Ensuring Trend Integrity in 2026 NAEP Reading and Mathematics: Assessing the Impact of Device, Framework, and Scoring Transitions

Goal

The purpose of this session is to provide the National Assessment Governing Board with an overview of the three major transitions NCES implemented during the administration of the 2026 NAEP reading and mathematics assessments: (1) the addition of approved school devices in assessment delivery; (2) updated assessment content to reflect framework changes; and (3) the implementation of automated scoring in reading. The session will highlight the evidence supporting score comparability and trend reporting.

Overview

Enis Dogan of the National Center for Education Statistics (NCES) will present an overview of the three major transitions implemented in the 2026 NAEP reading and mathematics assessments. The presentation will summarize the evidence supporting the comparability and validity of 2026 results, including findings from the device bridge study, sensitivity analyses evaluating the impact of framework changes on trend reporting, and reliability and fairness analyses supporting automated scoring, with a focus on how these efforts help maintain the integrity and interpretability of NAEP trend measures.

Background

The 2026 NAEP reading and mathematics assessments mark a significant modernization milestone, introducing three major operational and methodological changes simultaneously: (1) the transition from exclusive use of NAEP-provided devices to a mixed-device administration model that includes eligible school-owned devices, (2) implementation of updated reading and mathematics frameworks, and (3) the first operational use of automated scoring for constructed-response items in reading. Each of these changes has the potential to affect score interpretation and trend reporting, making a comprehensive series of evaluations and quality assurance steps essential.

To support the device transition, NCES implemented a nationally representative bridge study in which schools were randomly assigned to either a legacy administration condition using NAEP-provided devices or a mixed-device condition in which students in eligible schools use their own devices while ineligible schools continued to use NAEP-provided devices. The bridge study was designed to evaluate score comparability across administration modes and to provide the technical foundation for maintaining trend reporting.

The 2026 assessments also represent the first operational administration of updated reading and mathematics frameworks. The revised frameworks introduce new item types and changes in content and subscale weights, and, in reading, an expansion from two to three reporting subscales. To address questions about trend comparability under the new frameworks, NCES is conducting a series of sensitivity analyses to examine the impact of framework changes on overall results and subgroup performance.

Finally, 2026 marks the first operational use of automated scoring for reading constructed-response items. This transition builds on years of research, pilot testing, and special studies. Automated scores are being used alongside rigorous quality-control procedures, including human-scored validation samples and ongoing monitoring of scoring performance. Reliability and fairness analyses evaluate agreement between machine and human scoring overall and across subgroups.