

# Can Speech Recognition Help Children Learn to Read?

**A primer on the applications, risks —  
and aspirations—of speech recognition  
technology in the classroom**

---

By Evo Popoff

Foreword by Margery Mayer

Produced by



Commissioned by



---

## FOREWORD: An Idea Whose Time Has Come

---

In 1999, my colleagues at Scholastic Education and I launched a new reading intervention program called READ 180 that we'd hoped to equip with speech recognition capabilities. We imagined kids reading aloud to a computer program that could actually "listen" to them speak, generate feedback — and provide teachers with valuable insights to inform their practice.

We decided not to employ the technology for a variety of reasons. For starters, we were worried about the accuracy of the technology: we didn't want to tell kids they were mastering something when they weren't, and we understood the profound implications of telling them they got something wrong when they were right.

With good reason: At the height of the 1990s tech bubble, speech recognition for classrooms was still largely the stuff of science fiction (recall that none of us met Siri until 2010). The field of speech recognition, still in its infancy, hadn't gone mainstream, let alone mapped the terabytes of data required to block out ambient noise in busy classrooms or comprehend the complexity of tiny voices. Children's voices, after all, have different pitches, different speech patterns, different rhythms. They tend to over-elongate and over-enunciate. They sometimes pause frequently or even stumble over words they're not used to pronouncing.

Although READ 180 was about 20 years too early to benefit, recent advances in artificial

intelligence and machine learning have enabled companies like SoapBox Labs to create speech recognition technology that's purpose-built for the realities of the classroom. The technology has advanced to the point where it can not only recognize and process children's speech, but it can also account for differences in accents or dialects in ways that mitigate implicit and unintentional biases. This can have a profound impact on the trajectory of young learners.

Such advances could not have come at a more crucial moment.

Even before the COVID-19 pandemic, more than 80 percent of children from economically disadvantaged families failed to reach reading proficiency by third grade. After a year of separation from in-person instruction by skilled educators and vast gaps in digital equity, students had learned just 87 percent of the reading for a typical year, according to a report from McKinsey & Co. They lost an average of 1.5 months of learning during spring school closures.

Not surprisingly, reading losses were especially acute in schools that serve mostly students of color, where learning was just 77 percent of the historical average.

Individualized observation and assessment was already a persistent challenge before COVID-19, but became even more impractical for a teacher whose students were returning

to school with unprecedented educational and emotional challenges.

As students return to classrooms, speech recognition holds the potential to revolutionize the classroom—not to mention toys and other applications—by transforming the way children interact with technology. Voice-enabled literacy programs can further professionalize the field by taking the administrative work out of capturing and quantifying the multiplicity of variables teachers look for when a child reads aloud. Voice-enabled tools can generate invaluable insights into a student’s reading progress, pick up on patterns, or isolate areas that need improvement.

In response, dozens of education developers and publishers, such as Amplify, have released or announced plans for voice-enabled solutions in recent months. In the world of education

technology, the last year, in many ways, has been the “year of speech recognition.”

This paper considers not just the aspirations but the risks and applications of speech recognition in the classroom. It’s focused, in particular, on the role of speech recognition in supporting educators working with young readers, a topic that is especially near and dear to my heart. I found it to be a compelling and useful primer at a pivotal moment in education history.

In a few years, I have no doubt this technology will be part of every reading classroom, and educators will be able to say with certainty it accelerates reading skills in young students. Educators will find it enables them to be more strategic in their instruction. It’s an idea whose time has come.

---

### ***About Margery Mayer***

*Margery Mayer served for 25 years as President of Education at Scholastic, where she was the leading force behind the development of Read 180 along with a family of other successful edtech products. Today, Margery advises a number of innovative companies and serves as a board member of Teachers Pay Teachers and Cambium Learning Group. Margery also sits on the boards of non-profit schools including Public Prep, a life-changing network of charter schools in New York City, and the Middlebury Language Schools and the Bread Loaf School of English at Middlebury College.*

---

# TABLE OF CONTENTS

---

<b>Table of Contents</b>	<b>4</b>
<b>About the Author</b>	<b>6</b>
<b>Acknowledgements</b>	<b>7</b>
<b>Executive Summary</b>	<b>8</b>
<b>The Problem with Learning to Read Online</b>	<b>9</b>
<b>2011: A Dream Deferred</b>	<b>11</b>
<b>Why Now? From Vision to Reality</b>	<b>14</b>
• <i>AI Unlocks a New Age of Speech Recognition for Children</i>	
• <i>2021: The Year of Speech Recognition in Education</i>	
<b>Equity Implications</b>	<b>17</b>
<b>Translating Potential into Practice: Case Studies</b>	<b>19</b>
<b>Conclusion</b>	<b>24</b>



Whiteboard Advisors is a social impact agency for education.

For more than 15 years, our team of educators, policy wonks, and storytellers has brought an understanding of policy, technology, and practice to bear on to our collaboration with the most transformative organizations and entrepreneurs in education.

We're passionate about taking breakthrough ideas to scale to reimagine how we learn, work, and live.



SoapBox Labs powers voice-enabled experiences for kids ages 2-12 of all accents and dialects.

The SoapBox voice engine has been built using a privacy-by-design approach. Protecting kids' fundamental right to voice data privacy is a cornerstone of our mission and philosophy.

SoapBox licenses its specialist technology to third party education and entertainment companies, offering two solution suites:

**SoapBox Educate** for instruction, practice, screening, intervention, and assessment of literacy, math, and language learning.

