



Data Science Is Vital to Student Success. **So Why Are Outcomes Going Down?**

Zarek Drozda
Data Science 4 Everyone

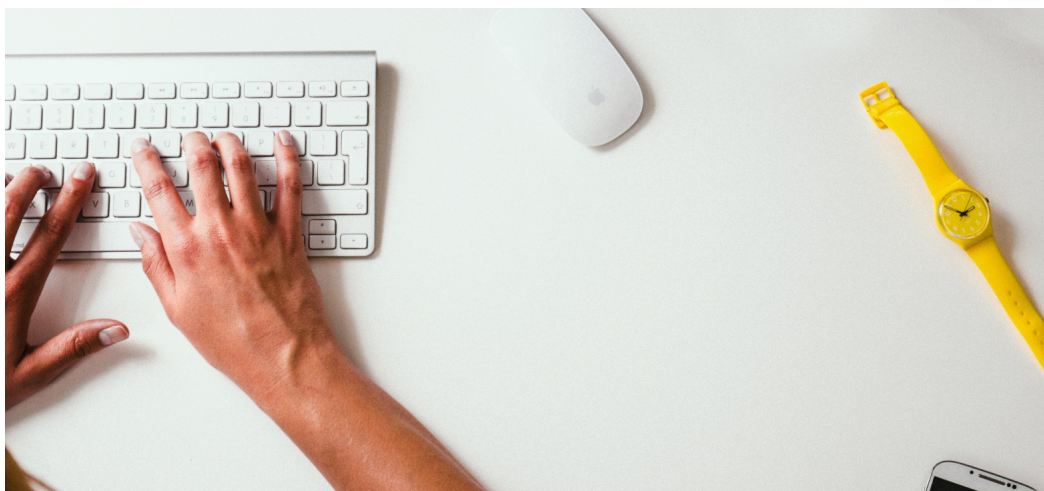
It is nearly impossible for anyone to go about their daily routine without encountering the impact of data science.

The tech boom of the last 30 years gave way to a data boom that continues to rapidly redefine the skills that tomorrow's generation needs. Indeed, the advent of computers, tablets, and smartphones has created an exponential increase in information and thrust data science and data literacy into the spotlight as essential 21st-century skills.

People in all sectors of society – from science labs, healthcare, and law enforcement, to business, retail, and politics – are tasked with understanding and interpreting data each day in order to make important decisions for society and for their success at work and in life.

In the last decade, data has become a new frontier that is ripe for exploration and analysis. And in order to live up to America's ideals, its spirit of innovation, and its push toward equality, today's students must be empowered with the data literacy skills they need to succeed. Yet when it comes to this particular challenge, our education system is still rooted in the past – when computers did not exist and data analysis and insights were calculated by hand.

Worryingly, despite the fundamental need for data science education, the latest assessment data shows that declines in student performance in data literacy skills are outpacing declines in other content areas.



Findings

In an effort to provide a better look at the status of data science education and implementation in K-12 schools, the [Data Science 4 Everyone](#) (DS4E) coalition recently examined results from the [National Assessment of Educational Progress](#) (NAEP), also known as the “nation’s report card” on schools. Special attention was paid to student data literacy performance and teachers’ self-reported efforts around data-related instruction.

The most recent NAEP results indicate that student achievement in data literacy skill sets declined significantly compared to other content areas. Between 2019 and 2022, scores in Data Analysis, Statistics, and Probability fell by a full 10 points for 8th-grade students, representing what [some experts consider a full grade level](#) in progress.

Results for high school (Grade 12) will be released in 2024. Prior course-taking patterns in high school would suggest little optimism for learning recovery in data literacy. In 2019, the NAEP [High School Transcript Study](#) (HSTS) revealed that only 17% of students completed a course in statistics or probability. Additionally, less than a quarter of students have completed any computer programming (which itself is broader than computer science), and less than one out of every ten high school students completed AP Statistics, according to NAEP student questionnaires. [Supplementary analysis](#) of the HSTS study revealed that less than 1% of students completed a dedicated course in modern data science or applied data analytics in high school. Historically low prioritization of this content likely continued during the pandemic, meaning students are at risk of graduating with record-low proficiency in data, statistics, and probability – let alone a near-total lack of opportunity in modern data science techniques.

However, these short-term declines only represent an extension of broader declines in student data literacy.

Given the broader technological acceleration in “big data,” artificial intelligence, automation of work, and blockchain, these data sound an alarm for an educational crisis that has been unfolding in slow motion.

Data literacy outcomes are down over the long term.

While the largest declines in data literacy outcomes occurred during the pandemic, there’s been a much longer-term decline in outcomes starting well before COVID. Over the last decade, scores in Data Analysis, Statistics, and Probability on the NAEP mathematics exam dropped 10 points for fourth-grade students and 17 points for eighth-grade students.

Data science vs. data literacy

It is important to note that NAEP measures partially assess components of data science and data literacy. We use both of these terms throughout the report, where data literacy refers to the basic *skills* and *habits* that allow us to read, work with, analyze and communicate with data effectively, and data science refers to the *process and modern techniques* of drawing insights from and exploring trends in data and using data to solve problems. This is an important distinction because data literacy skills and habits are foundational to a data science education. The NAEP content category, “Data Analysis, Statistics, & Probability,” captures some basic components of data literacy, in addition to a heavier emphasis on probability scenarios. This report recommends NAEP should expand, modernize, and clarify these areas within the exam, reflecting significant technology and practice changes that are not yet present in the test.

