



As Nation's Report Card Resumes for First Time Since Pandemic, Federal Testing Chief Admits She's 'A Little Nervous' about Results

March 28, 2022

LINDA JACOBSON

Almost 600,000 U.S. fourth- and eighth-graders are currently taking national reading and math tests for the first time since the pandemic began.

The prospect makes the federal official in charge of measuring student progress a bit anxious.

“The likelihood that the scores would be anything but down is pretty small,” said Peggy Carr, commissioner of the National Center for Education Statistics.

While performance among the lowest-scoring fourth- and eighth-graders on the National Assessment of Educational Progress was falling well before the pandemic, Carr predicted, “It’s more than likely we’re going to see the bottom drop even more.”

Known as the Nation’s Report Card and a chief “barometer” of educational achievement in the U.S., the congressionally-mandated NAEP is the only assessment with results broken down by gender, race and socioeconomic status that can be compared across all 50 states. As such, it is a major gauge of achievement gaps that are likely to have grown larger since COVID’s arrival. Carr’s predecessor, James “Lynn” Woodworth, described this year’s administration as “the most important NAEP assessment that’s ever been done in the history of this country.” In a wide-ranging interview with The 74, Carr fleshed out why, explaining that the pandemic had added layers of “noise” that could make results harder to interpret.

‘True change’

Even the most superficial alteration in a student’s testing experience can throw off their performance. In the mid-1980s, researchers determined that a change in the color of the ink on the test booklets contributed to an otherwise unexplained drop in reading performance for 9-

and 17-year-olds. In 2002, the mistake was accidentally repeated with a random sample of students, and again, scores dropped.

But the pandemic has exploded the universe of possible variables: The sample of test-takers includes masked and unmasked students, as well as smaller groups. Social distancing and other changes in the environment could also affect student performance.

“It makes me a little nervous about what we’re going to see, and how I’m going to be able to separate out what is noise and what is true change in students’ performance,” she said.

At the same time, collecting those results has been far from easy. From staff quitting due to illness to schools rescheduling because of students in quarantine, this round of testing is unlike anything the center has faced in the past.

“I’m getting notices every day that people are quitting or people have ... caught COVID in the schools,” Carr said.

In December, there were 3,560 NAEP staff members in the field. More than 850 have quit, with over half of those leaving in December and January as Omicron started to spread, according to NCES.

Because of COVID’s lingering interference, Carr said she was pleasantly surprised schools haven’t pulled out of the assessment. Only one district, Fresno Unified in California, opted not to participate in the Trial Urban District Assessment, which provides results for more than two dozen districts nationwide.

Nonetheless, state and district chiefs have already expressed concern about whether Carr can guarantee the validity of the results.

“They said, ‘Peggy, I’ve got 900 vacancies. I have people who normally teach art teaching some academic subject,’” she said.

To the doubters, she emphasized that this year’s tests include the same items used in 2019, which will further give the public a “solid trend line” through the pandemic years, she said.

NAEP, she stressed, is “still the standard by which other large-scale assessments judge themselves, and even in the context of COVID that has not changed.”

But because of the impact of the pandemic, it might seem as if this year’s results are setting a new “baseline,” Carr said. A baseline, which technically refers to official changes in the test, is the starting point researchers and policymakers use to track student performance over time.

“It’s a new day in many ways,” Carr said. “How tests are being administered, how students are being taught and how they learn in schools today is a little different than it was before COVID.”

Despite those challenges, NCES’s responsibility is to maintain the public’s trust in NAEP as an accurate measure, said Andrew Ho, an education professor at Harvard University and a former member of the National Assessment Governing Board, which sets policy for NAEP.

Carr must help parents and educators understand how the pandemic has affected fourth- and eighth-graders’ math and reading achievement, he said. “It’s not a new baseline if we do our job right,” he said. “It is a decline.”

He and Carr said the urban district results will be especially valuable when viewed against the backdrop of school closures at those sites. “There are always policy differences; they just haven’t been so confounded with historic health issues,” Ho said, adding that it’s inevitable the results will become fodder for political arguments over how leaders responded to the pandemic. “Everyone likes to attach a policy story to NAEP results.”

In addition to the reading and math tests, which were delayed a year because of school closures, NCES is testing eighth graders in civics and history and 9-year-olds as part of its long-term trend study. Nine-year-olds were also tested in 2019, which will allow NCES to provide pre- and post-pandemic results. Data from three years ago showed stagnant performance in both reading and math, except for girls, whose math scores dropped five points. Now researchers will be able to see how students in that age group, who were in first or second grade when schools shifted to remote learning, are performing. Next year, 13-year-olds will be assessed.

A [bipartisan bill](#) in the Senate, introduced this month, proposes that NCES add a new component to measure the long-term impact of COVID on a representative sample of students.

It could be much harder, however, to see how U.S. students fared during the pandemic compared to their peers in other countries. While states and districts generally participate in all non-mandated NAEP tests, such as those in history, economics and technology, Carr struggles to get an adequate sample — 350 schools — for the Program for International Student Assessment and other global comparisons.

School leaders are bombarded with requests to participate in surveys and an optional assessment can feel like one more burden. When Betsy DeVos was education secretary, Carr asked her to recruit schools for the international assessment. [Michael Casserly, who led the Council of the Great City Schools and pushed for the urban assessment results, also helped.](#)

[“When Betsy DeVos was here, we had her calling schools, and we got Mike Casserly, who’s a good friend of mine, calling schools and we barely made it,” Carr said. “It’s a hard sell. I’ve got to figure out another way to develop a relationship with the stakeholders on the ground and make it worth their while to participate.”](#)

Cloud-based tests and AI scoring

As Carr prepares to analyze this year’s NAEP data, she’s also overseeing a modernization of the program, which has been “fast-tracked” by the pandemic, she wrote in a recent [blog post](#), co-authored with Lesley Muldoon, executive director of the assessment’s governing board.

Future tests will be cloud-based and downloaded to districts’ own devices. And beginning in 2024, NCES will no longer hire 3,500 to 4,000 administrators to deliver devices with the assessment to schools. That model, which prevented the center from conducting mandated tests in 2021, typically costs about \$62 million.

While some field staff will still be on site, using local administrators could save \$22 million, according to [a report](#) released last week.

Also in 2024, NCES will begin using artificial intelligence to score students' essays. In January, the center announced four winners of a competition who showed AI can score an essay with 88% accuracy compared to trained individuals, Carr said.

“It may be good enough with a little bit of tweaking,” she said, adding that the center will still have human scorers in 2024 to remain “scientifically defensible.”

The center will also continue running its monthly School Pulse Panel — the result of an early Biden administration executive order to produce data on the impact of the pandemic. The survey tracks the percentages of students in in-person, hybrid and remote learning and has expanded to add questions on staff vacancies, quarantines and mental health.

The project has pushed the center toward a quicker turnaround — something the governing board and state and local leaders would like to see with NAEP as well.

“If I don't have to put together the full-blown report card with all the bells and whistles, maybe I can get it out faster,” Carr said. “But I'm not going to cut short the statistical analysis that I need to make sure we can stand behind the data. I'll put asterisks on it. I'll caveat it, and then ... whatever it says, I'm going to report it.”