



Artificial Intelligence

By Laura Jimenez and Ulrich Boser | September 16, 2021

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—this issue brief—reviews advancements in technology, with a focus on artificial intelligence that can powerfully drive learning in real time. And the third part looks at assessment designs that can improve large-scale standardized tests.

Despite the often-negative discussion about testing in schools, assessments are a necessary and useful tool in the teaching and learning process.¹ This is especially true when it comes to diagnostic and formative assessments, which give teachers real-time direction for what students need to learn to master course content. It is this space where the advancements of technology can particularly benefit teaching and learning, as there is growing recognition in the field of psychology that tests help students learn. Sometimes called the testing effect, this theory suggests that low-stakes quizzes help students gain knowledge—and improve instruction.²

Advancements in technology have led to new developments in the field, such as stealth assessments, that reduce some of the stress students may experience around assessments. This approach makes testing more ubiquitous and useful for teachers because the methods are woven into the fabric of learning and are invisible to students.³

But to get to a place where all teachers have access to such tests, there needs to be greater investment in testing research and development that results in better systems of diagnostic and formative assessments. This issue brief reviews developments in artificial intelligence (AI) and the types of advancements in diagnostic and formative testing it makes possible. This issue brief ends with recommendations for how the federal government can invest in testing research and development.

What is artificial intelligence in education?

At its most basic level, AI is the process of using computers and machines to mimic human perception, decision-making, and other processes to complete a task. Put differently, AI is when machines engage in high-level pattern-matching and learning in the process.

There are a number of different ways to understand the nature of AI. Two types of assessment include rules-based and machine learning-based AI. The former uses decision-making rules to produce a recommendation or a solution. In this sense, it is the most basic form. An example of this kind of system includes an intelligent tutoring system (ITS), which can provide granular and specific feedback to students.⁴

Machine learning-based AI is more powerful since the machines can actually learn and become better over time, particularly as they engage with large, multilayered datasets. In the case of education, machine learning-based AI tools can be used for a variety of tasks such as monitoring student activity and creating models that accurately predict student outcomes. While machine learning-based AI is still in its infancy, the approach has already shown impressive results when it comes to complex solutions not governed by rules, such as scoring students' written responses or analyzing large, complex datasets.

Within AI, there are other important distinctions, largely based on the technological use cases. One subfield revolves around natural language processing, which is the use of machines to understand text. Technology such as automated essay scoring uses natural language processing to grade written essays. Also important within AI are recommender and other prediction systems that engage in data-driven forecasting. For example, Netflix currently uses an AI-based recommender system to suggest new films to its users.

Vision-based AI is also an important field that can help with assessment. A number of assessment groups have used optical systems to grade students' work. Instead of a teacher grading a math equation that a student wrote, for example, the teacher can snap a picture of the equation, and a machine will grade it. Finally, there are AI systems based on voice recognition. These systems are the backbone of tools such as Siri and Alexa, and experts have been exploring ways to use voice-based AI to diagnose reading and other academic issues.

Despite the innovation that AI supports in assessment, concerns around bias may prevent some of these designs from seeing the light of day. This issue brief will discuss those concerns.

Who is using AI?

Uses of AI in education expand beyond student assessments and into other tools to support student learning, often using built-in stealth assessments that students do not even recognize as a test.

For example, researchers at Carnegie Mellon University’s Human-Computer Interaction Institute developed new ways to use rules-based AI through intelligent tutoring systems.⁵ Their method allows students and teachers to create tutors by entering problems and showing the ITS how to solve them. Once learned, the computer applies the solution; if incorrect, the human can fix it. Thereafter, the computer continues to build the rules, making the machine capable of applying solutions to other problems. This feature makes the tool much faster at building the tutoring system because humans no longer need to build the rules in the system. For example, a teacher can build a 30-minute lesson in about 30 minutes—all through a free tool.⁶ These systems are much more scalable than human-based tutoring, providing students with one-on-one support.

Today, the use of machine-based AI is already fairly widespread in education. For example, several testing companies, such as the Education Testing Service and Pearson, use natural language processing to score essays. Massive online open courses allowing unlimited participation through the web, run by companies such as Coursera and Udacity, have also integrated AI scoring to analyze essays within their courses. Most states also currently use natural language processing to score the essay portion of their yearly assessment.

