

Inside NAEP: Expanding NAEP Scales to Improve Measurement

The Board has recently heard about efforts that are underway to expand NAEP item pools and scales in order to more accurately measure the knowledge and skills of students across the ability distribution. At the May 2014 Board meeting, NCES provided an update on the Knowledge and Skills Appropriate (KaSA) studies that were intended to more reliably measure the mathematical knowledge and skills of students in Puerto Rico. For the KaSA studies, new items were designed and added to the mathematics item pool that targeted accurate measurement at the lower end of the ability distribution.

The KaSA studies exemplify an approach to targeted item development and administration that NAEP can build upon as the program begins to implement multi-stage adaptive testing designs. Multi-stage testing, or MST, is a test assembly and administration approach in which students are administered a set of items that are targeted to their ability levels. Specifically, based on responses to an initial set of test questions, lower-ability students would receive a set of relatively easier items that measure more accurately at the lower end of the ability distribution and higher-ability students would receive more difficult items that measure more accurately at the higher end. Targeting assessment content in this manner improves the measurement of students' abilities, as it more precisely hones in on what students know and are able to do. In contrast, administering items that are much too easy or too difficult for a student yields little information about that student's knowledge and skills.

The Board raised several questions about the expansion of NAEP scales, including whether administering easier or more difficult items might artificially increase or decrease scores. This presentation will address those questions by providing an overview of how NAEP scales can be expanded to improve measurement without compromising the program's ability to report valid scores that can be compared and interpreted on a common metric. The presentation will explain how a common scale is established despite the fact that students receive different sets of NAEP items that vary in difficulty. Specific topics to be addressed in the presentation include:

- The basic principles of the NAEP assessment design, including the assembly of test forms.
- How the relationship between students' ability and the assessment items, including difficulty, is characterized.
- The concepts of test information and measurement error, which tell us how well the assessment is estimating ability at specific levels of the ability distribution.