Center for American Progress

GETTY IMAGES/BEN HASTY

FUTURE OF TESTING IN EDUCATION

The Way Forward for State Standardized Tests

By Laura Jimenez and Ulrich Boser September 2021

Contents

Introduction and summary

> This series is about the future of testing in America's schools. Part one of the series presents a theory of action that assessments should play in schools. Part two reviews advancements in technology, with a focus on artificial intelligence that can powerfully drive learning in real time. And the third part—this report—looks at assessment designs that can improve large-scale standardized tests.

Federal law requires all public school students in grades three to eight to take an annual assessment in reading and math at the end of the year and requires students to take an assessment once during high school. The goal of this assessment is to measure the extent to which all students are meeting the state's academic standards. These standards must align with the knowledge and skills in reading and math that students need to succeed in first-year college reading and math courses. Ensuring all students are held to rigorous standards is a key goal of equity in education.

Yet many question the value of yearly standardized testing in schools since the opportunity to receive a high-quality education and graduate high school adequately prepared for college-level academics is still wholly inequitable. Students who are Black, Indigenous, and Hispanic graduate high school at lower rates than their white peers, and they require catch-up coursework in college more often.[1] What is more, the costs and time associated with assessments, delayed results, and failure of tests alone to improve students' academic results leave many to wonder if they are worth the effort at best, and at worst, if they harm students and punish teachers and schools.[2]

Still, there are ways to design an assessment to reduce the amount of time it takes to administer, ensure that it collects information about students throughout the year, or base the test on performing tasks. This report describes the advancements in testing technology that make such assessments possible, and it concludes with recommended changes in federal testing policy to make the use of these designs

effective. Apart from greater investments in research and development of new assessment designs, the federal government should also loosen regulations on the assessment pilot included in the recent reauthorization of the Elementary and Secondary Education Act.[3]

The Center for American Progress' companion report in this series, "Future of Testing in Education: Effective and Equitable Assessment Systems,"[4] separates fact from fiction regarding the criticisms against standardized testing. It also underscores CAP's theory that, when well designed, tests can provide insights into what students know and do not know, allowing education stakeholders to drive student learning forward. This information is critical to teachers in the design of daily instruction, as well as to school administrators and policymakers who decide on and fund supports when students need them.

The yearly state standardized assessment alone cannot ensure a high-quality education for every child. But without it, educators do not know their progress toward meeting that goal. Despite the valid concerns that remain about standardized assessments and their role in education, there are a number of upsides and value to standardized testing. For instance, it is the only common measure of grade-level academic standards for all public school students. As such, it is one measure to determine if students are on track for college or career readiness when they graduate high school. The Every Student Succeeds Act (ESSA)—a federal law requiring all students to be held to the same high standards—is one way to help ensure an equitable opportunity to a high-quality education.

New ways to test students in the spring

Advances in technology—and even some decades-old assessment designs—can reduce testing time and improve the quality of the standardized tests themselves by addressing the drawbacks discussed in CAP's issue brief in this series, "Future of Testing in Education: Artificial Intelligence."[5]

Long testing times, cultural bias, and limited usefulness to teachers are just some of the criticisms against today's state standardized assessment. However, advances in technology can alleviate some of these concerns. Some tests, for example, can use sampling testing techniques to reduce testing time. Overall, there are three new ways to assess students discussed in this report.

Matrix sampling cuts testing time

Today's state standardized tests require students to sit for eight to nine hours total, in two-hour segments. Matrix sampling—which provides individual students with a representative sample of assessment questions—reduces testing time. Rather than all students taking all test items, one approach to matrix sampling involves selecting a limited set of test questions in a way that allows evaluators to estimate results for the entire test. In other words, no individual student takes the entire exam. In addition to decreasing testing time, the results produce group-level information rather than individual test scores. This is the approach used by the National Assessment of Educational Progress, or NAEP—a test given to students since the 1970s that takes about 90 minutes to complete.[6]

Using another type of matrix sampling, test developers select specific test questions that will predict performance on the entire test. All students take these selected items.[7]

State education departments already use matrix sampling in current state tests, often to pilot new questions and see how students interact with them. For widescale use, a matrix sampling design would provide similar information about student performance at the student group level as current state tests do.

