
Review Teams course packet reviews	3/2/12–4/20/12
NAEP Expert Teams reviews	3/12/12–5/7/12
PILOT STUDY ACTIVITIES (CONCLUDING COURSES)	
Review Teams course packet reviews	4/24/12–5/4/12
NAEP Expert Teams reviews	5/4/12–5/16/12
REMAINING OCCUPATIONS ACTIVITIES (CONCLUDING COURSES)	
Review Teams course packet reviews	5/8/12–6/29/12
NAEP Expert Teams reviews	5/18/12–7/12/12
REPORTING	
COSDAM update report submitted to Governing Board	2/9/12
Draft pilot report submitted to Governing Board	3/31/12
COSDAM update report submitted to Governing Board	4/19/12
Final pilot report submitted to Governing Board	4/30/12
COSDAM update report submitted to Governing Board	7/5/12
Draft final report submitted to Governing Board	7/31/12
Final report submitted to Governing Board	9/7/12

PROGRESS UPDATE

An initial program analysis identified for each institution the number of qualifying degree programs as well as the courses required for degree completion. There is a total of 41 automotive master technician training programs in the study, and 38 institutions with 61 programs were invited to participate. Of these 61 programs, 12 programs within 10 schools have completed a total of 78 course submissions, including 40 introductory courses (21 reading and 19 mathematics courses) and 38 concluding courses (22 reading and 16 mathematics). Another 23 course submissions are in progress for 14 programs. EPIC continues to call and email instructors in the auto programs to assist with course document submission.

¹ Course artifacts are being gathered for introductory and concluding courses concurrently.

The automotive master technician Review Teams have completed reviews of six courses each in reading and mathematics. The remaining four occupations will begin submitting course documents the week of February 13. EPIC has confirmed the participation of NAEP Expert Teams to conduct the summary review of automotive master technician introductory course findings, and their review instrument is in development. EPIC is preparing a progress report with initial pilot findings to inform and improve data collection and analysis for the remaining four occupational areas.

**2011 National Assessment of Educational Progress in Writing
Achievement Levels Descriptions, Second Field Trial and
Achievement Levels-Setting Study (ALS)
CLOSED SESSION
Susan Cooper Loomis**

The first fully computerized National Assessment of Educational Progress (NAEP) was administered in 2011 for writing at grades 8 and 12. Writing tasks were presented in a variety of formats including written documents and instructions, video clips, and audio clips.

For the writing ALS process, the Body of Work (BoW) method was the process implemented for panelists to judge student performances relative to the NAEP achievement levels descriptions for Basic, Proficient, and Advanced. The Body of Work method is a holistic judgmental process whereby student work is judged according to criteria (NAEP achievement levels descriptions) and classified accordingly. Performances judged to be lower than the description of the Basic achievement level were to be classified as below Basic. Achievement levels-setting panels include teachers, other educators, and representatives of the general public. A total of 100 student booklets were evaluated by each panelist relative to the achievement levels descriptions. Panelists were first given 50 booklets to classify. They used the same 50 booklets to classify booklets two times, with feedback and discussions after each classification. Cut scores were computed for each panelist, and the grade level cut score was computed as the median of the panelists' cut scores for each grade. A different set of 50 booklets were distributed for the third round of classifications, and the third classifications were used to compute the final cut scores.

Achievement Levels Descriptions

Achievement levels descriptions (ALDs) were developed for use in the panel meetings, including the operational achievement levels-setting meetings. A lengthy process involving many writing experts was implemented to develop and evaluate the descriptions that were recommended for use in the 2011 achievement levels-setting process. The Committee on Standards, Design and Methodology (COSDAM) approved the achievement levels descriptions at the August 2011 meeting, with the understanding that final approval would be given in May 2012 to make these the “official” descriptions as part of the overall process of setting the achievement levels.

As a result of the pilot study implemented in November 2011, staff recommended that the achievement levels descriptions be reviewed again. Panelists commented on the ambiguity of some aspects of the descriptions, and staff observed that panelists had a tendency to translate less precise wording to mean lower performance requirements (e.g., “some” was taken to mean “any.”) Further, the results of a study comparing performance on the 2007 and 2011 assessments relative to the 2011 descriptions raised further concerns regarding the achievement levels descriptions. Although several factors could be contributing to the differences resulting from this comparison, COSDAM agreed that the achievement levels descriptions should be re-evaluated.

Further, COSDAM recommended that a small-scale study be conducted to try out the ALDs prior to the operational study, if modifications to the ALDs were made.

Staff worked with content experts to review and modify the achievement levels descriptions. The goal was to make the wording more precise and to further address the calibration of levels within and across grades. As always, the NAEP policy definitions were the primary guide for calibration of the descriptions of what students should know and be able to do. The modified descriptions were reviewed by eight additional content experts, most of whom had been part of the NAEP Writing Framework panel. The recommendations of these experts were evaluated and incorporated to further modify the ALDs.

