

## 12th Grade Preparedness Reporting

These materials have been prepared for the Board discussion about reporting 12<sup>th</sup> grade preparedness and are based on more than 30 studies implemented by Board staff and contractors.

Board members are asked to review the attached materials:

- A. Timeline for the Board's preparedness research and reporting
- B. Overview of the Governing Board's program of preparedness research and key findings
- C. Summary of recommendations for reporting  
(This document includes a summary of the purpose of the research, responses to a number of important questions, and recommendations about reporting college preparedness for particular NAEP scale scores, called reference points.)

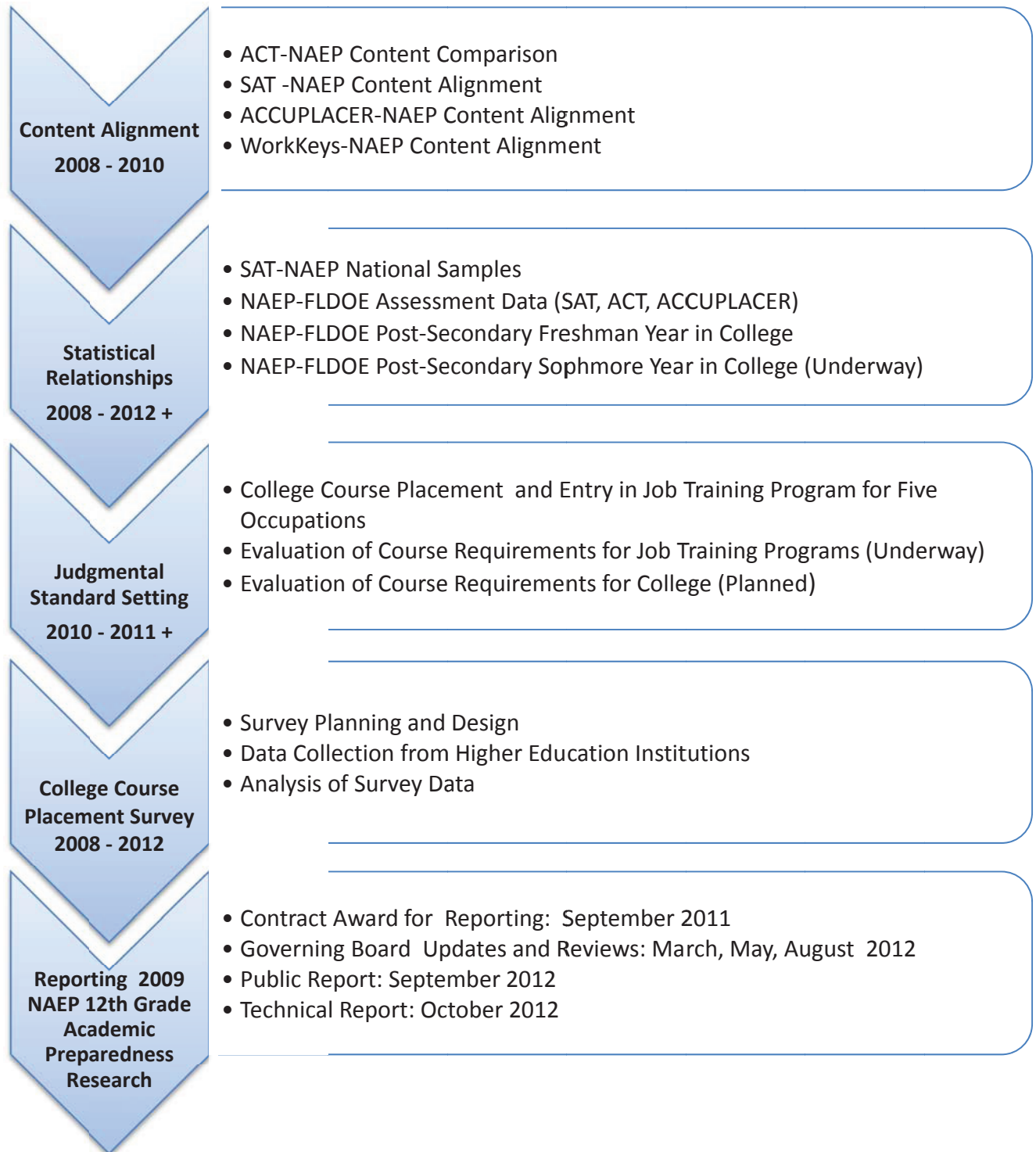
The staff recommendations are based on results from the preparedness studies, as well as input from the technical reviews collected to date. While the recommendations and proposed reporting are based on all available information, more research is needed to further clarify these findings and to expand the scope of what is known. Additional studies are planned to enhance and further explain some of the 2009 findings and to collect additional information from the upcoming 12th grade NAEP reading and mathematics assessments in 2013.

