

12<sup>th</sup> Grade NAEP as a Measure of College and Employment Preparedness

David T. Conley, Ph.D.  
Professor, University of Oregon  
Member, NAEP Technical Panel on 12<sup>th</sup> Grade Preparedness Research

## + Charge to the Panel

- Determine the technical feasibility of utilizing 12<sup>th</sup> grade NAEP results as measures of college and career readiness
- Assist the National Assessment Governing Board in planning research and validity studies that will enable the National Assessment of Educational Progress (NAEP) to report on the preparedness of 12<sup>th</sup> graders for postsecondary education and job training

## + Preparedness vs. Readiness

- *Preparedness* is defined as a subset of *readiness*. Specifically, *preparedness* focuses on academic qualifications, which are measured by NAEP
- *Readiness* includes behavioral aspects of student performance—time management, persistence, and interpersonal skills, for example—which are not readily measurable by NAEP

## + Definition of Preparedness

- *Preparedness for college* means that the student has the reading and mathematics knowledge and skills to qualify for entry level college credit coursework *without the need for remedial coursework* in mathematics or reading.
- *Preparedness for workplace training* means that the student has the reading and mathematics knowledge and skills needed to qualify for *job training*; it does not require that the student is ready to be hired for a job.
- *Preparedness* does not mean *success* in postsecondary education and training.
- *The operational definition of preparedness is limited to academic skills in reading and mathematics as assessed by NAEP*

## + Distinguishing Achievement from Preparedness

- Achievement and preparedness are conceptually different
  - E.g., the NAEP cut score indicating a proficient learner may be different from (higher or lower) than the cut score indicating performance that enables placement into a non-remedial credit-bearing general education course
- NAEP reporting for grade 12 reading and mathematics would include achievement levels as well as preparedness reference points, both on the NAEP scale
- Reference points can be thought of as areas on the NAEP score scale that identify indicators of preparedness for specific postsecondary contexts

## + Panel's Conclusions and Strategy

- *The Technical Panel concluded that, with the appropriate validity studies, reporting on preparedness using NAEP seems feasible*
- The panel debated two strategies to determine feasibility
  - Strategy A: Conduct research to identify whether the NAEP assessment has the properties needed to report on pre-specified statements of preparedness
  - Strategy B: Identify external indicators of preparedness and determine how those indicators relate to the NAEP scale
- The Panel chose Strategy B
  - The intent is to triangulate data from distinctly different sources to illuminate as fully as possible the relationships of interest

## + Multi-method Validation Approach 7

1. Content alignment
2. Statistical relationship
3. Criterion-referenced judgmental standard-setting studies
4. National surveys

## + 1. Content Alignment 8

- Content alignment studies help identify the degree to which the objectives or content descriptors are similar between tests, and the degree to which different types of items on NAEP and the other test are substantially similar or different
- A content alignment study should compare two assessments in terms of
  - the knowledge framework or domain of each
  - the technical specifications of each
  - the specific assessment objectives against which each is designed
- Content alignment studies will be conducted first for each assessment used as an indicator for reporting preparedness on the grade 12 NAEP scale
  - Full alignment should not be expected
- *Results identify the degree of overlap, which is essential for interpreting statistical studies of score relationships*

## + Key Research Questions for Content Alignment Studies 9

- What is the correspondence in the content domain assessed by NAEP and that of the specified assessment?
- To what extent is the emphasis of NAEP content proportionally equal to that of the other assessment?
- Are there systematic areas of difference between NAEP and the other assessment(s)? Are these differences such that entire content sub-domains are missing or not aligned?

## + 2. Statistical Relationship 10

- The strongest *feasible* form of linking should be used to establish statistical relationships between NAEP and the other assessments
- Equating, which is the strongest form of linking, involves relating scores between two tests built to the same specifications—same content, same difficulty, same reliability—which means that results are interchangeable
  - Because NAEP is a unique assessment with a different function and purpose, equating is not an option

## + Feasible Methods of Linking 11

- Concordance relates scores between two tests built to different specifications. Scores are comparable, but not interchangeable
- Equipercntile methods relate scores on one assessment to those on the other by aligning the distribution of scores such that the average score at the 90<sup>th</sup> percentile (50<sup>th</sup>, 25<sup>th</sup>, 10<sup>th</sup>, and so forth) is aligned to the score on NAEP at each of the target percentiles.

