

Continuing Commitment to All Students: Ensuring Fair and Unbiased NAEP Assessments and Reporting

NAEP is highly valuable as a measure of student achievement across the United States at the national, state, and select district levels. NAEP is unique in that it does not generate individual student or school level scores and does not impact high-stakes decisions. NAEP is especially useful for providing a consistent measure of student performance in various subjects across the nation, 53 states/jurisdictions, and 26 urban districts, across subgroups. As such, NAEP strives to maintain fair and unbiased assessments.

Both the Governing Board and the National Center for Education Statistics (NCES) have prioritized addressing fairness and equity in the context of NAEP. Equity is highlighted in the NAEP Innovation Agenda, a set of potential priorities developed in 2022 as a joint effort between the Board and NCES. Equity in the NAEP program starts with ensuring a continued commitment to fair and unbiased assessments. At recent meetings, individual Board members and outside speakers have encouraged the Board and NCES to consider the implications of equity for other aspects of the NAEP program, such as the collection of contextual data on participating students and the reporting of results. The purpose of the March 2023 session is to consider what is currently being done to incorporate equity across the full NAEP process and identify any opportunities for improvement. To maximize the Board's discussion time, only key points will be presented before breaking into small groups. Thus, we ask all members to prepare for the session by reviewing this document and the embedded links.

Equity and fairness are incorporated across the full NAEP process. The Board's legislative responsibilities include the development of frameworks to ensure valid and reliable assessments, design of the methodology, developing guidelines for reporting and disseminating results, and improving the form, content, use, and reporting of results ([Public Law 107-279](#)). NCES is responsible for operations including assessment development, sampling, test administration, and analysis and reporting, and the Board policy positions guide these operations.

NAEP Assessment Frameworks

NAEP frameworks provide the blueprint for the content and design of each NAEP assessment. They provide guidance on what should be assessed, how it should be assessed, and how it should be reported. To develop each subject-area framework, the Board works with panels of subject matter experts and practitioners, including many different stakeholder groups. The panels are highly valuable for ensuring the content in the assessments are responsive to changes in standards and curricula across the nation, while also remaining stable for as long as possible to allow the measurement of trends in student performance over time. For more information on the Board's responsibilities related to NAEP framework development, please refer to the [Assessment Framework Development policy](#).

It is important to rely on panels with specific content expertise for ensuring the assessment captures the important skills and knowledge students should know in a particular subject; however, some aspects of the frameworks must remain within the Board's domain and not be determined by a panel, e.g., the overarching intent of NAEP, and how equity is defined in the

context of the NAEP assessments. Similarly, it is within NCES' domain to define how equity is operationalized.

The Board identified a need to be more explicit in what equity means for NAEP based on discussions that arose during the process for the recent updates to the NAEP Reading and Mathematics frameworks. At framework panel meetings for these subjects, panel members discussed equity in various, and sometimes conflicting, definitions and approaches. Ultimately, the Board agreed with the following statements included in the 2026 NAEP Reading Framework to define equity for NAEP assessments:

“The Governing Board has a continuing commitment to equity in our educational systems. It advances this goal by designing assessments that are inclusive and accessible for the full diversity of students who are administered the NAEP Assessments. The assessments will align with the recent standards in large-scale assessment by continuing to strive to minimize test bias to the maximum extent possible (American Educational Research Association, American Psychological Association, and National Council of Measurement in Education, 2014; International Test Commission, 2019; IRA/NCTE Joint Task Force on Assessment, 2010). Finally, the Assessment will gather data that afford opportunities to examine malleable contextual variables that may lead to greater understanding of differential student achievement.” (pg. 4)

In addition, the 2026 NAEP Reading Framework indicates what this means for the Reading assessment specifically. For example:

“Range of Texts and Tasks Represented:

Selection of a diverse range of texts and tasks representing different student identities, interests, knowledge, and other backgrounds helps to ensure equity across diverse subpopulations of test takers. Such broad sampling facilitates equitable test items and scales.” (pg. 41)

The complete 2026 NAEP Reading Framework is available [here](#).

As part of the March Quarterly Board meeting, members will be asked to reflect upon the utility of the overarching statement included within the 2026 NAEP Reading Framework, and consider whether these words should be incorporated into the 2028 NAEP Science Framework and future NAEP frameworks. They will also have the opportunity to consider whether additional actions should be considered to ensure equity is incorporated to the framework process as intended by the Board.

NAEP Processes and Operations

Bias and sensitivity reviews are a key component of the NAEP [item development process](#). NCES purposely develops more than twice the number of items needed and only those that pass item reviews by experts and teachers and pilot tests with students make it onto the assessments. The bias and sensitivity reviews help ensure items do not unintentionally disadvantage student groups for reasons that are irrelevant to the skill or knowledge to be measured. During pilot and operational administrations, NCES conducts scoring reliability checks, item analyses, and [differential item functioning \(DIF\)](#) to ensure fair and reliable measures of performance in the

subject of the assessment. DIF reflects the differential probability of doing well on the item depending on group membership, even after controlling for overall performance.