### Discussion Questions:

1. What are the key takeaways you want people to have after reading the NAEP report on 12th grade preparedness?
2. Is the evidence from the Board's preparedness research persuasive in supporting the reference points recommended (e.g., likely to be successful)?
3. Are the reference points and their explanations meaningful, useful, and understandable?
4. Is the rationale sound and explained in a way the public will understand for using a probability of 80% to place the math SAT score of 500 at Proficient for NAEP mathematics and a probability of 50% to place the critical reading SAT score of 500 at Proficient for NAEP reading?
5. The research results to date do not support a reference point for "just academically qualified," which would reflect the Board's working definition for preparedness. This reference point is assumed to be in the range between the proposed reference points for "likely to be successful" and "needs remediation." Research planned for 2013 will be designed to help locate this point on the NAEP scales. Does this seem like a reasonable, supportable approach?

6. What is most important for the Board to address when communicating the preparedness findings to the public, education policymakers, teachers, researchers, the business community, etc.?
7. Do we have the proper balance, for this first report on preparedness research, between making solid claims vs. presenting caveats and plans for future research?

## 12<sup>th</sup> Grade NAEP Prepared Research - Timeline



## Overview of the Board's Preparedness Research Findings

Based on the recommendations of a panel of technical experts, the National Assessment Governing Board has conducted more than 30 studies from 2009 to 2012 to examine the feasibility of using the National Assessment of Educational Progress (NAEP) to report on the academic preparedness of 12th graders for college and job training. The technical panel, composed of seven experts<sup>1</sup> in a variety of measurement and policy areas, was headed by Michael W. Kirst, emeritus professor of education and business administration at Stanford University.

The studies on which this report is based represent the first cycle of the Board's research, using data from the NAEP 2009 assessments in reading and mathematics. Additional studies are planned over the next few years with data from the next administration of the reading and mathematics grade 12 NAEP in 2013.

The research completed so far has been in these four areas:

- ***Content comparisons*** between NAEP and widely-used examinations for college admissions, post-secondary course placement, and workplace skills.
- ***Statistical relationships*** linking NAEP performance data with data from college admissions and placement tests, when available.
- ***A nationally-representative survey of the tests and cut scores used for placement in remedial courses*** at two-year and four-year colleges.
- ***Judgmental standard setting*** by expert panels to identify the NAEP scores needed to qualify for job-training programs or for entry-level college credit courses without remediation.

Two other aspects of this research are worth noting in this summary. First, Florida<sup>2</sup>, one of the eleven states that volunteered for the state-level grade 12 pilot test administration in 2009, has provided substantial assessment and follow-up data on the representative sample of students taking NAEP as 12<sup>th</sup> graders in 2009. These data are being used as validity evidence along with the other studies. Because the Florida data analyses have demonstrated the utility of such data, studies linking data in additional states will be pursued based on the 2013 NAEP.

Secondly, plans for another category of study have not been fully implemented to date. *Benchmarking studies* involve administering NAEP to reference groups of interest, such as

---

<sup>1</sup> Members of the technical panel included John Campbell, David T. Conley, Michael Kane, Mark David Milliron, Robert Mislevy, and George C. Thornton, III. A copy of the technical panel's report is found on the Governing Board's website at <http://www.nagb.org/publications/PreparednessFinalReport.pdf>.

<sup>2</sup> The Board very much appreciates the data Florida has been able to provide. It's inclusion in this report is illustrative and not intended as commentary on Florida policies, programs, or students.

military recruits or enrollees in job-training or college programs, and these studies are planned for future implementation. One pilot study in this area was conducted in the fall of 2010 involving an administration of grade 12 NAEP reading and mathematics to a sample of Texas college freshmen. Although the degree of participation and cooperation from nine Texas public postsecondary institutions was excellent, student participation was too low to produce valid results. Therefore it was determined a full scale operational study with college freshmen was not feasible.