**SoapBox Play** for games, toys, apps, and entertainment applications for the home and consumer markets.

---

## THE AUTHOR

---

Named State Policy Maker of the Year by the State Education Technology Directors Association (SETDA), Evo previously served as Chief Innovation and Intervention Officer and Assistant Commissioner for the New Jersey Department of Education, where he oversaw the state's education technology and school and district improvement efforts.

Prior to joining the Department, Evo Popoff led the development of education technology products and school improvement solutions in collaboration with district and state leaders, and educators.

Before beginning his career in education, Evo practiced law at McDermott, Will & Emery where he worked on labor and employment, antitrust, and general corporate issues.

He holds a B.A. in Political Science from the University of Chicago and a J.D. from the George Washington University Law School.

---

## ACKNOWLEDGEMENTS

---

We are grateful for the expertise of friends and practitioners who have provided valuable support, insights, research, and perspectives to inform this paper, including:

- Dylan Arena, Vice President of Learning Science, McGraw Hill Education
- Dan Cogan-Drew, Co-Founder and Chief Academic Officer, Newsela
- Larry Berger, Founder and CEO, Amplify Education
- Kai-leé Berke, Co-Founder, Noni Educational Solutions
- Dan Cogan-Drew, Co-Founder and Chief Academic Officer, Newsela
- Jennifer Carolan, Founder, Reach Capital
- Krista Curran, Senior Vice President, Amplify
- Sara DeWitt, Vice President, PBS Kids Digital
- Martyn Farrow, CEO, SoapBox Labs
- Yolie Flores, National Campaign Director, Parent Nation
- Sunil Gunderia, Chief Innovation Officer and Head of Mastery & Adaptive Products, Age of Learning
- Jonathan Hume, Vice President of Global Sales, SoapBox Labs
- Michael Levine, Senior Vice President of Learning and Impact, Noggin
- Toby Mather, CEO, Lingumi
- Satya Nitta, CEO, Merlyn Mind
- Sean Ryan, President, McGraw Hill School Group
- Patricia Scanlon, Founder and Executive Chair, SoapBox Labs
- Ralph R. Smith, Managing Director, Campaign for Grade Level Reading

---

## EXECUTIVE SUMMARY

---

While billions of people use virtual assistants like Siri or Alexa each day, the largest tech companies have only recently started to invest in training their tools to understand and respond to children. The promise is there, but the impetus to fine-tune their technology for young users has not received the resources deployed to the far larger commercial applications focused on adults.

As a result of relatively recent advances and a growing body of research, child-specific speech recognition systems are now paving the way for a whole new range of voice-enabled tools for use by professional educators in the classroom, and children and families in the home.

Among them: Artificial intelligence-enabled reading products that teachers can use to listen, prompt, and encourage a child on their literacy journey. Such tools can not only encourage and enable regular practice, but they can also invisibly—to the child at least—assess their reading ability and progress, and provide professional educators with valuable insights. Reading solutions are, however, just the beginning, as speech recognition holds the promise for the creation of entirely new learning experiences for children across content areas and media.

Such technologies are not without risk, and technology is never a replacement for the critically important role of professional educators in a child's reading journey. Although embedding purpose-built speech recognition technology within education applications holds potential to capture data that can support teachers and mitigate the impact of implicit biases, the AI that undergirds speech recognition in classrooms could also exacerbate inequities, absent inclusive development and thoughtful research.

This white paper explores the evolution of speech recognition for children and the potential applications (and risks) of this technology for learners, as well as challenges that educators and policymakers should consider. We hope that it does a thoughtful job of highlighting both the vast opportunities and thoughtful concerns in an area of technology that is garnering increased attention.

Of course, the paper itself is not without bias and it is far from an exhaustive treatment of the topic. It was commissioned by a developer of speech recognition technology and produced by a firm that is enthusiastic about its potential. We hope that readers will consider it within that context, and treat it as a primer to spur further discussion and exploration of the topics raised.



---

## THE PROBLEM WITH LEARNING TO READ ONLINE

---

Even before the pandemic, an estimated [617 million](#) children worldwide were struggling to reach basic reading proficiency. Sadly, concerns that school closures would exacerbate longstanding educational inequities have proven founded.

“Over the last year, parents have felt the weight of their children’s education in our inexperienced hands. Few are comfortable with the responsibility that was unexpectedly thrust upon them. They saw, firsthand, the struggles that their children had with using technology to learn,” said Yolie Flores, a former member of the Los Angeles Unified School District and national campaign director for Parent Nation. “Sadly, the most insidious long-term effects of remote learning may have yet to be realized. But the perspectives of parents, based on their experiences over the last year, should play a big role in designing solutions for the future.”

Research now shows that [nearly one-third of U.S. second-graders returned to school in fall 2020](#) two or more grade levels behind in reading, 6 percentage points higher than the historical average.

“That’s because reading instruction requires a blend of art and science that even trained educators struggle to replicate online,” according to Dylan Arena, vice president of Learning Science at McGraw Hill Education. “Until recently, capturing, synthesizing, and spotting

“

---

Over the last year, parents have felt the weight of their children’s education in our inexperienced hands. Few are comfortable with the responsibility that was unexpectedly thrust upon them. They saw, firsthand, the struggles that their children had with using technology to learn.” - *Yolie Flores, a former member of the Los Angeles Unified School District and national campaign director for Parent Nation*

the complex patterns exhibited by early readers has been nearly impossible for computers.”

