

On NAEP and 12th Grade Preparedness

Discussion Draft: Hypothetical Statements for NAEP Reports

Ray Fields
July 31, 2009

At its March 2009 meeting, the Governing Board adopted a program of research to serve as validity evidence for statements about 12th grade preparedness for postsecondary education and training that might be made in NAEP reports. The Governing Board also adopted a working definition of “preparedness” for use in connection with those studies. At the May 2009 Board meeting, members of the Committee on Standards, Design and Methodology requested that staff prepare examples of statements about 12th grade student preparedness that might be made, based on possible outcomes of the program of research.

On the pages that follow are examples of statements related to 12th grade student preparedness that plausibly could be made in NAEP reports, depending on the outcomes of the Board-adopted program of research. Three scenarios are provided: one in which the research results are convergent and confirmatory, one in which the results are mixed, and one in which there are no convergent results. The intent is to provide examples for discussion purposes and to illustrate a range of possibilities. The statements presented are not intended to be exhaustive and, for the sake of brevity, focus only on postsecondary education and do not include training for occupations. Further, the statements do not include subgroup results, which will be very important to report.

The Technical Panel on 12th Grade Preparedness Research recommended that reports of 12th grade results include a definition and explanatory context of preparedness as used in NAEP. The purpose is to ensure that readers will understand what the statements about 12th grade preparedness mean and the limits of what NAEP can assess. An example follows of what a context statement might contain. The example focuses on preparedness for postsecondary education and includes some assumptions about the outcomes of validity research that is just beginning to get underway. Clearly, a context statement applicable to preparedness for training for occupations is missing and would need to be added. The example is offered for the purpose of critique and discussion:

- Are the right elements present?
- Are there missing elements that should be added?
- Should anything be deleted?
- Does it have the right tone?
- Should this kind of text should be included in the body of NAEP reports on preparedness or elsewhere (e.g., a technical appendix)?

Following the context description are the three scenarios describing possible research outcomes and a limited set of potential statements for each that might be made in NAEP reports. Some data mentioned in the scenarios, while based on the approved program of research, may not be available in time to inform initial reporting on 12th grade preparedness. Among the scenarios, a range of hypothetical statements is presented, from the very narrow and specific to the broad and general.

For example, a narrow, specific statement might be

X percent of U.S. 12th graders in 2009 scored at or above [NAEP Score] on the NAEP reading assessment. A panel of experts who make postsecondary placement decisions found that the knowledge and skills represented by a score at or above [NAEP Score] in reading are consistent with the types of reading demands found in textbooks required for entry level college credit courses in subjects such as history or psychology.

A general statement might be

X percent of U.S. 12th graders in 2009 are prepared for entry level credit bearing courses in college mathematics.

The research results may support only one narrow statement, more than one narrow statement, a sweeping statement, or no statement. Consequently, judgments will be needed in distilling and interpreting the research results, in shaping and refining the potential supportable statements, and in deciding which statements, if any, should be included in a NAEP report. No discussion is provided in this document about a process for making these judgments, but such a process will be needed and is suggested as a topic for future consideration.

As reflected in the potential statements associated with the scenarios below, these judgments may involve subtle distinctions in language to ensure accuracy and thus, may be subject to misinterpretation. For example, potential statements may be framed in probabilistic terms (e.g., “likelihood of placement...”) or in terms of expert judgment (e.g., “a panel of experts found...”), but be interpreted in secondary reporting as a broad generalization (e.g., X percent of students are prepared...”). Consequently, affirmative steps should be taken to help readers avoid mis- or over-interpretation of preparedness statements in NAEP reports.

Finally, in addition to fidelity to the data, the meaningfulness and utility of preparedness statements to the public, policymakers, and educators should be an uppermost consideration in their formulation and the decision on whether such statements should be included in NAEP reports.

EXAMPLE CONTEXT STATEMENT FOR PREPAREDNESS FOR POSTSECONDARY EDUCATION

Postsecondary education and training are becoming increasingly important for individuals and for the nation. For individuals, postsecondary education and training provide prerequisites for employment in occupations that offer advancement and good pay, a fulfilling life, and participation as effective citizens in our democracy. For the nation, postsecondary education and training provide the prospect of a workforce able to compete in the global economy and a citizenry with the capacity to understand the complex issues and choices faced by public policy makers.