## Content Findings

The ***content comparison or alignment studies*** compare NAEP to other assessments that measure academic preparedness for college and job-related skills. They involve a systematic comparison of the extent to which NAEP and another test measure similar content in a similar way. Although a 100 percent alignment between two different tests cannot be expected, a substantial overlap gives confidence when a strong statistical relationship or correlation between the two is established. Substantial correlations indicate performance on the two tests represents about the same level of achievement in the content, that is, performance on one test represents about the same knowledge, skills, and abilities as performance on the other.

The preparedness research studies compared the test frameworks, specifications, and items for NAEP with four other assessments in mathematics and reading: SAT and ACT (used for college admissions and course placement), ACCUPLACER (used for college course placement), and WorkKeys (used to assess job-related skills).

The studies found similar content in NAEP and the three college-related examinations, providing a firm basis for statistical linking. The content overlap was more apparent in mathematics than in reading. Although the text presentations for NAEP reading and the reading contexts in the other exams are quite different, the alignment studies found that the tests emphasize many of the same or closely related specific skills.

In the comparison between NAEP and WorkKeys, some similarities were found, but also significant differences in both focus and rigor. NAEP is broader and assesses higher level mathematics content and skills than WorkKeys. WorkKeys Applied Mathematics and Reading for Information focus on the application of academic skills in the workplace; NAEP does not.

## Statistical Findings

A ***statistical relationship study*** was conducted between NAEP and SAT, using the scores of over 15,000 public school students in grade 12 in 2009 that took both tests. The matched sample accounted for about a third of the twelfth graders in the NAEP national sample. Nationally, 36% of all twelfth grade public school students take the SAT and the percent of match to NAEP in this study is consistent with that figure. In addition, a study was

conducted in Florida, linking NAEP data to the Florida Department of Education (FDOE) longitudinal database.

The national NAEP-SAT linking study found a very strong correlation (0.91) between the mathematics scales on the two exams and a moderate correlation (0.74) in reading. As a result, two statistical methods (projection and concordance) were implemented to relate NAEP and SAT scores and for placing the College Board's SAT College and Career Readiness Benchmarks on the NAEP scale. The Florida results are similar and confirming of the national results, showing the same pattern of a higher correlation in mathematics than in reading. This is a common pattern in studies correlating performance between two mathematics assessments and between two reading assessments.

Florida was one of the 11 states in the first-time pilot of 12th grade state NAEP in 2009. Via a data confidentiality agreement, the Board's research includes following the state-representative sample of 12th grade students that took NAEP reading and mathematics assessments in 2009 for the first five years after high school. The analyses are based on data provided by the Florida Department of Education, which maintains a longitudinal database that includes college entrance and placement test scores and college performance data for students attending all public colleges in Florida, as well as a complete compilation of data for these students throughout the period of attending public elementary and secondary schools in Florida. In 2009-2010, about 54 percent of the 3,400 students in the 12th grade state NAEP sample enrolled in Florida's public postsecondary institutions, about two-thirds of them in community colleges. The Board will continue the analysis of post-secondary data for Florida.

Although Florida is only one state and its results cannot be generalized to the nation, it has the nation's fourth largest public school enrollment and a highly diverse student population. The Florida analyses are generally consistent with the NAEP scores (scale reference points) obtained through the NAEP-SAT link and the higher education survey, and thus provide confirmatory evidence of the validity of the data being reported. For example, the average performance on NAEP of Florida 12<sup>th</sup> graders is slightly below the national average. Parallel to this, public college freshmen in Florida with an overall grade point average (GPA) of at least B- have an average 12th grade NAEP score slightly below the College Board Readiness Benchmark derived from the national NAEP-SAT link.

## Survey Findings

The Board's *survey of post-secondary institutions* is the first ever undertaken to obtain nationally representative data on the tests and cut scores used by colleges for placement in entry-level credit-bearing courses or remedial courses. The study was conducted for the Board in fall 2011, and included a nationally representative sample of 1,700 two-year and four-year colleges. An excellent response rate of 86 percent was obtained.

The survey asked the institutions to identify the placement tests used, and the cut scores in reading and mathematics below which students were deemed to need remedial courses

and at or above which they would be placed in credit-bearing college entry-level courses without remediation. At two-year colleges data were requested only for courses in programs for which four-year colleges award transfer credits.