Unlike content-driven subjects like math or science that have spawned thousands of online apps and games, reading instruction requires careful listening by an adult who helps a child by encouraging, correcting, and prompting them along their reading journey. Because all students start their reading journeys with different abilities and learn at a different rate, individualized screening and assessment is crucial.

Assessing reading levels and screening for reading difficulties are also observational, meaning educators must sit with each child individually and listen as they read, evaluating their ability to recall, pronounce sounds, and blend, decode, and identify words. For more mature readers, educators must count the words correct per minute, noting any deletions, substitutions, and insertions, to assign each child a level of reading fluency.

Assessment needs to occur with high frequency to catch struggling readers before they begin to fall behind their peers. With good reason: Research suggests it takes four times longer to intervene with an 8-year-old student who ideally should be approaching reading fluency, than it does with a 5-year-old in the early stages of the reading journey.

“Teaching reading is already hard in person; teaching children and assessing their ability to read independently is especially hard and can be incredibly time consuming,” said Kai-leé Berke, co-founder of Noni Educational Solutions and a lifelong educator and curriculum and assessment developer. “In a typical classroom, teachers are expected to compute a child’s fluency by presenting them with a passage to read aloud and calculating the words per minute, minus errors such as deletions, substitutions, insertions. That not only places a sometimes untenable administrative burden on teachers that eats away at instructional time, but it’s also limited by our innate computing capacity, subjectivity, and implicit biases.”

To put the challenge into perspective: 15 minutes of individual time per child in a class of 25 would take over more than six hours of a teacher’s day, every day.

“

---

Teaching reading is already hard in person; teaching children and assessing their ability to read independently is especially hard and can be incredibly time consuming.”

- *Kai-leé Berke, co-founder of Noni Educational Solutions and a lifelong educator and curriculum and assessment developer*

The levels of literacy assessment and instruction required to redress learning loss and put students on a path to proficiency are, as a result, limited by the realities of the modern classroom. The discrete learning challenges faced by individual children are often missed by educators in under-resourced and crowded classrooms where teachers attend to a wide range of academic and socioemotional needs. All of that is compounded by the fact that reading assessments designed to catch struggling readers early are often only delivered once or twice a year.

“At present, even the best models for literacy instruction can be derailed by the logistical hurdles associated with delivery,” said Ralph R. Smith, Managing Director at the Campaign for Grade Level Reading. “This is unfortunate, unnecessary and unacceptable. We can and must do better.”

That may soon be changing.

---

## 2011: A DREAM DEFERRED

---

“

In the early years of our company, people used to say, ‘Wouldn’t it be cool if our technology could actually listen to children read, and give them a score, just like that?’ And I used to explain why that just wasn’t possible. Then I would explain our equity concerns: The kids who have an accent or whose voice is different in one way or another would get a differentially lower score on a measure that needs to be equitable. We were monitoring the world of speech recognition intensely at the time to see if anybody was ready, and we’d test the latest solutions against a bunch of kids with different accents and voice tonalities, and it just wasn’t ready. — *Larry Berger, founder and CEO of Amplify*

A 2011 report by Dr. Marilyn Jagar Adams from the Joan Ganz Cooney Center proposed using automatic speech recognition (ASR) technology to enable automated reading tutors that could support children’s literacy development. The appeal of ASR technology was clear: It provided the only path for giving every student access to one-to-one guided oral reading, the best practice for developing fluency. ASR technology could solve the basic numbers problem of there being too little time in the day to provide one-to-one support to every student.

By solving this numbers problem, a challenge that is especially acute in lower-income schools, ASR holds the promise to democratize guided oral

reading, making it freely available to all students. In such applications, the report states, ASR can be used to track the reading position of the child; detect oral reading errors like substitution, omission, or hesitation; prompt the child where necessary; and assess reading fluency and comprehension of the text being read.

The report cites numerous other applications beyond guided oral reading that deploy ASR technology for the benefit of young learners, including prosody and understanding; managing interactive learning; and reading assessment. Given the paper’s conclusion that ASR technology is “up to the task” despite a handful of challenges Adams identifies with the then-

state-of-the-art technology, a reader would be forgiven for assuming that, in 2011, we were on the cusp of a new golden age of ASR-powered early literacy tools.

But a decade later, we're only now at the beginning of the era of speech recognition in education.

What happened?

The Cooney Center paper got a lot of things right. First, the challenge of providing adequate time for one-to-one reading support for young learners is very real, as any elementary school teacher will tell you. There simply isn't enough time to provide the level of one-to-one support, including guided oral reading, to every student in a class—particularly in schools with more struggling learners.

Second, speech recognition holds immense promise for education in general and early literacy in particular. And the potential to provide every student with access to powerful literacy supports is significant.

Finally, the paper wasn't entirely wrong about the state of the technology: Apple's Siri was introduced in 2010, followed by Amazon's Alexa. As speech recognition matured, it became increasingly commonplace in industry after industry—and in our homes. Today, it's estimated that [over half of adults in the U.S. who own a smartphone use voice assistants, and one in every four Americans owns a smart speaker](#), such as Amazon Echo or Google Home.

“Although billions use virtual assistants like Siri or Alexa each day, the largest tech companies have only recently started to invest in training their tools to understand and respond to

children,” said Jennifer Carolan, founder of Reach Capital. “The promise is there, but the impetus to fine-tune their technology for young users has not received the resources earned by far more lucrative commercial applications.”