In the United States today, there is no single, agreed upon definition of “prepared for postsecondary education.” Postsecondary education in the U.S. is a complex mix of institutions, public and private, that have different admission requirements and different procedures and criteria for placing individual students into education programs.

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

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

Further, readiness for postsecondary education involves more than academic proficiency; it includes personal qualities and habits of mind as well, such as motivation, persistence, and time management, to name a few, qualities that NAEP does not measure.

NAEP’s strengths as a source of data about 12th grade preparedness reside in the high quality and rigor of its test frameworks and questions; the fact that it is the only continuing, nationally representative source of data about 12th grade achievement in reading, mathematics, and other core subjects; and the integrity of its testing and quality control procedures. However, NAEP does not produce individual student scores, only group results. The absence of student level data precludes direct longitudinal studies of NAEP test takers, which limits the kinds of preparedness validity research that can be conducted. In addition, the fact that NAEP does not provide results to test-takers, along with the fact that there are no direct benefits or consequences for either high or low performance, suggests that different factors related to motivation may be present than would be the case with tests that do provide individual results and have consequences, such as college admission and placement tests.

Given the complex nature of postsecondary education, the lack of agreement on a single definition of “preparedness,” and the strengths and limitations of the National Assessment of Educational Progress (NAEP), great care has been taken to ensure that information about the preparedness of 12th grade students for postsecondary education is valid and in a form useful to the public and policymakers.

The definition of preparedness used in NAEP focuses on the academic qualifications generally required of entry level students for placement into credit-bearing courses that meet general education requirements in postsecondary education institutions. Twelfth-grade students who are prepared for postsecondary education would not require placement into remedial or developmental coursework in reading or mathematics as entry level college students.

Although the term “preparedness” is intended to reflect academic qualification for placement, it is not intended to imply success or predict grades in first year courses nor predict the likelihood of obtaining a degree. This is because, as noted above, there are many other factors beyond academic qualifications that are associated with success in college and the likelihood of completing a degree, and NAEP does not measure these factors.

The definition of and the statements about preparedness made in NAEP reports were validated through a thorough and comprehensive program of research. This program of research involved nineteen studies encompassing four types:

- Studies of the degree of content alignment between NAEP 12th grade reading and mathematics assessments and standardized tests used for college placement;
- Statistical linking studies between NAEP 12th grade reading and mathematics assessments and tests used for college placement;
- Judgmental standard-setting processes involving postsecondary experts to set cut-scores on NAEP 12th grade reading and mathematics assessments, at and above which represent “preparedness” and below which represent a high likelihood of placement into remedial/developmental courses;
- A survey of two-year and four-year institutions’ use of tests and cut-scores for placing students into entry level college credit coursework and into remedial/developmental courses.

A detailed description of the validity studies is provided in technical appendix A (Note: technical appendix A would be included in a NAEP report but is not included here). The results of the research studies should be interpreted with caution, because the purposes, administrative procedures, testing environment, and motivation levels of students under the various testing programs and NAEP are different. The statements about preparedness have been carefully developed to reflect accurately the validity evidence on which they are based. These statements should not be revised or simplified in secondary reporting of NAEP results.

RESEARCH SCENARIOS AND POTENTIAL STATEMENTS FOR NAEP REPORTS

Scenario 1

Research Outcomes: Validity research results for postsecondary education are convergent and consistently confirmatory. The content alignment studies show a high degree of alignment between NAEP and the comparison tests (e.g., SAT, ACCUPLACER), although none are perfectly aligned and some differences are cited. Subject matter experts perform judgmental standard-setting work yielding a cut-score on the 12th grade NAEP mathematics scale representing the skills and knowledge to qualify for placement into entry-level college credit mathematics courses and below which students would require remedial/developmental courses.

The statistical linking studies of NAEP to the comparison tests produce scale concordances. The survey of tests and cut-scores for postsecondary placement decisions yields identical mean and median cut-scores for 2-year and 4-year institutions that also match college readiness benchmarks and standards in the public domain.

Course placement decisions for Florida 12th grade students who participated in NAEP in 2009 are analyzed in terms of average NAEP scores and average admission/placement test scores of students placed into entry level credit-bearing courses and into remedial/developmental programs. For the placement tests, the average scores match those in the survey findings. The NAEP cut-score set by the subject matter experts is within one standard error of the average NAEP score of Florida students placed into entry-level credit bearing college mathematics courses. The same NAEP cut-score matches the score point at which NAEP and college benchmarks are linked.