Furthermore, experts advise that with additional innovative techniques, matrix sampling could measure growth for individual students by giving them enough test questions in common and applying additional statistical analysis to the questions.[8] Other statistical techniques could also produce individual student reports. These methods could also support results comparison between students and from one year to the next, allowing policymakers and administrators to identify trends.

In sum, a matrix sampling assessment design could give enough of the benefits of a full-length assessment without the significant drawbacks that long testing time has on students. However, without further innovation on current matrix sampling designs, the lack of individual test scores poses a significant barrier. The authors discuss the current policy barriers to using a matrix sampling design in a later section of this report.

Through-year assessments would eliminate a summative assessment in the spring

Through-year tests have been piloted in some states involved in the assessment pilot created by ESSA. The concept of this design is simple: Develop tests that are administered throughout the year, and aggregate the results of some questions into a summative score.

Two states involved in this pilot are using through-year assessments but have different theories of action for how the tests should support student learning. These differences lead to distinct test designs.[9] Louisiana, for example, wants to eliminate the gap between what students learn and what they are tested on. To do so, Louisiana bases its test on optional statewide curricula that are aligned with the state's standards. Experts in the content of Louisiana's curriculum create the tests, with the involvement of teachers. The state designed this approach to allow teachers to go deep on the standards as well as other skills such as critical thinking.

Georgia's approach focuses on developing tests that meet students where they are, regardless of their grade level. When students are behind, the state standardized assessment tests knowledge and skills that these students cannot perform. Thus, the results have a limited ability to inform teachers and other education stakeholders of the path to help students catch up. For students who are ahead, testing only grade-level material provides a disincentive to push students further.

The idea in both states is that the tests will return data throughout the year that teachers can use to shape instruction while kids are still in the classroom, rather than getting scores after the school year is over. When provided with this information, teachers can intervene earlier and use data to meaningfully close gaps for students, work toward the goal of achieving grade-level proficiency throughout the year, and adjust their instruction if the gaps are not closing as they hoped. The potential to test information closer to students' real-time learning and intervene sooner if students are not meeting critical benchmarks are significant upsides to through-year test designs.

Both states are very early in their development of these new tests, and it is not yet known how well these methods meet states' goals for their new tests. For example, they are still working to develop a single, summative student score at the end of the year as federal law requires.

Currently, through-year assessments face three challenges. First, it is too early to know if these tests can still innovate despite the design constraints required by law, as the pilots are still in their early phases. Second, it is also too early to know if through-year tests help ease the anxiety around testing or if they amplify it, because instead of just one test that will be used for accountability, there will be three. Test designers hope that by giving more frequent tests that are more tightly connected to what students learn and are closer to students' academic level, this practice, over time, may reduce test anxiety. Third, through-year tests work best in states where students are learning the same things at the same time. However, decisions about what students learn and when they learn it happen at the district level, not the state level. The only state that currently has symmetry of curriculum and instructional materials across some school districts is Louisiana.[10]

Performance-based assessments use hands-on tasks where students can demonstrate their knowledge and ability

Researchers consider performance tasks to be authentic measures of standards because the tasks are extended performances of student work, showing multiple stages of the thought process and how students arrived at the solution.[11] Students do not just learn the specific academic content; they develop a range of skills in the process of completing complex tasks, such as presenting and defending their work, leading or participating in individual or group projects, and performing other multifaceted tasks.

There are two approaches to designing tests based on performance tasks: (1) standardize the tasks, meaning all students perform the same tasks, or (2) standardize the scoring rubric, meaning students perform different tasks that are scored using a common scoring tool.

New Hampshire uses the first approach in its program called Performance Assessment of Competency Education, or PACE.[12] The New York Consortium on Performance Based Assessment uses the second approach:[13] Participating educators come together a few times a year to develop the scoring rubric, which also functions as professional development for teachers on how to use standards-based grading. To do this, educators review a sample student work product, such as an essay, to produce a common scoring rubric for all participating schools.[14]

Proponents argue that performance-based tests are more motivating to students and allow for more holistic review of student work.[15] As a result, the performance-based assessments can analyze students' skills at a deeper level and are better suited to measuring crosscutting skills such as critical thinking and teamwork.[16] Thus, performance-based tests can be tools for learning as well as measures of learning.