Overall, the median score below which students were identified as likely to need remediation was 470 on the SAT mathematics test and 453 on the SAT in critical reading. These scores are the basis for the “likely to need remediation” reference points on the NAEP scale. The SAT, generally used as an admissions test, was used for remedial placement by only 17 percent of post-secondary institutions in math and 11 percent in reading. However, the analyses of data from Florida are generally consistent with the SAT survey scores reported for remediation.

## Standard Setting Findings

The *judgmental standard-setting studies* were designed to identify reference points on the NAEP scale for entry into job training programs or placement into entry-level college credit courses that meet general education requirements.

The five occupational areas included were for automotive master technician, computer support specialist; heating, ventilation, and air conditioning technician (HVAC); licensed practical nurse (LPN); and pharmacy technician. All of these occupations require at least three months of job training beyond high school but not a bachelor’s degree. All are large and growing fields of employment, and have the potential for reasonable wages. All have a “cross-walk” to military occupations and training programs. The Board is still seeking an agreement with the military to further explore the relationships with NAEP performance.

Nearly 200 job-trainers and 40 college faculty members were assembled separately in groups of approximately 40 for each occupation and college for a week of standard-setting sessions. Each group was divided into two replicate panels in reading and two in mathematics. The job trainers were asked to define the point on the NAEP scale that represents the minimal level of knowledge and skills needed to enter their training programs, that is, to be academically prepared. The faculty from a range of colleges and curricular areas went through a similar exercise for placement in entry-level college courses that fulfill general education requirements.

For more than half of the replicate panels, the average cut scores on NAEP differed significantly from each other. The members of some panels differed widely among themselves. In several cases the recommended scores did not align with other information; overall, the recommended scores were not confirmed by other information on preparedness.

The judgmental studies confirmed a finding from the test content comparisons: The NAEP assessments in mathematics and reading cover the academic skills needed for college entrance and credit-course placement, but the content of 12th grade NAEP may not be well aligned to job training requirements. NAEP assesses content not required for the job training programs; and for content that is required, NAEP measures it in a more academic or

theoretical—not applied—manner. Comments from some of the job-training participants suggested that the 8<sup>th</sup> grade mathematics NAEP may provide a better assessment of the knowledge, skills, and abilities needed for their programs.

Because of the internal inconsistencies in the standard setting studies and the lack of mutually confirmatory information from the other preparedness studies, results of the standard-setting studies are not being used in this report to identify NAEP scale scores or reference points for the reporting of academic preparedness.

## **Technical Reports**

An online compendium of technical reports for all studies will provide full information about all of the Board’s preparedness research.

## Summary of Recommendations and Rationale NAEP 12<sup>th</sup> Grade Preparedness Research %

### Primary Research Purpose

The fundamental purpose of this research was to determine whether it is possible to use NAEP reading and mathematics at grade 12 to report on the academic preparedness of U.S. students for college and job training. The Board's working definitions of preparedness which follow have guided the development of the research being reported.

Students have the reading and mathematics knowledge and skills to qualify for ...

- entry-level college credit courses that meet general education requirements, without the need for remedial coursework.
- a job training program without remediation.

### Important Questions, Answers, and Rationales

Below you will find the staff response to important questions about the research findings along with supporting and explanatory statements.

#### **Q1. %Can NAEP be used to inform the national discussion about the academic preparedness of U.S. students?**

A1. **Yes**, but some qualifications are needed.

- More is known about preparedness for college than for training programs.
- The content of NAEP is a good measure of college preparedness, but some NAEP content does not appear to be relevant for some job training programs.
- The level of academic preparedness for college coursework appears to be different from that required for entry in job training in the programs included in the NAEP research.
- The NAEP reporting of 12<sup>th</sup> grade preparedness for college is intended to apply to the placement (not admissions) policies of "typical" postsecondary institutions (i.e., institutions of moderate selectivity) and recognizes that admissions and placement policies vary for postsecondary institutions by selectivity and for specific courses or majors.

**Q2. %Will more than one preparedness reference point on the NAEP scales be established?**

A2. **Yes**, but the meaning of these reference points for reporting on the 2009 data will apply only to preparedness for college.