As a result of comments from panelists in the study to try out the ALDs (Field Trial 2, described below), the ALDs were modified yet again. Panelists considered the description of Basic level performance to be more rigorous than the policy definition for Basic performance. Content experts again worked with staff to modify the descriptions. Finally, agreement was reached that the descriptions were clear and appropriately calibrated across levels and grades. The descriptions used in the operational achievement levels-setting process are included in Attachment B-1. These descriptions are recommended for consideration by COSDAM as part of the overall set of information to be used in reporting results for the 2011 Writing NAEP.

Field Trial #2

One field trial was planned to study the logistics of using two computers in the ALS process. Field trial #1 was implemented in September 2011 with panelists for grade 12 only.

A second field trial was implemented to try out the revised achievement levels descriptions and one other modification to the process. A paper selection process has been used in most NAEP achievement levels-setting procedures, and it was added to the Body of Work process to be implemented for the 2011 Writing NAEP. The addition of this procedure was recommended by the Governing Board staff, and the Technical Advisory Committee for Standard Setting approved the plan presented by Measured Progress (the Writing ALS contractor) for incorporating the training step into the procedure. The paper selection procedure provides an opportunity for panelists to review samples of student work, judge the level of achievement most closely matched to the performance exhibited in the paper, and discuss their judgments with others in the group. After all the papers have been discussed, panelists are then given the scores for the papers. Scores are presented to help panelists understand that there is no systematic correspondence between the rubric scores and the achievement levels represented by the student responses.

The field trial was conducted on January 27, 2012 with 39 panelists: 19 for grade 8 and 20 for grade 12, all recruited within a 50-mile radius of the meeting location in Dover, New Hampshire. This was a sample of convenience in that most panelists serve or had served as scorers for Measured Progress. Panelists included the three types required for the NAEP ALS process: grade-level teachers, non-teacher educators including higher education faculty in the subject matter, and representatives of the general public who are trained in or work in the subject matter

area. The distribution of panelists in each of the three categories was approximately the same as that required for achievement levels setting: 55% teachers, 15% non-teacher educators, and 30% general public representatives. The study was conducted for the 2011 Writing NAEP for grades 8 and 12. Panelists made only one round of judgments in which they classified 50 student booklets into achievement levels categories.

The results of this study using the revised achievement levels descriptions and incorporating the paper selection process as additional training were generally similar to those for the pilot study. For grade 8, the classifications in this single round by field trial panelists resulted in 4% fewer students performing at or above the Basic level than was the case for the pilot study, about 16% fewer performing at or above the Proficient level, and 2% more at the Advanced level. The most noticeable difference was a sizable increase (12%) in the percentage of 12th grade students that would be classified at the Advanced level. The percentage performing at or above the grade 12 Basic level was the same, and the percentage performing at or above the Proficient level increased by 9% over the level for the pilot study.

Operational Achievement Levels-Setting Study

The operational achievement levels-setting study was implemented in St. Louis February 7-10, 2012, with a special study added at the end implemented February 10-11, 2012. The goal was to recruit 30 panelists for each grade level to include 55% writing teachers at the grade level; 15% writing educators, such as college writing instructors, curriculum directors, and other educators who are not currently teaching writing in the K-12 system; and 30% general public representatives who are trained in writing and or currently employed in a position requiring a significant writing component. The content facilitators considered the panelists to be exceptionally strong in terms of their writing experience and knowledge of writing. The panelists included a mayor and several authors as well as outstanding teachers of writing and leaders of writing programs. The panelists for the operational achievement levels-setting study are described in the Table 1 below.

Two additional modifications were made for the operational ALS process. For previous achievement levels-setting studies for the 2011 Writing NAEP, student booklets were presented to panelists in order from lowest student performance to highest. The Technical Advisory Committee for Standard Setting concurred with the Governing Board staff recommendation to change the order of presentation from highest to lowest performance. Panelists were told that they could classify booklets in any order, and they were encouraged to look at one or two booklets at a few locations across the range before they began classifying. The second modification was to have panelists record their level of confidence in the classification of each booklet for the second and third classification. The second classification was made with the same set of booklets used for round 1, but the third classification involved a completely new set of booklets. Staff recommended that the confidence rating be collected to provide information of the relative confidence in the two sets of ratings, as well as to provide data for further research regarding this methodology because the Body of Work method had not been used previously in NAEP achievement levels-setting. Again, this recommendation was discussed with the Technical Advisors, and they approved of the plan.