But despite the advantages of allowing students creativity in demonstrating their work, performance tasks suffer from some drawbacks. Current performance-based assessments may not serve as a complete replacement for summative tests. The complexity of the performance tasks themselves and the sheer number of academic standards may make it difficult for the tasks alone to measure the full range of standards without significant effort to group the standards. As a result, New Hampshire uses a combination of performance tasks, local tests, and the statewide test to measure the full range of standards.[17]

Performance tasks also have challenges with scalability. For example, a state would need to create a process to norm the scoring rubrics statewide, based on a large set of sample work; it would then need to create a process for how educators become certified graders for their school. If the state chose to standardize the tasks as well, they would repeat a similar process. There are also challenges in hand-scoring and observing any student demonstrations occurring in real time, such as an essay defense—not just in the time it takes to complete, but also to ensure that no bias occurs.

The performance tasks also have implications for how teachers teach and how schools are organized to deliver instruction. That is, schools must redesign their approach to teaching in order to help students build the prerequisite skills and allow for hands-on learning. For example, schools would need to invest significant time and resources in creating learning experiences designed to improve skills such as critical thinking, time management, collaboration, and communication, in addition to teaching content based on the state's academic standards.

Some of these factors explain the slow-growth approach New Hampshire intentionally takes in its PACE program. In its first 10 years, four districts out of 167 total joined the PACE effort.[18] During a 2016 visit of New Hampshire schools involved in the program, then-state assessment director Paul Leather said the program would be optional for districts given the complexities involved in implementing the program successfully.[19]

Finally, it is unclear how much more effective performance-based assessments are at improving student outcomes. Studies of New Hampshire's PACE program show small observed differences and less improvement among low-income students, students with disabilities, and male students.[20]

How the law can allow states to test and use these innovations in testing

Technical requirements in existing laws and regulations could prevent states from trying some of the innovative designs highlighted above. However, this report does not advocate for wholesale waivers of these requirements. It offers points of consideration for the U.S. Department of Education and states as they develop policies that allow for innovation.

This section discusses the following standardized testing requirements:

- Validity, reliability, and comparability
- Grade-level measurement
- Individual score reports

Defining validity, reliability, and comparability, and why they matter

Validity refers to how accurately and fully a test measures the skills it intends to evaluate.[21] For example, if an algebra test includes some geometry questions, the test would not be a valid measure of algebra. Reliability refers to the consistency of the test scores across different testing sessions, different editions of the test, and different people scoring the exam.[22] Reliability informs how consistently the test measures the knowledge and skills it is intended to measure. Comparability allows for comparing of test scores, even if students took the test at different times, in different places, and under different conditions.[23] For example, test developers will design a test that may be given via computer or paper and pencil to account for these differences so results can be compared.[24]

These three technical qualities apply to the ESSA pilot program for state assessments as well. Piloted designs must meet the highest standards for validity, reliability, and comparability, as state tests do. Lawmakers did this not only

because the piloted tests were required across states to fulfill the yearly student and school performance requirements in federal law, but also because they are fundamental to test quality, fairness, and equity.[25]

While these technical requirements help ensure a high-quality test, they also constrict states' abilities to try new approaches to testing. As a result, the requirements confine the pilots to look and behave like today's standardized tests.[26]

Where can policymakers be flexible on test reliability and comparability?

State tests should always be valid measures of student knowledge and skills of interest. But state tests can be reliable and comparable enough while still serving as high-quality measures of what students know and can do. For instance, to provide flexibility on how reliable scores must be from one student to the next, policymakers could allow scores to be less consistent as long as the test still measures the same skills. Test developers would call this maintaining comparability of the standards.

Test makers can compare scores even if they are not 100 percent equivalent. In fact, test developers compare different tests all the time through a concordance table, which allows the scores from different kinds of tests to be equated. Universities use these, for example, when they accept scores from both the ACT and the SAT as part of freshman application requirements; a concordance table can indicate what scores on each test are roughly equivalent.[27] The same is true for states when they change from one yearly summative test to another. They create concordance tables that compare the same test construct rather than the results of different tests altogether. While these tables do not perfectly convert scores from one test to another, they make the closest comparison possible. Although this is not a complete solution to achieving 100 percent comparability of scores, as state tests currently have, this offers some degree of comparison. Policymakers could set a minimum threshold for comparability of scores that is less than 100 percent.