The challenge stems from the fact that the technology that enables most popular speech recognition is designed primarily with adult voices, for adult voices.

For children in classrooms and informal learning environments where accuracy is important, the speech recognition used in voice assistants simply isn't up to the task. Children's voices are physically different from adults. Children have shorter, thinner vocal tracts; smaller vocal folds; and an immature larynx. Children's language is also vastly different from adults' language and changes

“

---

Although billions use virtual assistants like Siri or Alexa each day, the largest tech companies have only recently started to invest in training their tools to understand and respond to children. The promise is there, but the impetus to fine-tune their technology for young users has not received the resources earned by far more lucrative commercial applications.” - *Jennifer Carolan, founder of Reach Capital*

greatly as they develop. Children’s behaviors are also wildly erratic when compared to the predictability of adult speech. For all of these reasons, the younger the child, the worse the voice system’s performance.

Until recently, most speech recognition applications for kids simply modified technology originally designed for adults. The paper’s author references “hacks” that can be used to mimic the “shorter vocal tracts” of kids. The challenge, as any parent can attest, is that the differences between adults and children come down to more than just size. There are physiological differences beyond the size of the vocal tracts, in addition to significant behavioral differences, that result in lower accuracy when applying the technology to young children.

Systems did exist that were trained on kids’ voices, but they, too, had shortcomings. They used technology “trained” with data sets of children’s voices recorded in studio conditions with microphones and were only accurate

under similar conditions—meaning students needed to use them with microphones in quiet settings. By their nature, children inhabit noisy environments like kitchens, playgrounds, and classrooms. Therefore, to become an accepted and commonplace tool in early childhood and K-12 education, speech technology needed to work for children on all devices and in everyday classroom and home settings.

Moreover, the data sets didn’t take into account a broad spectrum of accents and dialects, exposing these systems to be biased toward certain users.

“

---

Voice gives kids agency and allows them to learn and play in the most natural way possible—using their voices. - *Dr. Martyn Farrows, CEO of SoapBox Labs*

---

# WHY NOW? FROM VISION TO REALITY

## Artificial Intelligence Unlocks a New Age of Speech Recognition for Children

---

The Cooney Center report, in many ways, ushered in a new age of speech recognition for children. Among the many people who read the piece was an engineer and researcher in Dublin, Ireland, who had dedicated her professional life to working on speech technology.

As a mother of young children when the paper was published, Dr. Patricia Scanlon appreciated its vision for speech technology in education. She had witnessed the shortcomings of the early reading apps her children were using and wondered why they didn't make use of speech recognition. Exploring the answer to that question would take her on a journey that included launching a company dedicated to the development of the first speech technology designed specifically for kids.

Founded in 2013, SoapBox Labs, which commissioned this paper, develops speech recognition for kids that's modeled on their speech, language, behaviors, environments, and privacy needs. The company's proprietary speech recognition powers immersive play and learning experiences, including literacy, math, and language learning tools, for children ages 2 to 12.

Unlike speech recognition technology that's not designed specifically with children in mind, SoapBox Labs began by collecting thousands of hours of children's voice data from 192 unique countries. It was able to do so while still complying with the U.S.-based Children's

Online Privacy Protection Act (COPPA) and the European-based General Data Protection Regulation (GDPR) by incorporating a concept called privacy-by-design, which means none of the data collected was identifiable by child or tied to a unique user profile that could follow a child across a lifetime. Moreover, as part of their privacy-by-design approach, the company committed early on to never share or sell data to third parties or use data collected for marketing or advertising purposes.

That treasure trove of speech data enabled Dr. Scanlon and her team to develop and scale a speech recognition technology solution that is accurate for kids speaking in English of all ages, backgrounds, accents, and dialects. What further sets the SoapBox Labs' dataset apart are voice recordings collected in places where kids interact naturally—classrooms and kitchens, playgrounds, front yards.

The diversity of these datasets enables the speech recognition to understand a variety of accents and dialects, including those from racial, ethnic, and linguistic minority groups. The SoapBox voice engine has been independently validated to be free of bias—both racial and socioeconomic—making it inclusive for all children speaking English.

Driven by artificial intelligence (AI), the speech recognition technology developed by SoapBox Labs gives the computer or device “ears,” allowing

it to listen, prompt, correct, and encourage the child as they read, just as a helpful adult would do. This immediate and accurate feedback from the voice-enabled reading tool or app allows the child to progress autonomously, practice regularly, and assess their own reading ability and areas for improvement. For their part, teachers can use a voice-powered reading tool to invisibly assess a child's reading progress, adapt the learning path for each child's needs, and act upon the need for timely interventions when necessary.

“In many ways, that can be a freeing experience for children who struggle with reading aloud,” said Dan Cogan-Drew, co-founder and chief academic officer at Newsela, an instructional content platform that serves students in grades 2 through 12. “There’s real interpersonal stress for a student who has difficulty reading and is reading to another person. In some cases, it’s really much better that you’re reading to a machine than a human, because the machine passes no judgment. It just says, ‘This is what I heard you say.’”

But there is even further untapped potential here when it comes to AI and a child's reading journey. Human reading assessment is, after all, highly subjective. Recent studies have shown that variances of up to 18 percent can be caused by assessor bias.