- The NAEP item pool and the NAEP frameworks used in the 2009 12th grade reading and mathematics assessments were similar to nationally recognized tests used for college admissions and placement.
- Two reference points are recommended for reporting the 2009 results:
  - Likely to be successful in college, and
  - Likely to need remediation.
- The research results to date do not support a reference point for “just academically qualified,” which would reflect the Board’s working definition for preparedness. This reference point is assumed to be in the range between “likely to be successful” and “needs remediation.” Research planned for 2013 can be designed to help locate this point on the NAEP scales.
- The meaning of the reference points must be explained and understood, because the statistical and policy considerations used as the basis for locating the reference points differs somewhat for reading and mathematics.
- The reference points rely primarily on the linking study for NAEP and SAT scores and the higher education survey results, but the other studies provide confirmatory evidence and have informed the placement of these reference points.

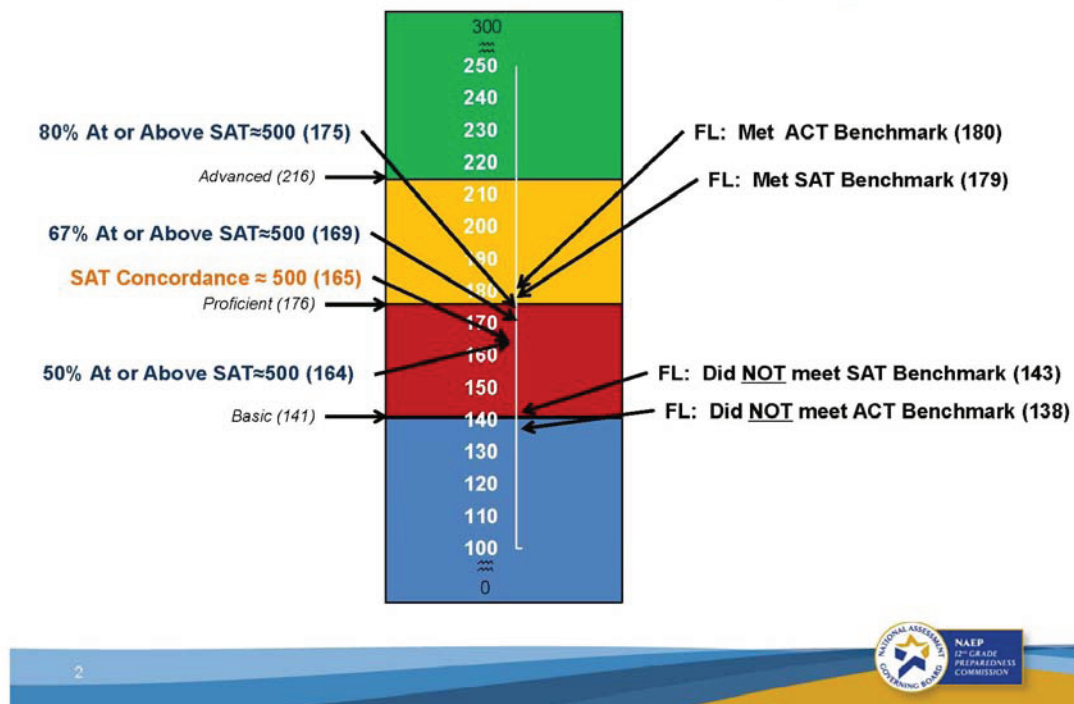
**Q3. % What process was used to determine the recommended reference points and which findings to report?**