How do tests' technical requirements affect the future of testing?

Each of the new ways to test students highlighted in this report may be constrained by the requirements discussed above. As a result, the U.S. Department of Education may need to grant states some flexibility to try these designs.

For example, states would need a waiver from ESSA section 1111(b)(2)(B)(x)—which requires individual student score reports—in order to try a matrix design. Although matrix design assessments do not typically provide individual scores, test developers may be able to apply additional statistical analysis to approximate student results for the full range of standards, allowing them to provide insights into individual students' performance.

Additionally, states could include information in an individual student report about the specific part of the domain of standards, or the specific knowledge and skills, on which the student was tested. For example, administrators can let a parent know their child will be randomly assigned an assessment at the end of the year that deals with one of these areas or domains. The parents would receive a report about how students do in the area they are tested in and information about how the school performed as a whole. However, this approach requires flexibility both in reporting and potentially in the provision that all students get the same test.

Recommendations: The path forward for states and ways the federal government can help

Regarding state assessments, states can choose between two divergent paths. First, they could reduce the footprint of their annual assessment by making it shorter and leaving more time for instruction through a matrix design. Proponents of this approach also say that integrating summative test questions into tests given throughout the year, such as with through-year exams, reduces the footprint by making the testing experience more normalized, especially since the assessment tests the content students have just learned. Alternatively, they could increase its footprint through a performance-based design, but in a way that would provide a broader array of data about what students know and can do.

Congress and the U.S. Department of Education can put the following policies into place to allow for more innovation within state summative assessments.

Congress should:

- **Revise the innovative assessment pilot policy.** The pilot requires states to use their piloted designs statewide within five years. Rather than instituting an arbitrary deadline, Congress should allow states to determine the length of time for their pilot. It should also recognize that not all pilots will be successful and allow states to abandon their pilots if they are not working well. Effective research in testing requires money, so Congress should also fund this pilot. Finally, pilot results should inform the next authorization of ESSA. In addition to using the pilot evaluation for this purpose, Congress should hold a hearing with pilot participants to gather their perspectives.

- **Give additional funding to states for testing and related research and development to support cutting-edge technology.** Congress can do so by increasing funding for three programs. First is the Competitive Grant for State Assessments (CGSA) program, which currently provides about $8 million every year to states wanting to develop additional assessments. In order to reach the full potential of these new technologies to improve the teaching and learning

process, the CGSA program needs far more funding.[28] Second is the Grants for State Assessments and Related Activities program, which helps fund states' yearly assessment systems and the activities needed to carry it out; any increase should not be reallocated to the competitive grant program as current law does.[29] Third is the Small Business Innovation Research program, which provides up to $1.1 million in awards to develop education-related learning technologies.[30] Congress could also orient this program to have more of an assessment angle rather than one focused on general education technology. As technologies evolve, there could be ways to better use them to make assessments more effective. These innovations could challenge the norms for how today's test developers and researchers are trained. Congress should also fund doctoral student training for future psychometricians who will build the assessments of tomorrow, emphasizing their education in new and emerging technologies as one way to continue advancing technological innovations.

The U.S. Department of Education should:

- **Make three revisions regarding applications to participate in the assessment pilot.** It should ask states to:
  - » Design their pilot and respond to the existing technical requirements for assessments according to their theory of action. For example, if proposed designs have less strict requirements for score comparability, states should describe how this improves student learning, such as through higher-quality interactions between students and teachers based on student work.
  - » Develop their application in collaboration with testing experts and the pilot site communities. This helps piloted designs reflect what local communities want and need.
  - » Propose a realistic timeline for the pilot and its analysis, revisions, and rollout statewide.

- **Improve the evaluation of the assessment pilot.** Current plans for the evaluation include reviewing existing documents from piloting states and asking about participant experiences through a survey.[31] It is not clear that these methods will provide the most useful information about the technical merits and challenges of the tested designs or the policy and design choices and their impacts. The final regulations for this evaluation should include a much more robust plan for gathering this kind of information, evaluating the extent to which states realized their theories of action and their intended outcomes.