AI-driven speech recognition doesn't have the same limitations as humans: it can drive scalable and, most importantly, objective formative assessments. It also has the potential to generate deeper insights into a child's reading ability and progress and to retrospectively map assessment data points to outcomes, which generate new key indicators. Voice data represents a whole new category of invaluable data about how individuals, groups, and large numbers of

“

---

In many ways, that can be a freeing experience for children who struggle with reading aloud. There's real interpersonal stress for a student who has difficulty reading and is reading to another person. In some cases, it's really much better that you're reading to a machine than a human, because the machine passes no judgment. It just says, “This is what I heard you say.” - *Dan Cogan-Drew, co-founder and chief academic officer at Newsela*

children learn. It can be used to compare a child's ability to regional, national, and even global benchmarks that are automatically and objectively generated.

Speech recognition holds tremendous promise for significantly improving reading outcomes, a goal that predated the COVID-19 pandemic and has only become more urgent.

### **2021: The Year of Speech Recognition in Education**

Integrating speech recognition into the classroom is far from a new idea in K-12 education. Every few years over the past two decades, a promising new voice-powered solution would be introduced into K-12 schools—and would fail to catch on in classrooms.

But necessity is the mother of invention, which explains why, in 2021, speech recognition is finally making its way to the forefront of the conversation and becoming a top priority for developers of reading products. After all, the pandemic provided the perfect opportunity for innovations that, for example, allow reading instruction to continue even when students are learning from home or moving between homes and classrooms.

That's why companies and educators alike are looking to voice-powered resources in 2021 as a way to increase the quality and efficacy of digital learning and as a way to complement in-person and blended instruction.

During a May 2020 SXSW EDU [panel](#) discussion titled "Will Speech Technology Transform Education?", Sunil Gunderia, Chief Innovation Officer and Head of Mastery and Adaptive Products at Age of Learning, a provider of digital resources for children ages 2-5, noted that "the ability to use something like speech recognition to serve as a way to provide kids meaningful feedback on how they can improve in their reading, fluency, or their pronunciation is really compelling."

During that same SXSW EDU panel, Satya Nitta, former head of IBM's research team on AI for

learning and founder and CEO of Merlyn Mind, said, "SoapBox Labs has laid the groundwork for edtech developers by building speech recognition that's accurate enough to work in an educational setting and by licensing it to other companies."

In a follow-up conversation, Nitta added, "SoapBox Labs has shown it's committed to continuing to develop features that cater to the needs of their audiences. As more capabilities come online, more and more solution providers will come to the table and build bespoke applications for this space."

As stated by Sean Ryan, president of McGraw Hill School Group: "Language is a uniquely human tool for thinking. And speech is the vital adjunct that allows us to gain and grant access to one another's thoughts. Beyond reading and writing, speaking connects two or more people in real time to create meaning between the building blocks of community." As Ryan continues, "At McGraw Hill, we are just beginning to explore the power of automated speech recognition. And so far the technology is exceeding the demanding expectations of our pioneer teachers. The primary determination of mass adoption is trust, built one experience at a time, over time. I see this accelerating as teachers witness the time savings and, more importantly, gain deeper insight and drive positive student outcomes."

“

---

At McGraw Hill, we are just beginning to explore the power of automated speech recognition. And so far the technology is exceeding the demanding expectations of our pioneer teachers. The primary determinant of mass adoption is trust, built one experience at a time, over time. I see this accelerating as teachers witness the time savings and, more importantly, gain deeper insight and drive positive student outcomes.

— *Sean Ryan, president, McGraw Hill School Group*



---

## EQUITY IMPLICATIONS

---

In March 2020, a [New York Times](#) article sent shockwaves through the speech recognition community. Researchers at Stanford University released a [study](#) revealing that speech language systems from Amazon, Apple, Google, IBM, and Microsoft misidentify words from white speakers about 19 percent of the time. That number jumps to 35 percent for Black speakers.

The idea that the AI undergirding speech recognition in classrooms could exacerbate inequities was deeply troubling, but it also helped to shine a light on profound opportunities. In an [op-ed](#) about that research, Dr. Yaacov Petscher and Dr. Nicole Patton Terry from the Florida Center for Reading Research explained that such an assertion wasn't surprising since linguists have long bemoaned the lack of diverse voice samples in speech recognition programs. Given society's increasing reliance on speech recognition technology and the misguided assumption that there's one right way to speak, they concluded that "biases like these are not just annoying; they're dangerous."

But Dr. Petscher and Dr. Terry explained that humans often are just as biased in their own speech recognition, if not more so. Some people haven't been exposed to a variety of accents and dialects, they wrote, and others simply believe there's one "right" way to speak. And those people are more likely to misidentify words spoken in non-standard English.

Biases are especially concerning when applied to children's speech, such as in a reading program that measures what a child says against what the system expects them to say.

Regardless of the evaluator—poor-quality AI or a biased human—the ramifications for the speaker are significant, since people judge others by the way they talk. "The reality is Black children are not broken and neither is the AI," they wrote. "The AI algorithms operate exactly as they are programmed. Perhaps a problem lies in aspects of the design itself."