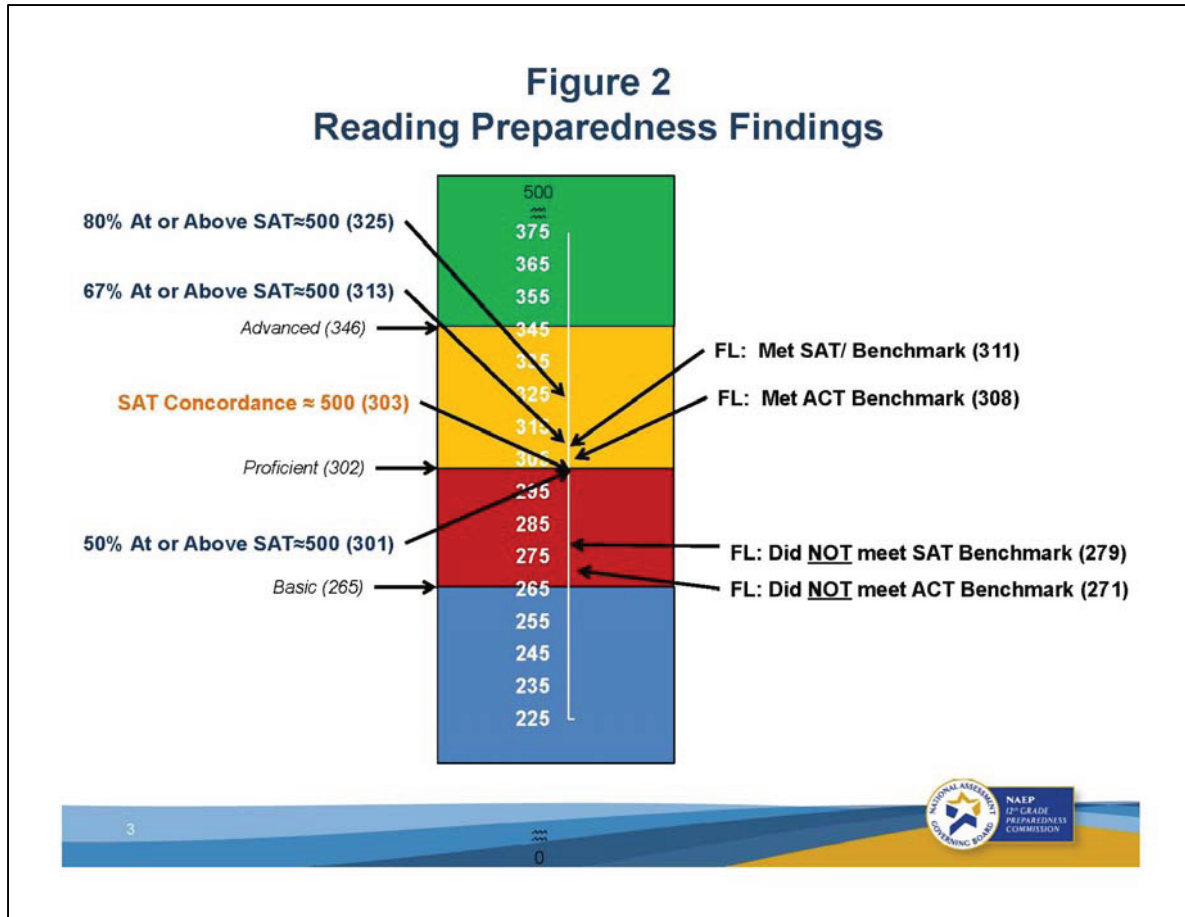
A3. Staff and technical advisors have reviewed and discussed the results of the studies and other available evidence to determine the extent to which findings are mutually confirming. From these reviews, staff developed evidence-based score interpretations for the NAEP 12<sup>th</sup> grade reading and mathematics scales related to academic preparedness and staff reviewed these interpretations with technical advisors. Figures 1 and 2 illustrate some of the results of the data analyses and Figures 3 and 4 illustrate the interpretations of the recommended reference points on the NAEP scales.

- It is important that the reporting and presentation of the preparedness research results be transparent and complete. When results from a single study or analysis are not confirming of results from other studies, this should be included in the report to help readers better understand the limitations of the validity evidence and provide a complete context of preparedness.
- Further research has been identified to explore more deeply the reasons for the findings. For example, a review of course content has been implemented to provide more understanding about the minimal level of academic preparedness needed for the five occupational areas and college.

- There is mutually confirmatory evidence for reporting both reading and mathematics reference points for college.
- The reference points for reading and mathematics do not require the use of identical probabilities to be useful.
- Reference points on the NAEP scale for job training programs are not supported by the research findings to date.
- Plain language should be used to identify and describe the reference points so they are meaningful and useful. The language should not obfuscate the concept with sophisticated statistical and probabilistic wording or caveats.
- Limit the number of reference points identified and have them be generally representative of a “typical” institution.
- Reference points should be informed by, but not necessarily equivalent to the college readiness benchmarks identified by the College Board or ACT.
- Information provided about the reference points should not make claims that imply precision that is not supported by the analyses.

**Figure 1**  
**Mathematics Preparedness Findings**





**Q4. What statements can be made about preparedness for job training?**

A4. Although no reference points for the NAEP scale are supported by the results of the standard-setting studies, some statements about job training programs and NAEP are supported by the research.