- **Create the following priorities for the CGSA program to facilitate parts of new test designs—not just entirely new tests—that any state or district could use.** These grants could develop:
  - A bank of test questions suitable for diagnostic tests, as well as formative and summative exams.
  - A bank of test questions that align with the highest quality textbooks and curriculum, as rated by experts. The test questions could also take cultural relevancy and cultural representation into account.
  - Large datasets to be used for machine learning applications, such as automated essay scoring. For example, the Automated Student Assessment Prize awarded scientists funding to train machines to score essays similarly to how human experts would.[32] Datasets like these are key to the training of new tools and models.
  - Easier to understand and more actionable student and school score reports.[33] About half of states still do not report federally required data such as levels of teacher experience and student college attendance rates.[34] Innovation in this area could make it easier for states to produce consumable data.
  - Training for educators and education leaders in states, districts, and schools to use test results in ways that are more useful to student-teacher interactions and ensure that students get the support they need.
  - More tools and funding for formative assessments and the ways that formative assessment can be rolled up into summative assessments.

- **Revisit the assessment peer review process to ensure it provides flexibility while maintaining a high bar—for example, by evaluating grade-level mastery based on as few items as necessary.** Regulations could also allow states to show comparability through concordance tables, allowing for greater differences in scores from one student to the next.

- **Provide technical assistance to states.** Specifically, the Education Department can guide states on how to write requests for proposals in ways that result in the creation of tests that are more innovative.

Conclusion

Advancements in assessment technology can make state standardized tests more streamlined and capable of providing better information about what students know and can do. If states are encouraged and funded to take the new approaches described in this report, they can increase the value that testing data provide to educators, parents, and policymakers.

About the authors

Laura Jimenez is the director of standards and accountability on the K-12 Education team at the Center for American Progress.

Ulrich Boser is a nonresident senior fellow at the Center.

Acknowledgments

The authors would like to thank the following people for their advice in writing this report:

- Abby Javurek, Northwest Evaluation Association
- Alina Von Davier, Duolingo
- Ashley Eden, New Meridian
- Bethany Little, Education Counsel
- Edward Metz, Institute of Education Sciences
- Elda Garcia, National Association of Testing Professionals
- Jack Buckley, Robox and the American Institutes for Research
- James Pellegrino, University of Illinois at Chicago
- John Whitmer, Institute of Education Sciences
- Krasimir Staykov, Student Voice
- Kristopher John, New Meridian
- Laura Slover, Centerpoint Education Solutions
- Margaret Hor, Centerpoint Education Solutions
- Mark DeLoura, Games and Learning Inc.
- Mark Jutabha, WestEd
- Michael Rothman, Independent consultant formerly of Eskolta School Research Design
- Michael Watson, New Classrooms
- Mohan Sivaloganathan, Our Turn
- Neil Heffernan, Worcester Polytechnic Institute
- Osonde Osoba, RAND Corp.
- Roxanne Garza, UnidosUS
- Sandi Jacobs, Sean Worley and Scott Palmer of Education Counsel
- Terra Wallin, The Education Trust
- Tim Langan, National Parents Union
- Vivett Dukes, National Parents Union

Endnotes

1 National Center for Education Statistics, "Public High School Graduation Rates," available at https://nces.ed.gov/programs/coe/indicator/coi (last accessed July 2021); National Center for Education Statistics, "Remedial Coursetaking at Public 2- and 4-year Institutions: Scope, Experience and Outcomes" (Washington: U.S. Department of Education, 2016), available at https://nces.ed.gov/pubs2016/2016405.pdf.

2 Kirwan Institute for the Study of Race and Ethnicity, "Standardized Testing and Stereotype Threat," March 12, 2013, available at https://kirwaninstitute.osu.edu/article/standardized-testing-and-stereotype-threat; Richard J. Shavelson and others, "Problems with the use of student test scores to evaluate teachers" (Washington: Economic Policy Institute, 2010), available at https://www.epi.org/publication/bp278/; Christian Barnard, "To Ensure Equitable Funding for Low-Income Students, Fixing Title I Isn't Good Enough — It Needs to Be Rebuilt From Scratch," The 74, June 18, 2019, available at https://www.the74million.org/article/barnard-to-ensure-equitable-funding-for-low-income-students-fixing-title-i-isnt-good-enough-it-needs-to-be-rebuilt-from-scratch/.