“

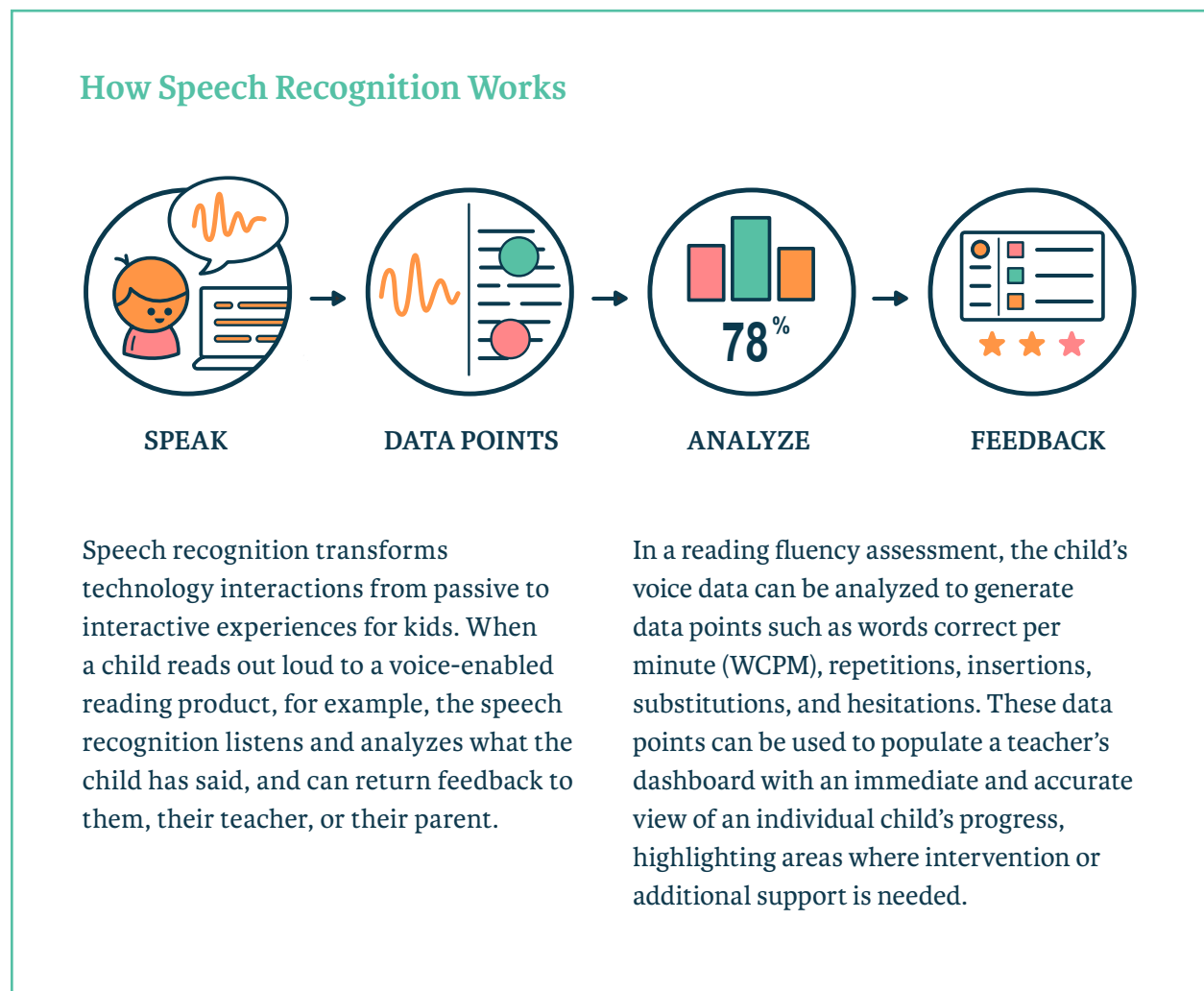
---

As AI-enabled technologies like speech recognition systems become more and more prevalent in everyday life, biases ... are not just annoying — they're dangerous. In the classroom, those dynamics present both profound risk — and opportunity. — *Dr. Yaacov Petscher and Dr. Nicole Patton Terry, Florida Center for Reading Research*

However, AI also holds the promise for weeding out bias, particularly through speech recognition that “proactively design[s] for diversity,” Petscher and Terry wrote. They cite as an example African-American English, which linguists have studied for decades, documenting its specific rules and systems, recognizing it as just as valid as Bostonians who “pahk the cah.” But it’s not enough to simply increase the voice samples of random Black Americans. It takes specialized research on linguistic variations of African-American English to reduce opportunities for bias, they wrote.

Dr. Petscher and Dr. Terry’s research in using dialect variations to improve speech recognition is proving promising. Early results show it may help reduce the disparities pointed out in the Stanford University research.

“If we are successful,” they wrote, “we may unlock a new frontier of assessment—and inform instruction.”



---

# TRANSLATING POTENTIAL INTO PRACTICE: CASE STUDIES

---

With speech recognition advances catching up on their promise, we can now begin to imagine a world where there's no longer a need to press, swipe, or type to engage in learning experiences. The following are among the exciting potential uses for speech recognition in education.

- **Early literacy and fluency:** The first solutions that hit the market focused on early literacy and fluency: areas that benefit from speech recognition's ability to listen

to and understand language, unlocking far more natural and nuanced ways of gauging comprehension than traditional quizzes and tests. Speech recognition, which has an important role to play at each stage of the literacy journey, is also increasingly being leveraged to support the process of “observational assessment” and to help diagnose reading challenges, including dyslexia, at an earlier stage and while students are learning at home.