## + Key Research Questions for Statistical Relationship Studies 12

- Can NAEP results be statistically linked to performance on the other assessment of interest?
- Do statistical relationships hold across the full NAEP score scale?
- What scores or score ranges on the NAEP score scale are related to preparedness cut scores on the other assessment?
- How do the statistical analysis results relate to findings of other studies (both within the statistical relationship studies and across other study types in NAEP preparedness research)?

### + 3. Criterion-referenced Judgmental Standard-setting Studies

- Criterion-referenced judgmental studies make use of rigorously developed statements of required knowledge and skills to guide the judgments and increase consistency within and across judges
- These studies would use a criterion-referenced judgmental standard-setting process to identify points on the NAEP scale that indicate preparedness for entry-level general education courses offering transfer credits

### + Key Research Question for Criterion-Referenced Judgmental Standard-Setting Studies

- Based on expert judgments through a rigorously designed standard-setting process, what scores on grade 12 NAEP represent the knowledge, skills, and abilities in reading and mathematics required to demonstrate academic preparedness in the subject area for a particular post-secondary activity?

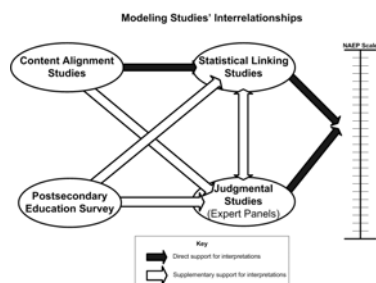
### + 4. National Surveys

- Collect data from a nationally representative sample of two- and four-year postsecondary education institutions
- The survey would collect information about the assessments used for course placement and the cut score(s) on widely used standardized tests for placement into college credit coursework, placement into remedial programs in reading and mathematics, and exemption from placement tests
- The survey results will yield descriptive information related to results from other studies and provide a context for reporting NAEP preparedness research

### + Key Research Questions for the Survey Study

- For each widely used "standardized test," what are the cut scores used to make placement decisions in postsecondary education institutions across the country?
- For each widely used "standardized test," are there systematic patterns and central tendencies in cut scores for institutions or programs of certain types, i.e., level of selectivity, college major, etc.?

### + Relating and Integrating the Results of the Studies



### + Types of Findings from Each Study

- *Content alignment studies* support interpretation of the statistical studies.
- *Statistical relationship studies and criterion-referenced judgmental standard-setting studies* both provide results that produce scores on the NAEP scale.
- *The postsecondary education survey* will provide contextual evidence that can be used to understand how the scores and score ranges identified in the statistical and judgmental studies relate to college course placement cut scores used across the country.

## + Integrating the Findings and Making Sense of the Results

- A logic model is needed *a priori* to ensure that the evidence to support preparedness statements is, in fact, available
- The minimum level that constitutes alignment needs to be specified *a priori* as part of the logic model and as a function of overall procedural validity
- When forming conclusions about procedural validity, the following elements will be included:
  - Consideration of evidence from all types of studies
  - The strength of each type of evidence, e.g., the size of an index of alignment; the representativeness of a survey sample or response; the qualifications of subject matter experts in a judgmental study; and the level of agreement or consistency of judgments
  - The relative strengths of evidence to support inferences about preparedness for different postsecondary education and training endeavors

19

## + The Array of Study Results: Potential Outcomes

- Hypothetical Scenario #1: The Ideal Clustering of NAEP Preparedness Reference Points



20

## + The Array of Study Results: Potential Outcomes

- Hypothetical Scenario #2: NAEP Preparedness Reference Points Cluster with Some Outliers



21

## + The Array of Study Results: Potential Outcomes

- Hypothetical Scenario #3: NAEP Preparedness Reference Points with Minimal Discernible Patterns



22

## + Summary Observations

- Not all comparisons are entirely technical in nature; some skill in combining methods and interpreting results is still required
- Statistical methods, while important, are often not enough in and of themselves to determine relationships of interest
- Validity issues are highlighted in comparisons of this nature, and validity analyses are often more multi-faceted
- The reality is that comparisons of this nature are of interest to assessment users, and methods to make such comparisons need to be developed and refined even if they are technically complex and potentially a bit messy

23