Such technology can also be used to drive down the cost of assessment. Using a mix of machine learning and natural language processing, several experts such as Neil Heffernan at Worcester Polytechnic Institute are looking at ways to automatically generate new, high-quality test items around a body of knowledge. Heffernan calls the items “similar but not the same,” and he argues that they are key in truly understanding if a student understands a domain.⁷ In some cases, experts believe that machines will soon be able to generate assessment questions that are personalized to a student’s interests. For a student who loves baseball and is learning the concept of 5 plus 3, the machines might generate a problem about baseball (for example, “The batter hit five line drives and three homeruns. How many total hits did they have?”). These efforts on item generation also have the benefit of driving down the costs of assessment.

While natural language processing does not “understand” language in any technical sense, it can be used to evaluate the quality of essays in ways that make formative assessment much more powerful. For instance, most word processing and email programs use natural language processing to suggest greetings or specific words.⁸ Commercial products such as Grammarly also use natural language processing technology to act as a virtual writing assistant. These approaches are

particularly important when it comes to improving formative assessment, and one of the authors of this issue brief has a forthcoming tool that will automatically evaluate a student summary of a reading assignment. Other organizations such as Revision Assistant and MIWrite also use natural language processing to evaluate the quality of argumentative essays.⁹

When it comes to recommender systems, one use case is credit transfer. Researcher Zachary Pardos has created recommender systems that help students transfer credits from community colleges to four-year colleges.¹⁰ Another use case is recommending instructional practices after an assessment. For instance, a recommender system would outline a specific instructional path for a student to take after an assessment. This is important given the often limited practical utility of many end-of-year state exams.

Such predictive systems, also known as early warning systems, can help track students who are in danger of weak academic performance. About half of public high schools and 90 percent of colleges use an early warning system to track student grades, attendance, and other factors to identify when students veer off track.¹¹ These systems are powerful because they can rely on other performance data—such as attendance—to predict student success, allowing counselors and other faculty to intervene early.

Vision-based AI systems can also help with assessment and are being rolled out in a number of areas. Assessment groups such as Pearson have used optical systems to grade students' work, and some, such as the team at the education technology firm Bakpax, envision a world in which teachers use the camera on their cell phones to take a picture of a child's homework, which is then automatically graded.¹² Finally, there are AI systems based on voice. These systems are the backbone of tools such as Siri and Alexa, and experts such as John Gabrieli, a neuroscientist at the Massachusetts Institute of Technology, and Yaacov Petscher, a professor at Florida State University, have been exploring ways for voice-based AI tools to be used to diagnose reading issues.¹³

The benefits and challenges of AI

Artificial intelligence can help students learn better and faster when paired with high-quality learning materials and instruction. AI systems can also help students get back on track faster by alerting teachers to problems the naked eye cannot see. In some cases, such as automated essay scoring, teachers and students do not directly experience the benefits of the tools. Rather, the state grades the exams in a faster, more efficient manner. In other cases, teachers are the direct beneficiaries. Scholars, such as Scott Crossley at Georgia State University, are experimenting with ways that natural language processing-based assessments can be embedded into writing programs so that teachers can get data reports on their students' writing quality.¹⁴

Despite these benefits, there are clear concerns. One major issue is around privacy. How do these tools protect user privacy? How do schools gain consent of both students and parents when introducing them? Should data that have been anonymized be shared with researchers and other external groups? Another issue is the value of social and emotional ties and the very human experience of education. Put simply, AI will not replace teachers.¹⁵ Experts also point to bias as a drawback of AI. Scores computed by machines will be based on the results of thousands of tests. But as noted in this issue brief, test results can more often reflect a lack of opportunity rather than lack of ability. Machine scoring will not be able to make these distinctions.

Defining bias in testing, AI, and big data

Bias occurs when student inputs are misinterpreted and, in turn, misevaluated and scored differently.

- Bias in AI and big data comes in four forms:¹⁶
- The incoming data contain built-in bias. That is, poor outcomes such as low scores may result from fewer opportunities for students to learn, rather than differences in ability.
- Poor past performance predicts poor future performance. For example, students who performed poorly in the past will repeat it.
- The use of AI breeds a lack confidence that the outcomes are fair. Since the incoming data can have biases, the outcomes may as well.¹⁷
- The use of AI continues past inequities, and gaps in access to opportunities to achieve at high levels continue.