## CASE STUDY: AMPLIFY

As SoapBox Labs' pioneering work began to catch the attention of the education community, Amplify, a leading U.S. curriculum and assessment company that serves 7 million students across 50 states, saw the opportunity to integrate SoapBox Labs' speech recognition technology into its literacy tools. Together, they plan to release a voice-enabled reading assessment for grades 1 through 6 in time for the 2021-2022 school year. The assessment, called “mCLASS Express,” will be the first literacy assessment to enable remote evaluation of oral reading fluency, accuracy, and comprehension—key indicators of whether students are making the transition from learning to read to reading to learn.

Like most early literacy assessments, however, evaluating oral reading fluency typically requires in-person observation by trained educators. By enabling them to understand and evaluate fluency, accuracy, and comprehension development, mCLASS Express can help educators tailor instruction and avoid reading loss during a critical developmental stage. This is true regardless of educational setting.

“In the decades since we first experimented with the use of mobile computers to support observational assessment, we always imagined speech recognition could unlock a step

change in the early literacy field. But the realities of teaching and learning in a pandemic accelerated the need for tools that allow educators to capture insights from a child's independent reading when they cannot be together in person," said Amplify CEO Larry Berger. "By building speech recognition that actually works with children's speech, SoapBox Labs has opened the door to not just remote, but more equitable assessment by reducing the risk of longstanding issues like inter-rater reliability and bias."

To perform the assessment, students will read two texts aloud: one at the student's grade level and another adapted based on the student's performance on the first text. Students will then answer text-dependent comprehension questions. Data from the assessment, including reading rate and specific error types, provide educators with near real-time feedback, which they can use to identify areas of focus and cater lessons to each child's needs.

Extensive beta testing over a six-month period shows a high correlation of 0.96 between SoapBox Labs' assessments and human scoring, according to Krista Curran, Amplify's senior vice president and general manager of assessment and intervention.

And while the rate of agreement exceeds market standards, Curran said, it's important to remember the technology is a "supplement, not a replacement."

"There's not something this machine can do that a person cannot. This is not about the machine replacing a teacher. It is, instead, about getting educators the information they need in a way that supports them and saves them time," she said.

This is especially true for time-constrained teachers in the upper grades who still need to prioritize reading proficiency. "The technology saves teachers the time it would take to administer the assessment and listen to each student one-on-one on their own," she said. "It allows the teacher to, on their own time, review the data, including the errors the system flagged; grade the assessment like any other assignment; and steer interventions as they see fit."

- **Language learning:** Millions of adults have used popular speech recognition-enabled applications like Duolingo or Rosetta Stone to learn a foreign language. For this reason, it's not surprising that one of the earliest success stories for speech recognition for kids was

in language learning. Unlike adult-focused systems, kid-specific speech recognition can listen while a kid reads aloud and can immediately return pronunciation scores and encouraging feedback—just as a supportive adult or tutor would.

## CASE STUDY: LINGUMI

Studies have shown it's easier for preschool-age children to learn a foreign language. [Linguists](#) say that's because kids are hard-wired to acquire language, even referring to the time between infancy and puberty as the “critical period” for language development. Despite the research, parents of infants, toddlers, and preschoolers have had few options for language-learning tools outside of the school setting.

That changed in April 2019, when SoapBox Labs partnered with Lingumi to help hundreds of thousands of young children learn English. Lingumi, an English-language learning platform for children ages 2 through 6, learned about SoapBox Labs' speech recognition technology through an online search. Lingumi embedded that technology into the app, enabling it to “listen” as children complete speaking tasks. The app then assesses pronunciation and fluency, providing the same kind of immediate and accurate feedback an adult might give.

Using SoapBox Labs' technology not only improved the app's quality and consistency, but it transformed the way it reinforces learning in its child-users, said Lingumi CEO Toby Mather. The company's goal beyond teaching English to children is building confidence through positive reinforcement. Lingumi's automated pronunciation assessments, enabled by SoapBox Labs' technology, are a particularly helpful way of delivering that, according to Mather.

“When we plugged in the SoapBox [Application Programming Interface], it gave children a visual representation of that score very quickly,” he said. “We found that that really motivated them to improve and engage. It has been a great success.”

- **New Learning Experiences:** In every episode of children's television show *Daniel Tiger's Neighborhood*, the protagonist breaks the fourth wall to ask the viewer a question, which invariably results in a response.

But what if Daniel Tiger didn't just continue with the storyline, but actually listened to the child's input and responded back to them? As any parent or teacher knows, the mere act of having a conversation with a child offers a profound learning experience. And when

the conversation includes a beloved character and is powered by technology, it creates a whole new world of learning experiences that weren't possible before.

Take the experience of Sara DeWitt, vice president of PBS KIDS Digital. When she first started working at PBS in 1999, she and her colleagues were trying to help 4-year-old and 5-year-old children use a mouse and keyboard. It wasn't an easy user experience for a child, she said. Touch screens made

technology more transparent and gave kids the ability to navigate content more easily.

“Technology—when used properly—holds the potential to remove barriers between kids and their learning experiences,” DeWitt said, discussing the evolution of technology and learning over the past 20 years. Speech technology, however, allows them to harness a much more natural mechanism for engagement: their voices.

It’s a way to empower children by putting them in control of their own learning.

“Voice is putting us on the edge of that in terms of how kids can find our content more quickly,” DeWitt said. “A kid can use a voice-powered app or assistant to say, ‘I want to watch Daniel Tiger.’ That’s something we want to get into.”