3 Every Student Succeeds Act, Public Law 114-95, 114th Cong., 1st sess., December 10, 2015, available at https://www.govinfo.gov/content/pkg/BILLS-114s1177enr/pdf/BILLS-114s1177enr.pdf.

4 Laura Jimenez and Jamil Modaffari, "Future of Testing in Education: Effective and Equitable Assessment Systems" (Washington: Center for American Progress, 2021), available at https://www.americanprogress.org/?p=502607.

5 Laura Jimenez and Ulrich Boser, "Future of Testing in Education: Artificial Intelligence" (Washington: Center for American progress, 2021), available at https://www.americanprogress.org/?p=502663.

6 Emmanuel Sikali and Cadelle Hemphill, "Focus on NAEP: National Assessment of Educational Progress Sampling," Nation's Report Card, available at https://www.nationsreportcard.gov/focus_on_naep/files/sampling_infographic.pdf (last accessed July 2021).

7 Ibid.; Edward Roeber, "What does it mean to use matrix sampling in student assessment?", Michigan Assessment Consortium, available at http://michiganassessmentconsortium.org/wp-content/uploads/ThinkPoint_MatrixSampling3.pdf.pdf (last accessed July 2021).

8 Summary of interview with Jack Buckley, head of assessment and learning, Roblox, interview via video conference, October 7, 2020, on file with author; Summary of interview with James Pellegrino, Liberal Arts and Sciences distinguished professor and distinguished professor of education at the University of Illinois at Chicago, interview via video conference, December 29, 2020, on file with author.

9 Abby Javurek, Northwest Education Association (a testing company helping states carry out the pilots), interview via video conference, October 5, 2020 and December 18, 2020, on file with author.

10 A few years ago, Louisiana provided incentives to encourage districts to adopt curriculum from a small handful of high-quality curricula, reviewed and vetted by educational experts in the state. Louisiana Department of Education, "Curriculum," available at https://www.louisianabelieves.com/academics/curriculum (last accessed March 2021).

11 Rosario Martinez Arias, "Performance Assessment," *Papeles de Psicologo* 31 (1) (2010): 85–96, available at http://www.psychologistpapers.com/English/1799.pdf; Emily R. Lai, "Performance-based Assessment: Some New Thoughts on an Old Idea," *Pearson Research Bulletin* 20 (2011): 1–4, available at http://images.pearsonclinical.com/images/tmrs/Performance-based-assessment.pdf.

12 New Hampshire Department of Education, "Performance Assessment of Competency Education," available at https://www.education.nh.gov/who-we-are/division-of-learner-support/bureau-of-instructional-support/performance-assessment-for-competency-education (last accessed March 2021).

13 New York Performance Standards Consortium, "The Assessment System," available at http://www.performanceassessment.org/how-it-works (last accessed March 2021).

14 Ibid.

15 Lai, "Performance-based Assessment."

16 Ibid.

17 New Hampshire Department of Education, "Performance Assessment of Competency Education."

18 Ballotpedia, "List of school districts in New Hampshire," available at https://ballotpedia.org/List_of_school_districts_in_New_Hampshire (last accessed July 2021).

19 Information conveyed to the author during a visit to New Hampshire in 2016. Notes on file with authors.

20 Carla M. Evans, "Effects of New Hampshire's Innovative Assessment and Accountability System on Student Achievement Outcomes After Three Years," *Education Policy Analysis Archives* 27 (10) (2019), available at https://www.education.nh.gov/sites/g/files/ehbemt326/files/files/inline-documents/effectnhpace3years.pdf.

21 Fiona Middleton, "The four types of validity," Scribbr, September 6, 2019, available at https://www.scribbr.com/methodology/types-of-validity/.

22 Samuel A. Livingston, "Test reliability—basic concepts" (Princeton, NJ: Educational Testing Seervice, 2018), available at https://www.ets.org/Media/Research/pdf/RM-18-01.pdf.