Table 1
NAEP Writing 2011 Achievement Levels Setting Panelists

Demographic Variable	Attributes	Grade 8		Grade 12		All	
		n	%	n	%	n	%
Panelist Type	Teachers	16	59	15	54	31	56
	Non-teacher Educators	5	19	5	18	10	18
	General Public	6	22	8	29	14	25
Gender	Female	22	81	19	68	41	75
	Male	5	19	9	32	14	25
Race/Ethnicity	Caucasian	23	85	25	96	48	91
	Non-Caucasian	4	15	1	4	5	9
NAEP Region	Midwest	6	22	8	29	14	25
	Northeast	5	19	4	14	9	16
	South	6	22	6	21	12	22
	West	10	37	10	36	20	36

Special Study: A special study was conducted at the conclusion of the achievement levels-setting study, February 10-11, 2012. The purpose of the study was to produce estimates of the performance of students on the 2007 Writing NAEP comparable to that on the 2011 Writing NAEP.

All panelists were invited to participate in the study, but unavailability to participate in the special study did not preclude participation on the ALS panel. A total of 35 panelists (approximately 2/3 of the ALS count) agreed to participate: 17 for grade 8 and 18 for grade 12. The distribution of panelists participating in the special study is presented in Table 2 below.

Table 2
NAEP Writing 2011 Achievement Levels Setting Panelists for Special Study
with 2007 Writing NAEP

Demographic Variable	Attributes	Grade 8		Grade 12		All	
		n	%	n	%	N	%
Panelist Type	Teachers	8	47	8	44	16	46
	Non-teacher Educators	4	24	4	22	8	23
	General Public	5	29	6	33	11	31
Gender	Female	13	76	12	67	25	71
	Male	4	24	6	33	10	29
Race/Ethnicity	Caucasian	15	88	16	94	31	91
	Non-Caucasian	2	12	1	6	3	9
NAEP Region	Midwest	5	29	7	39	12	34
	Northeast	1	6	4	22	5	14
	South	4	24	1	6	5	14
	West	7	41	6	33	13	37

The special study was designed to provide data for comparing performance on the 2007 Writing NAEP to that on the 2011 NAEP when judged relative to the 2011 achievement levels descriptions. The special study was implemented following the pilot study; but because the achievement levels descriptions were modified for use in the operational ALS, it was necessary to implement the study design again using the revised descriptions.

Closed Session March 2, 2012

Outcomes of the achievement levels-setting study and the special study will be reviewed by the Technical Advisory Committee on Standard Setting at a meeting scheduled for February 23-24, 2012. Their recommendations, along with complete study results for the added field trial, the operational ALS, and the special study will be shared with COSDAM for discussion at the meeting on March 2, 2012.

The current schedule calls for a recommendation on 2011 writing achievement levels by COSDAM for action by the Governing Board at the May 2012 meeting in San Antonio.

WRITING ACHIEVEMENT LEVELS DESCRIPTIONS FOR GRADE 4

BASIC

Fourth-grade students writing at the Basic level should be able to address the tasks appropriately and at least partially accomplish their communicative purposes. Texts should be appropriately structured. Many of the ideas in the texts should be developed, and their texts should include supporting details and examples that are relevant to the topic, purpose, and audience. Most sentences should be well structured, and texts may be composed mostly of simple sentences. Many of the words and phrases should be appropriate to the topics, purposes, and audiences. Spelling, grammar, usage, capitalization, and punctuation skills should be sufficiently accurate to convey general meaning, although there may be some errors that detract from meaning.

PROFICIENT

Fourth-grade students writing at the Proficient level should be able to address the tasks appropriately and accomplish their communicative purposes. Texts should be appropriately structured and coherent. Most of the ideas in their texts should be developed effectively, and their texts should include supporting details and examples that support the main ideas. Texts should have well structured sentences and a variety of sentence types—simple, compound, and complex. Words and phrases should be thoughtfully selected and appropriate to the topics, purposes, and audiences. Spelling, grammar, usage, capitalization, and punctuation should be sufficiently accurate to communicate clearly with the reader. There may be some errors in the texts, but these errors should not impede meaning.

ADVANCED

Fourth-grade students writing at the Advanced level should be able to address the tasks appropriately and accomplish their communicative purposes in effective ways. Texts should be well structured and coherent. The ideas in the texts should be developed fully and effectively. Their texts should include supporting details and examples that are closely related to the topic, purpose, and audience and that enhance communicative effectiveness. Sentences should be well structured, and texts should include a variety of sentence types (simple, compound, and complex) to enhance their communicative effectiveness. Words and phrases should be chosen skillfully, and they should both enrich meaning in the texts and enhance communicative effectiveness. Spelling, grammar, usage, capitalization, and punctuation should be mostly accurate and well developed, and they should be used appropriately. Grammatical, mechanical, and usage choices should contribute to communicative effectiveness. There may be a few errors, but they should not impede meaning.