Potential Statements for NAEP Reports:

Mathematics

1. X percent of U.S. 12th graders in 2009 scored at or above [NAEP Score] in mathematics. According to research conducted by the National Assessment Governing Board, the academic skills and knowledge represented by a score of [NAEP Score] or higher are consistent with those required for placement into entry level college-credit mathematics courses in postsecondary education institutions.
2. X percent of U.S. 12th graders in 2009 scored below [NAEP Score] in mathematics. According to research conducted by the National Assessment Governing Board, the academic skills and knowledge represented by a score below [NAEP Score] are generally not sufficient for placement into entry level college-credit coursework in postsecondary education institutions and suggest a lack of proficiency requiring remedial or developmental coursework.
3. X percent of U.S. 12th graders in 2009 scored at or above [NAEP Score] in mathematics. Scores at or above [NAEP Score] suggest a high likelihood of qualifying academically for entry level college credit courses in mathematics. Scores below this level suggest a high likelihood of deficiency requiring placement into remedial or developmental coursework in mathematics.
4. X percent of U.S. 12th graders in 2009 are likely to be placed into standard, entry level credit bearing mathematics courses, if admitted as first year college students.
5. X percent of U.S. 12th graders in 2009 are likely to be placed into remedial or developmental mathematics courses, if admitted as first year college students.
6. X percent of U.S. 12th graders in 2009 have the mathematics skills and knowledge considered necessary for placement into standard, entry level college mathematics courses.

7. X percent of U.S. 12th graders in 2009 are prepared for entry level courses in college mathematics courses.

Analysis/Discussion:

In this scenario, a wide range of potential statements might be considered supportable. The differences in the statements are somewhat subtle and intended to reflect a situation in which, even where there is consistent confirmatory evidence, more and less conservative interpretations could be made of the set of research results.

The first three statements report the percentage of students at a particular NAEP score level and then, on the basis of the research, provide an interpretation of the meaning of that score without directly attributing those qualities to the students at or above/below the designated score. The remaining four potential statements do assign attributes related to preparedness to a designated percentage of students, either in a probabilistic formulation related to placement (4 and 5), as a statement that qualifying skills and knowledge are possessed by the students (6), and in the form of a conclusion about the percentage of students that are “prepared” (7). Statements 4 and 5 would rely more on the results of the empirical studies, whereas statement 6 would rely more on the results of the judgmental studies.

Statement 7 represents the least conservative interpretation of the study results and, at the same time, may best reflect the ultimate goal of reporting on 12th grade preparedness. The question is—would evidence beyond that provided in the scenario be needed to include statement 7 in a NAEP report?

Scenario 2

Research Outcomes: Validity research results for postsecondary education are mixed. The content alignment studies show a moderate degree of alignment between NAEP and the comparison tests (e.g., SAT, ACCUPLACER), but not so little alignment as to call into question the conduct of statistical linking studies between NAEP and the comparison tests. The statistical linking studies of NAEP to the comparison tests yield inconsistent results, especially for demographic subgroups.

The survey of the use of tests and cut-scores for postsecondary placement decisions yields lower mean and median cut-scores for 2-year than for 4-year institutions. College course placement decisions affecting Florida 12th grade students who participated in NAEP in 2009 were analyzed in terms of average NAEP scores and average admission/placement test scores of students placed into entry level credit-bearing courses and into remedial/developmental programs. For the placement tests, the average scores match the means for 2-year institutions in the survey, but are below the means for 4-year institutions.

Subject matter experts perform judgmental standard-setting work yielding a cut-score on the 12th grade NAEP reading scale that represents the reading skills to qualify for placement into entry-level college credit coursework, such as history or psychology, and

below which students would require remedial/developmental courses. This cut-score is one standard error above the average NAEP score of Florida students placed into entry-level credit bearing college courses in 2-year institutions but more than two standard errors below the average NAEP score of Florida students in credit bearing courses in 4-year institutions. The cut-score set by the subject matter experts is one standard error above the cut score for Proficient in reading at the 12th grade.