- The NAEP 2009 12th grade reading and mathematics assessments cover a range of content that appears to be much broader than needed to qualify for some training programs that do not require a two- or four-year college degree.
- The participants in the job training standard-setting studies described minimal academic preparedness for job training programs as being less rigorous and having a different content focus than the NAEP at 12<sup>th</sup> grade. They found many of the NAEP items in reading and mathematics to be irrelevant to the knowledge and skills needed to qualify for the respective training programs.

## Reference Point Recommendations

### Mathematics Reference Points (Shown in Figure 3)

#### **176—Likely to be successful in college** (NAEP Proficient Level cut score)

EXPLANATION: This point is based on the linking study between NAEP and the SAT. Students who score at 175 on NAEP have an 80 percent probability of meeting the College Readiness Benchmark<sup>i</sup> of 500 on SAT mathematics as established by the College Board. The SAT benchmark defines success as a 65 percent probability of achieving a freshman year grade point average of B- or higher. The average NAEP score of Florida students who met the SAT benchmark was 179. The NAEP proficient cut score is 176, just one point higher than the score from the SAT study and three points lower than the results of the Florida study. Given the confirmatory nature of these results and the error inherent in sample estimates, the staff recommendation is to use the proficient cut score as the reference point.

12<sup>th</sup> graders scoring at or above **176** in 2009

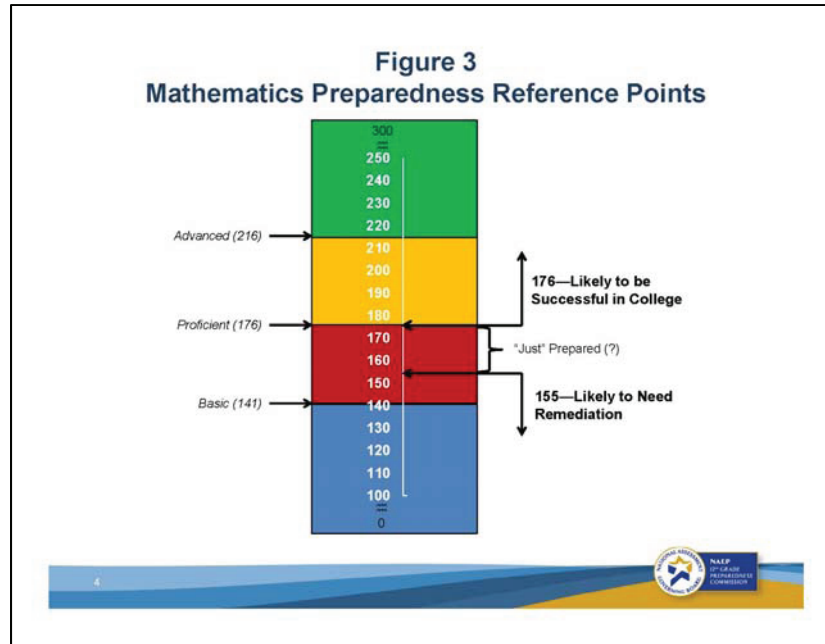
**26 percent**

#### **155—Likely to need remediation** (In the middle of the NAEP Basic Level)

EXPLANATION: This point is based on the median score from the higher education survey<sup>ii</sup> indicating an SAT score of 470 as the score below which freshmen are deemed to need remediation to be prepared for entry-level credit-bearing courses in college mathematics. (At two-year colleges scores were requested only for courses which would qualify for transfer credit at four-year institutions.) The score is linked to the NAEP scale based on the linking study between NAEP and the SAT. A conservative estimate of the percent of students likely to need remediation is more prudent as this point in the reporting of preparedness; therefore, the recommended score of 155 on NAEP is for students who have a 50 percent probability of attaining an SAT score of 470. The average score of Florida students who did not meet the SAT college readiness benchmark is 143, 12 points lower.

12<sup>th</sup> graders scoring below **155** in 2009

**51 percent**



Reading Reference Points (Shown in Figure 4)

Introductory comments: Use of a lower probability of meeting the SAT College Readiness Benchmark for reading (i.e., 0.50) than for mathematics (i.e., 0.80) is recommended. There is more uncertainty in the relationship between performances on the two reading tests than on the mathematics tests. The 50 percent probability was chosen for reading to ensure against underestimating the percentage of students likely to be successful in college and overestimating the percent in need of remediation. In addition, the NAEP-SAT content comparison studies, while finding significant overlap in both subjects, found more convergence in mathematics than reading.