WRITING ACHIEVEMENT LEVELS DESCRIPTIONS FOR GRADE 8

BASIC

Eighth-grade students writing at the Basic level should be able to address the tasks appropriately and mostly accomplish their communicative purposes. Their texts should be coherent and effectively structured. Many of the ideas in their texts should be developed effectively. Supporting details and examples should be relevant to the main ideas they support. Voice should align with the topic, purpose, and audience. Texts should include appropriately varied uses of simple, compound, and complex sentences. Words and phrases should be relevant to the topics, purposes, and audiences. Knowledge of spelling, grammar, usage, capitalization, and punctuation should be made evident; however, there may be some errors in the texts that impede meaning.

PROFICIENT

Eighth-grade students writing at the Proficient level should be able to develop responses that clearly accomplish their communicative purposes. Their texts should be coherent and well structured, and they should include appropriate connections and transitions. Most of the ideas in the texts should be developed logically, coherently, and effectively. Supporting details and examples should be relevant to the main ideas they support, and contribute to overall communicative effectiveness. Voice should be relevant to the tasks and support communicative effectiveness. Texts should include a variety of simple, compound, and complex sentence types combined effectively. Words and phrases should be chosen thoughtfully and used in ways that contribute to communicative effectiveness. Solid knowledge of spelling, grammar, usage, capitalization, and punctuation should be evident throughout the texts. There may be some errors, but these errors should not impede meaning.

ADVANCED

Eighth-grade students writing at the Advanced level should be able to construct skillful responses that accomplish their communicative purposes effectively. Their texts should be coherent and well structured throughout, and they should include effective connections and transitions. Ideas in the texts should be developed logically, coherently, and effectively. Supporting details and examples should skillfully and effectively support and extend the main ideas in the texts. Voice should be distinct and enhance communicative effectiveness. Texts should include a well-chosen variety of sentence types, and the sentence structure variations should enhance communicative effectiveness. Words and phrases should be chosen strategically, with precision, and in ways that enhance communicative effectiveness. An extensive knowledge of spelling, grammar, usage, capitalization, and punctuation should be evident throughout the texts. Appropriate use of these features should enhance communicative effectiveness. There may be a few errors, but these errors should not impede meaning.

WRITING ACHIEVEMENT LEVELS DESCRIPTIONS FOR GRADE 12

BASIC

Twelfth-grade students writing at the Basic level should be able to respond effectively to the tasks and accomplish their communicative purposes. Their texts should be coherent and well structured. Most of the ideas in their texts should be developed effectively. Relevant details and examples should be used to support and extend the main ideas in the texts. Voice should support the communicative purposes of the texts. Texts should include appropriately varied simple, compound, and complex sentence types. Words and phrases should be suitable for the topics, purposes, and audiences. Substantial knowledge of spelling, grammar, usage, capitalization, and punctuation should be clearly evident. There may be some errors in the texts, but these errors should not generally impede meaning.

PROFICIENT

Twelfth-grade students writing at the Proficient level should address the tasks effectively and fully accomplish their communicative purposes. Their texts should be coherent and well structured with respect to these purposes, and they should include well-crafted and effective connections and transitions. Their ideas should be developed in a logical, clear, and effective manner. Relevant details and examples should support and extend the main ideas of the texts and contribute to their overall communicative effectiveness. Voice should be relevant to the tasks and contribute to overall communicative effectiveness. Texts should include a variety of simple, compound, and complex sentence types that contribute to overall communicative effectiveness. Words and phrases should be chosen purposefully and used skillfully to enhance the effectiveness of the texts. A solid knowledge of spelling, grammar, usage, capitalization, and punctuation should be evident throughout the texts. There may be some errors in the texts, but they should not impede meaning.

ADVANCED

Twelfth-grade students writing at the Advanced level should be able to address the tasks strategically, fully accomplish their communicative purposes, and demonstrate a skillful and creative approach to constructing and delivering their messages. Their texts should be coherent and well structured; they should include skillfully constructed and effective connections and transitions; and they should be rhetorically powerful. All of the ideas in their texts should be developed clearly, logically, effectively, and in focused and sophisticated ways. Supporting details and examples should be well crafted; they should skillfully support and extend the main ideas; and they should strengthen both communicative effectiveness and rhetorical power of the texts. A distinct voice that enhances the communicative effectiveness and rhetorical power of the texts should be evident. Texts should include a variety of sentence structures and types that are skillfully crafted and enhance communicative effectiveness and rhetorical power. Words and phrases should be chosen purposefully, with precision, and in ways that enhance communicative effectiveness and rhetorical power. A highly developed knowledge of spelling, grammar, usage, capitalization, and punctuation should be evident throughout the texts and function in ways that enhance communicative effectiveness and rhetorical power. There may be a few errors in the texts, but they should not impede meaning.