Potential Statements for NAEP Reports:

Reading

1. X percent of U.S. 12th graders in 2009 scored below [NAEP Score] in reading. A panel of experts in postsecondary placement found that the knowledge and skills represented by a score below [NAEP Score] in reading is insufficient for the types of reading demands at the college level for entry level courses, such as history or psychology. In a study that followed the college course placement decisions affecting Florida students who took the NAEP 12th grade reading assessment in 2009, an average score below [NAEP Score] in reading was associated with an increased likelihood of being placed into a remedial/developmental course in reading as a first year college student.
2. X percent of U.S. 12th graders in 2009 scored at or above the Proficient level in reading. The reading skills at or above the Proficient level are generally consistent with those required for placement into entry-level, credit-bearing courses in 2-year institutions.

Analysis/Discussion:

The case of mixed results could present special challenges in formulating a defensible statement for use in NAEP reports. Of particular concern is the question of how to address disconfirming data. Should such data be included as a part of the statement, as a footnote, in an appendix, etc.?

Also, there are questions (intended for discussion) about whether the statements should be further refined. In statement 1, the [NAEP score] is one standard error above the average NAEP score of students in credit-bearing courses in 2-year institutions and two standard errors below the average NAEP score of students in credit-bearing courses in 4-year institutions

Thus, while it is accurate to say that “an average score below [NAEP Score] in reading was associated with an increased likelihood of being placed into a remedial/developmental course in reading as a first year college student,” there are other NAEP scores for which this statement also may be true (e.g., for average scores between [NAEP score] and the average for students in credit-bearing courses in 4-year institutions). Should the statement be revised to limit it to “first-year college student in a 2-year institution”? Revised in some other way? Removed from consideration as not supported by the research outcomes?

In statement 2, “generally consistent” is included in the statement because the cut score set by the subject matter experts is one standard error above the cut-score for Proficient. Is this an adequate and appropriate qualifier? The qualifier “2-year institutions” is based on the survey results for 2-year institutions and the confirmation in the Florida data—is this sufficient to support the statement? Should the statement be revised? Removed from consideration as not supported by the research outcomes?

Scenario 3

Research Outcomes: The data are too divergent to support statements about preparedness. Under this scenario, an alternate approach might be selected, using selected items as benchmarks for reporting, or selected research results.

Potential Statements for NAEP Reports:

Selected Item:

1. X percent of U.S. 12th graders in 2009 received full credit on an extended constructed response question in reading (or mathematics) that exemplifies the ability to perform tasks generally expected of entry level college students, such as
 - i. explain the basis for questioning an assertion that is not logically supported;
 - ii. evaluate data or sources for the quality of content;
 - iii. construct well-reasoned arguments to explain phenomena or defend a point of view;
 - iv. analyze competing and conflicting descriptions of an issue to determine the strengths and flaws in each description;
 - v. determine the type of precision appropriate to a particular task; or
 - vi. generate strategies to solve routine and non-routine problems¹)

(Note: the statement would be framed in terms of the content of the selected test item and accompanied by the test item along with rubrics, exemplar responses, and student data.)

Selected Research Result:

2. X percent of U.S. 12th graders in 2009 scored at or above [NAEP Score] in reading on the NAEP 12th grade reading assessment. A panel of experts in college placement determined that the reading skills represented by a score at or above [NAEP Score] are consistent with the reading demands found in the most widely used textbooks used in an introductory college course, such as history or psychology.

Analysis/Discussion:

Item level reporting was the initial approach used in NAEP. It did not support generalizations beyond performance on the item itself and left the public to its own devices to infer the meaning of the results. The unsatisfactory nature of this approach resulted in the evolution in NAEP reporting that has included the development of a NAEP reporting scale for each subject and grade, anchor-level descriptions of the content

¹ Adapted from Conley, D.T. (2007). *Toward a comprehensive conception of college readiness*. Eugene, OR: Educational Policy Improvement Center.

at particular points on the NAEP scale, and achievement levels. Would item level reporting be a satisfactory and sufficient approach for preparedness? Should it be considered, if necessary, as an initial step, with the intent to continue conducting additional preparedness research?

Using selected research results may be problematic because of the divergence and potentially disconfirming nature of other research results. The basis or criteria on which the selection would be made should be considered. How the non-selected research results would be addressed or explained in a NAEP report also should be discussed.

Finally, whether either of these approaches meets the goal of being meaningful and useful to the public would need to be decided.