### **302—Likely to be successful in college (NAEP Proficient Level cut score)**

EXPLANATION: This point is based on the linking study between NAEP and the SAT. Students at this level have a 50 percent probability of meeting the college readiness benchmark<sup>i</sup> of 500 on SAT critical reading as established by the College Board. The SAT benchmark defines success as a 65 percent probability of achieving a freshman year grade point average of B- or higher. Florida students who met the SAT College Readiness Benchmark scored 311 on NAEP.

12<sup>th</sup> graders scoring at or above 302 in 2009

**38 percent**

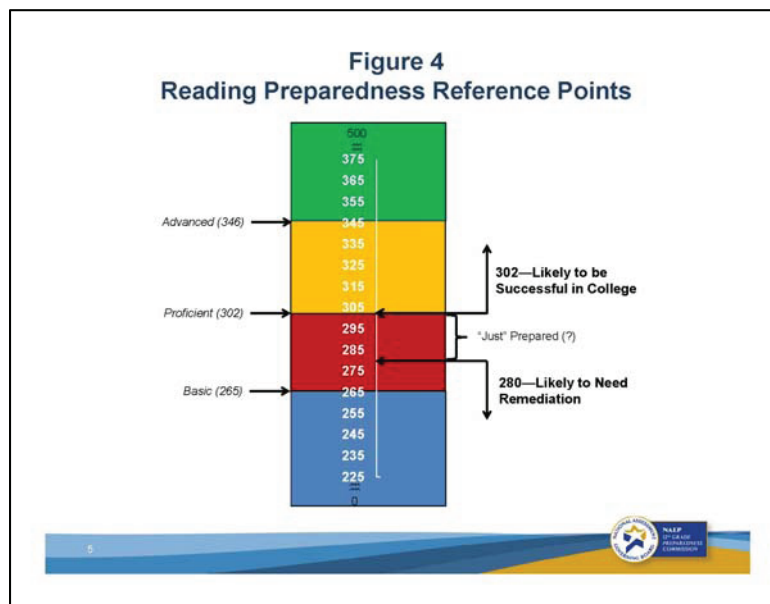
### **280—Likely to need remediation (In the middle of the NAEP Basic Level)**

EXPLANATION: This point is based on the median score from the higher education survey<sup>ii</sup> indicating an SAT critical reading score of 450 as the score below which

freshmen are deemed to need remediation in reading to be prepared for entry-level credit-bearing courses. (At two-year colleges data was requested only for programs awarding transfer credit recognized at four-year institutions.) The score is linked to the NAEP scale based on the linking study between NAEP and the SAT critical reading exam. Florida students who did not meet the SAT College Readiness Benchmark scored 279 on NAEP. Results from other published data about enrollment in remedial education study confirm this score interpretation.

12<sup>th</sup> graders scoring below 280 in 2009

**39 percent**



<sup>i</sup> According to a College Board statement, the SAT college and career readiness benchmark “indicates a 65 percent likelihood of achieving a B- grade-point average or higher during the first year of college, which in turn is indicative of a high likelihood of college success and completion.” The benchmark score on the SAT tests was computed using logistic regression and based on a 65 percent probability of obtaining a freshman-year GPA of at least B- minus, using pooled data for all different academic majors and coursework. The data were gathered in a national sample of 110 four-year colleges and universities using the SAT. There has been no College Board study of these benchmarks involving two-year community colleges or occupations that do not require a four-year college degree. Although the College Board terms them as benchmarks for “college and career readiness,” the definition of the benchmarks makes no reference to success in those institutions or those career fields.

<sup>ii</sup> The working definition of preparedness was the basis of a nationally representative survey of two-year and four-year colleges about the exams and cut scores used for placement into remedial

---

or developmental courses at post-secondary education institutions. The survey, sponsored by the National Assessment Governing Board, asked institutions to identify the tests used [mathematics or reading] to make placement decisions and the “score below which developmental or remedial ...courses were needed.” The instructions explain: “Students scoring below this level are in need of remedial or developmental courses. Students scoring at or above this level are placed into entry-level courses.” Two-year colleges were instructed that responses were to be with respect only on the tests used to evaluate “entering students who are pursuing a degree program that is designed for transfer to a four-year institution.”