The NAEP [sampling procedures](#) intend to ensure NAEP results are a fair and unbiased representation of student achievement for all reported subgroups. The samples of schools and students for each assessment are selected to ensure proper representation by Census region, sex, race/ethnicity, degree of urbanization of school location, parent education (for grades 8 and 12) and participation in the National School Lunch Program. When assessments are to report out at the state and/or TUDA level, procedures are designed to produce representative samples at the state and/or TUDA levels and thus require a much larger overall sample than when only national results are reported. Because the schools and students who participate in the assessment represent only a portion of the larger population of interest, the assessment results are weighted to make appropriate inferences about the populations from the student, school, and district samples. Sampling weights are adjusted to account for the disproportionate representation of some groups in the selected sample. This assessment design includes oversampling of schools with high concentrations of students from certain racial/ethnic groups and the lower sampling rates of students who attend very small schools.

Because NAEP is intended to provide an accurate representation of all students, it is important to be as inclusive of all student populations as possible, including English learners and students with disabilities. To make this possible, NAEP includes a number of universal design elements available to all students, and accommodations for students with disabilities and English learners. The accommodations are generally similar to those allowed on state assessments, though they may differ as required by the NAEP frameworks. A summary of the NAEP accommodations is available [here](#).

Finally, fairness and equity are considered in NAEP reporting. The NAEP Report Card currently presents results for various student groups (i.e., select race/ethnic groups, sex, NSLP eligibility, various school characteristic variables, disability status, and English learner status). In addition, NAEP reports score differences by demographic characteristics of students as well as contextual data, such as access to technology. The majority of contextual items appear across years on the contextual questionnaires, so NAEP can report trends for these score differences, illustrating how gaps widen or narrow over time. For example, contextual variables asked administrators, teachers, and students about pandemic experiences, so context for lower-performing students could be compared to higher-performing students. NAEP routinely collects data from [student, teacher, and school questionnaires](#) that can be useful for exploring educational equity. Report Card readers can select two groups to discern differences in performance; NAEP data tool users can refine analyses further. See the [2022 NAEP Report Card for Reading, grade 4](#) for an example of how NAEP results are presented to examine student groups and make comparisons between groups. NAEP reporting is largely driven by the reporting guidance included in the NAEP frameworks.

Aside from the initial release of data, NCES publishes several studies that relate various student and environmental factors to achievement differences. The School Composition and the Black-White Achievement Gap report revealed that performance differences between Black and White

students were larger in schools with a majority of Black students. Since 2003, NCES has compared state-level proficiency with NAEP standards and spotlighted how proficiency widely varies throughout the country. NCES is in the process of deploying an achievement gap regression tool that breaks down score gaps between racial/ethnic groups by gender, socioeconomic status, and students' affective dispositions.

NAEP as a Tool for External Researchers

The NAEP Report Cards are not the only source of information available from NAEP data. NAEP provides achievement and contextual data to external researchers and encourages its use for conducting education research. The [Resources for Researchers](#) website provides information on data and tools available for secondary analysis. Some of the current Board members are regular users of NAEP data for secondary analyses. The Board should consider whether encouraging secondary analyses of NAEP data for exploring equity and fairness in education is something they would like to pursue.

Board 2021 Equity Panel and Discussions

The Board included a session on equity during the March 2021 Quarterly Board Meeting that may be useful to consider going into the March 2023 discussion. A plenary session was convened titled: [*Reflections on Recommendations from the National Academies' Committee on Developing Indicators of Educational Equity*](#). The purpose was for the Board to consider whether and how NAEP could play a role in national efforts to create more equitable outcomes for students through the use of data. The plenary session was moderated by Tonya Matthews (former Board member), and included presentations by Rucker Johnson (Chancellor's Professor of Public Policy, University of California, Berkeley), Gerunda Hughes (NAEP Validity Studies member and Professor Emerita at Howard University), and Christopher Edley (Distinguished Professor at University of California, Berkeley Law School and Chair of the National Academies' Committee on Developing Indicators of Educational Equity).

Key points panelists made relevant to NAEP included:

- Hughes suggested collecting and/or reporting student grouping data not specifically listed in the legislation, such as additional variables to address (a) societal, (b) socioeconomic, (c) cultural, (d) familial, (e) programmatic, (f) staffing, (g) instructional, (h) linguistic, and (i) assessment inequities in the educational system.
- Hughes highlighted where an equity lens can be applied to NAEP, namely in: (a) sampling, (b) assessment design and development, (c) administration, (d) accommodations, (e) data analysis and reporting, (f) reporting and interpretations, and (g) use of results. She concluded with the five "E's" of equitable educational assessment: empathy, engagement, equity, evaluation, and equality.
- Edley asked the Board to consider how NAEP and related data should be used to provide context and how NAEP-related data should be used within an Equity Indicators System, that would incorporate existing data to help measure and compare schools across jurisdictions. NAEP could be included as an indicator of test performance disparities.