23 Amy I. Berman, Edward H. Haertel, and James W. Pellegrino, "Comparability of Large-Scale Educational Assessments: Issues and Recommendations" (Washington: National Academy of Education, 2020), available at https://naeducation.org/wp-content/uploads/2020/06/Comparability-of-Large-Scale-Educational-Assessments.pdf.

24 Phoebe C. Winter, "Evaluating the Comparability of Scores from Achievement Test Variations" (Washington: Council of State School Officers, 2010), available at https://files.eric.ed.gov/fulltext/ED543067.pdf.

25 Office of Elementary and Secondary Education, "A State's Guide to the U.S. Department of Education's Assessment Peer Review Process" (Washington: U.S. Department of Education, 2018), available at https://www2.ed.gov/admins/lead/account/saa.html#Standards_and_Assessments_Peer_Review; Office of Elementary and Secondary Education, "Application for New Authorities under the Innovative Assessment Demonstration Authority" (Washington: U.S. Department of Education, 2020), available eat https://www2.ed.gov/admins/lead/account/iada/iadaapplication2020.pdf.

26 Office of Elementary and Secondary Education, "Application for New Authorities under the Innovative Assessment Demonstration Authority" (Wasihington: U.S. Department of Education, 2020), available at https://www2.ed.gov/admins/lead/account/iada/iadaapplication2020.pdf.

27 Megan Stubbendeck, " New SAT/ACT Concordance Tables," ArborBridge, June 14, 2018, available at https://blog.arborbridge.com/new-sat-act-concordance-tables.

28 U.S. Department of Education Office of Elementary and Secondary Education, "Competitive Grants for State Assessments," available at https://www2.ed.gov/programs/cgsa/index.html (last accessed July 2021).

29 Ibid.; U.S. Department of Education, "School Improvement Programs: Fiscal Year 2021 Budget Request" (Washington: 2021), available at https://www2.ed.gov/about/overview/budget/budget21/justifications/d-sip.pdf; U.S. Department of Education, "President's FY 2021 Budget Request for the U.S. Department of Education," available at https://www2.ed.gov/about/overview/budget/budget21/index.html (last accessed July 2021).

30 Institute of Education Sciences, "ED/IES Small Business Innovation Research," available at https://ies.ed.gov/sbir/ (last accessed July 2021).

31 U.S. Department of Education, "Agency Information Collection Activities; Comment Request; Evaluation of the Innovative Assessment Demonstration Authority Pilot Program-Survey Data Collection," *Federal Register* 85 (171) (2020): 54541-54542, available at https://www.govinfo.gov/content/pkg/FR-2020-09-02/pdf/2020-19421.pdf.

32 Kaggle, "The Hewlett Foundation: Automated Essay Scoring Develop an automated scoring algorithm for student-written essays," available at https://www.kaggle.com/c/asap-aes (last accessed July 2021).

33 Data Quality Campaign, "Show Me the Data: DQC's Annual Analysis of Report Cards," available at https://dataqualitycampaign.org/resource/show-me-the-data-reports/ (last accessed July 2021).

34 Data Quality Campaign, "Show Me the Data 2020," available at https://dataqualitycampaign.org/showmethedata-2020/ (last accessed July 2021).

Our Mission

The Center for American Progress is an independent, nonpartisan policy institute that is dedicated to improving the lives of all Americans, through bold, progressive ideas, as well as strong leadership and concerted action. Our aim is not just to change the conversation, but to change the country.

Our Values

As progressives, we believe America should be a land of boundless opportunity, where people can climb the ladder of economic mobility. We believe we owe it to future generations to protect the planet and promote peace and shared global prosperity.

And we believe an effective government can earn the trust of the American people, champion the common good over narrow self-interest, and harness the strength of our diversity.

Our Approach

We develop new policy ideas, challenge the media to cover the issues that truly matter, and shape the national debate. With policy teams in major issue areas, American Progress can think creatively at the cross-section of traditional boundaries to develop ideas for policymakers that lead to real change. By employing an extensive communications and outreach effort that we adapt to a rapidly changing media landscape, we move our ideas aggressively in the national policy debate.

Center for American Progress

1333 H STREET, NW, 10TH FLOOR, WASHINGTON, DC 20005 • TEL: 202-682-1611 • FAX: 202-682-1867 • WWW.AMERICANPROGRESS.ORG