Dewitt continued, “The technology is also creating a whole new category of educational and entertaining media for kids—part TV show, part game, part curriculum—that is interactive, authentic, instructive, and highly engaging, and more than the sum of its parts.”

Newsela’s Cogan-Drew discussed another way in which speech recognition could usher in a new era of authentic gamified learning

“

---

Technology—when used properly—holds the potential to remove barriers between kids and their learning experiences. - *Sara DeWitt, vice president of PBS KIDS Digital*

experiences. Despite much promise over the last two decades, the benefits of gamification have proven to be elusive. To Cogan-Drew, the problem comes down to the gap between the intent of the learning experience designer (i.e., to learn a particular skill) and the objective of the student (i.e., to earn points). However, this is a problem speech recognition can solve.

“By closing the daylight between an assessment and the intended learning behavior, speech recognition can make it impossible for a student to essentially ‘game the game,’” Cogan-Drew explained. “A student who ‘performs’ understanding of reading is actually reading, not just learning how to fool the machine into thinking they are reading.”

## THE VALUE OF VOICE DATA IN EDUCATION

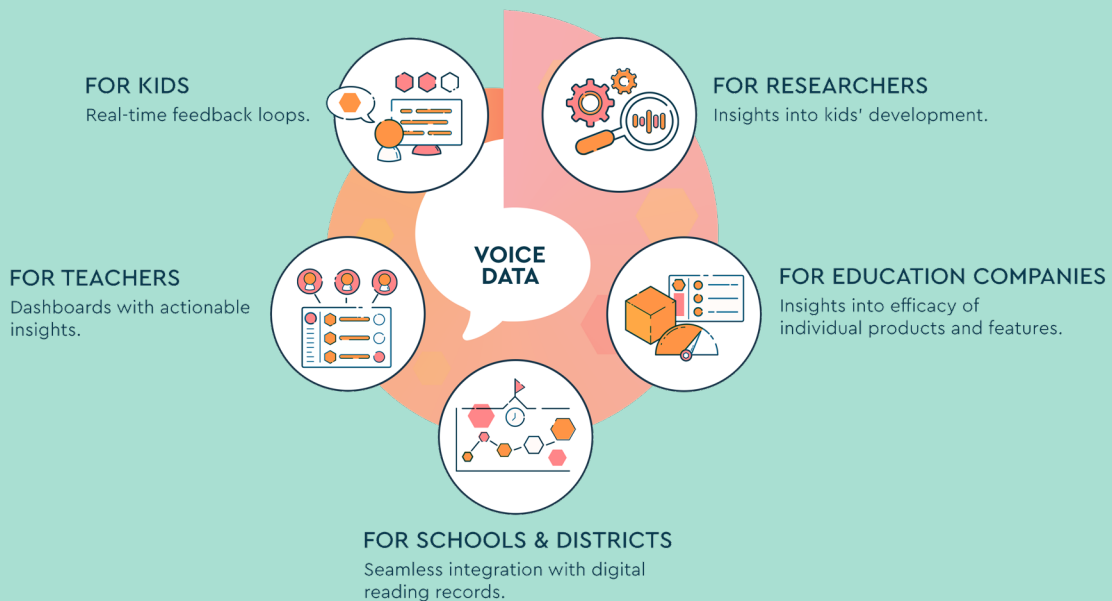
Whether it's the level of confidence in a spoken word, substituting or deleting certain words from sentences, or the length of time it takes to start and stop speaking, speech recognition generates unique data points.

And developers often want to know how the data generated by voice-enabled tools can improve learning outcomes for kids and their teachers.

“For the first time, using voice-enabled assessment tools, we can accurately measure the expressive language and literacy skills of a child. Historically, teachers needed to listen while the child read, annotate what they heard, and then input it into a system,” said SoapBox Labs CEO Dr. Martyn Farrows. “Voice-enabled solutions significantly reduce the time required for that manual process.”

Voice-enabled reading and language tools offer kids and teachers immediate feedback loops that provide intervention insights, support personalized learning paths, and improve learning outcomes.

What's more, Dr. Farrows added, “Over time, the data generated from speech recognition also provides longitudinal views for schools and districts and more informed product design decisions for edtech companies.”



## BUYER'S GUIDE

As speech recognition begins to emerge as an indispensable tool in the teaching kit, it's vital for the market—from those developing products to those purchasing them—to deepen their knowledge of the technology and, in particular, its applications with children in learning environments.

So what should education companies and publishers consider before adding speech recognition technologies to their tools and platforms for the classroom? Learn more in the SoapBox Labs [Buyer's Guide](#).

---

## CONCLUSION

---

Although speech recognition has become commonplace in myriad aspects of our lives, its application in education remains nascent. Within K-12 schools, speech recognition technology has the potential to drive equity, scale assessments, and enable children to interact with technology more independently. It is already being used to help people learn to read, learn math, and more confidently speak in new languages.

The technology offers capabilities for monitoring student performance and for providing intervening, corrective, and enriched learning.

“We need to be more thoughtful with our voice data, and even more so again when it comes to the voice data of our children,” Dr. Farrows wrote in a recent op-ed. “Voice is a much more important tool for children than it is for adults. For kids, voice is not just a convenience; it's a need, it's an opportunity.”

He continued, “Voice gives kids agency and allows them to learn and play in the most natural way possible—using their voices.”