Given how fast computer programs operate, they can apply biases more quickly and efficiently than humans can.

Experts agree that bias in testing, AI, and big data will always exist. Therefore, eliminating bias may be the wrong goal. Instead, policymakers who oversee testing systems must ask themselves how much and what type of bias is tolerable, as well as how to ensure that bias does not disproportionately affect students based on race, ethnicity, income, disability, or English learner status.

[How to reap the benefits of AI to improve testing](#)

Three steps will get educators and students closer to reaping the benefits of AI and its uses in student assessments.

First, Congress must invest in research to better understand where and how bias occurs in testing. Test results should be fair and accurate reflections of what students know and can do against a common and fair measuring stick. But when test results consistently exhibit racial patterns—and do not reflect true differences between the groups—they are biased.¹⁸ Bias could occur in what is being measured or in how it is being measured and scored. Research can point to where in the testing process bias is occurring and help discover remedies.

Second, Congress should invest in the development of new kinds of technology-driven assessments. Given the size and scale of investment needed, this can only come from the federal government. Thus, Congress should provide additional funding to states for testing and related research and development on cutting-edge technology such as AI-based tools, learning games, and virtual reality. This could take the form of increased funding for the Grants for State Assessments and Related Activities program in the Every Student Succeeds Act. Congress should also increase funding of a little-known program called the Small Business Innovation Research program, which provides up to \$1.1 million in individual grant awards to develop education-related learning technologies.¹⁹ Congress should also orient this program to have more of a focus on assessment strategies rather than general education technology.

Third, the federal government should invest in teacher professional development on the effective use of assessments. Teachers should be experts in creating their own assessments as well as in using the results of any assessments to customize learning supports for students. Countries such as Finland and Australia invest heavily in supporting teachers to effectively use assessments and could be a model for the United States to follow.²⁰

Conclusion

Well-designed formative assessments that take advantage of the latest advancements in technology can help students learn faster and better. These mechanisms are also a critical part of the teaching and learning process. From intelligent tutoring, stealth assessments, games, and virtual reality, mini-tests built by artificial intelligence can provide a wide variety of ways to use this technology to build engaging tools. To get there, the education system needs stronger investments in the research and development of new testing technologies that can provide teachers and students with the tools they need.

Laura Jimenez is the director of Standards and Accountability for K-12 Education at the Center for American Progress. Ulrich Boser is a senior fellow at the Center.

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

- 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 Corporation
- 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 Bari Walsh, "When Testing Takes Over: An expert's lens on the failure of high-stakes accountability tests—and what we can do to change course," Usable Knowledge, November 3, 2017, available at <https://www.gse.harvard.edu/news/uk/17/11/when-testing-takes-over>.
- 2 Teach.com, "What is the Testing Effect and How Can You Apply It in Your Approach to Educating," available at <https://teach.com/what/teachers-know/testing-effect/#:~:text=The%20testing%20effect%20is%20a,ideas%20and%20information%20long%20term> (last accessed August 2021).
- 3 Val Shute, "Stealth assessment in video games" (Tallahassee, FL: Florida State University, 2015), available at https://research.acer.edu.au/cgi/viewcontent.cgi?article=1264&context=research_conference#:~:text=Very%20simply%2C%20stealth%20assessment%20refers,the%20learning%20or%20gaming%20environment.&text=Evidence%20needed%20to%20assess%20the,%2C%20the%20processes%20of%20play.
- 4 James A. Kulik and J.D. Fletcher, "Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review," *Review of Educational Research* 86 (1) (2015): 42–78, available at https://www.researchgate.net/publication/277636218_Effectiveness_of_Intelligent_Tutoring_Systems_A_Meta-Analytic_Review.
- 5 Roybi Robot, "How Intelligent Tutoring Systems are Changing Education," Medium, July 31, 2020, available at [https://medium.com/@roybirobot/how-intelligent-tutoring-systems-are-changing-education-d60327e54dfb#:~:text=Intelligent%20Tutoring%20Systems%20\(ITS\)%20are,one%20Don%20Done%20curriculum](https://medium.com/@roybirobot/how-intelligent-tutoring-systems-are-changing-education-d60327e54dfb#:~:text=Intelligent%20Tutoring%20Systems%20(ITS)%20are,one%20Don%20Done%20curriculum).
- 6 TutorShop, "Home," available at <https://tutorshop.web.cmu.edu/> (last accessed August 2021); Byron Spice, "New AI Enables Teachers to Rapidly Develop Intelligent Tutoring Systems," *Carnegie Mellon University News*, May 6, 2020, available at <https://www.cmu.edu/news/stories/archives/2020/may/intelligent-tutors.html>.
- 7 Andrew Burnett, "The Power of Immediate Feedback," 7th Grade Math with Mr. Burnett, August 27, 2020, available at <https://burnettmath.wordpress.com/category/improving-teaching-and-learning/>.
- 8 Grammarly, "How We Use AI to Enhance Your Writing," May 19, 2019, available at <https://www.grammarly.com/blog/how-grammarly-uses-ai/>.
- 9 Turnitin.com, "About," available at <https://www.turnitin.com/about> (last accessed August 2021); Measurement Incorporated, "About Us," available at <https://measurementinc.com/about-us> (last accessed August 2021).
- 10 UC Berkeley Graduate School of Education, "Zachary A. Pardos," available at <https://gse.berkeley.edu/zachary-pardos> (last accessed August 2021).
- 11 U.S. Department of Education Office of Planning, Evaluation, and Policy Development, "Issue Brief: Early Warning Systems" (Washington: 2016), available at <https://www2.ed.gov/rschstat/eval/high-school/early-warning-systems-brief.pdf>; Hanover Research, "Early Warning Systems in Higher Education" (Washington: 2014), available at <https://www.hanoverresearch.com/wp-content/uploads/2017/08/Early-Alert-Systems-in-Higher-Education.pdf>.
- 12 Bakpax.com, "Why Bakpax?," available at <https://www.bakpax.com/> (last accessed August 2021).
- 13 Gabrieli Laboratory, "Publications," available at <https://gablabs.mit.edu/research/publications/> (last accessed August 2021); Florida State University College of Social Work, "Yaacov Petscher," available at <https://csw.fsu.edu/person/yaacov-petscher> (last accessed August 2021).
- 14 Georgia State University, "Scott Crossley," available at <https://education.gsu.edu/profile/scott-crossley/> (last accessed August 2021).
- 15 Robert F. Murphy, "Artificial Intelligence Applications to Support K-12 Teachers and Teaching: A Review of Promising Applications, Challenges, and Risks" (Santa Monica, CA: Rand Corporation, 2019), available at <https://www.rand.org/pubs/perspectives/PE315.html>.
- 16 Kelly Capatosto, "Bias in Big Data: Predictive Analytics and Racial Justice," available at <https://www.purdue.edu/diversity-inclusion/documents/msp-wg-data-bias.pdf> (last accessed August 2021).
- 17 Tatiana Lukoianova and Victoria L. Rubin, "Veracity Roadmap: Is Big Data Objective, Truthful and Credible?," *Advances in Classification Research Online* 24 (1) (2014): 4–15, available at <https://journals.lib.washington.edu/index.php/acro/article/view/14671/12311>.
- 18 Cecil R. Reynolds and Lisa A. Suzuki, "Bias in Psychological Assessment: An Empirical Review and Recommendations," in Irving B. Weiner, ed., *Handbook of Psychology, Second Edition* (Hoboken, NJ: John Wiley & Sons Inc., 2013), available at <https://onlinelibrary.wiley.com/doi/full/10.1002/9781118133880.hop210004>.
- 19 Small Business Innovation Research, "About," available at <https://www.sbir.gov/about> (last accessed August 2021).
- 20 Victoria State Government, "Assessment," available at <https://www.education.vic.gov.au/school/teachers/teachingresources/practice/Pages/assessment.aspx> (last accessed August 2021); Hannele Niemi, "Teacher Professional Development in Finland: Towards a more holistic approach," *Psychology, Society and Education* 7 (3) (2015): 279–294, available at https://www.researchgate.net/publication/301225639_Teacher_Professional_Development_in_Finland_Towards_a_More_Holistic_Approach.