FIGURE 1: Change in average scores between 2019 and 2022 for fourth-grade and eighth-grade students in NAEP mathematics, by mathematics content area.

Grade 4

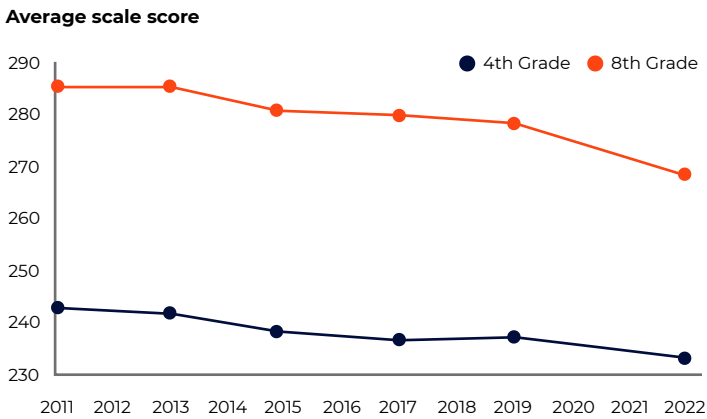
Content area	2022 score	2019 score	Change
Number properties and operations	239	243	5 ↓
Measurement	235	239	5 ↓
Geometry	230	234	4 ↓
Data analysis, statistics, and probability	233	237	4 ↓
Algebra	239	243	4 ↓

Grade 8

Content area	2022 score	2019 score	Change
Number properties and operations	271	279	7 ↓
Measurement	272	280	9 ↓
Geometry	273	278	6 ↓
Data analysis, statistics, and probability	269	279	10 ↓
Algebra	281	289	8 ↓

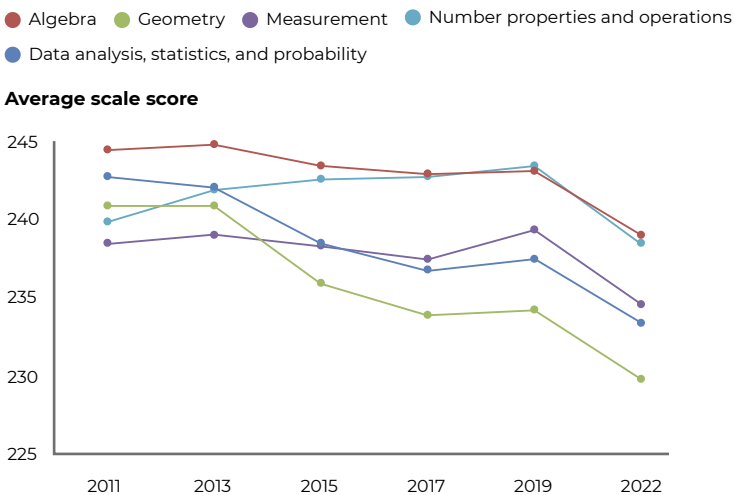
This drop in exam scores is significant. Most experts [believe](#) 10 points is the equivalent of a single grade level, or a year's worth of academic growth. Put differently, today's fourth graders have the data literacy of third graders from a decade ago, and today's eighth graders nearly have the data literacy of sixth graders from a decade ago. What's more, this decline in data literacy is occurring at a faster rate than other math content areas in the NAEP assessment.

FIGURE 2: NAEP grade 4 and grade 8 performance in Data Analysis, Statistics, and Probability, 2011-2022



Furthermore, data literacy has declined even more rapidly for some historically underrepresented students. For instance, eighth-grade Black students scored 18 points lower in Data Analysis, Statistics, and Probability compared to the 2011 exam.

FIGURE 3: NAEP math grade 4 performance, 2011 - 2022, by content area



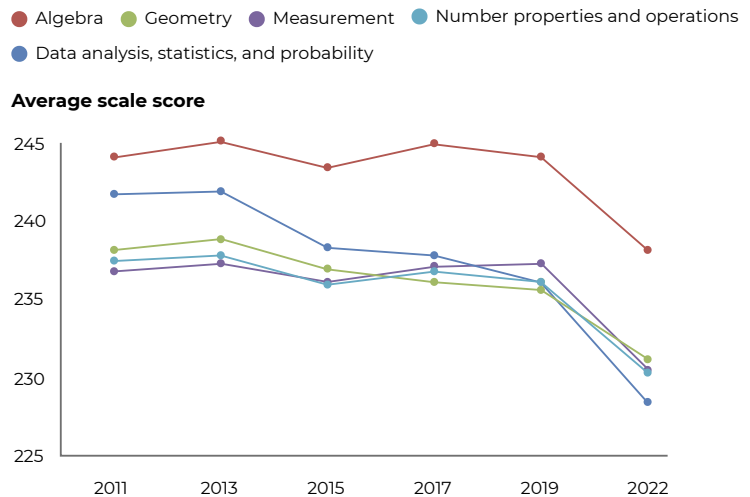
Low-income students are also facing a loss in data literacy. On the 2011 fourth-grade math exam, scores for low-income students dropped 10 points, meaning low-income fourth graders are currently performing at the level of low-income third graders from a decade ago. Among eighth-grade students, their data literacy score dropped 17 points, a loss of nearly two years' worth of learning.

Several states have also seen a particularly large decline in data literacy outcomes. From 2011 to 2022,

eighth-grade scores in Data Analysis, Statistics, and Probability dropped:

- The equivalent of two and a half grade levels in **Texas**, from 292 to 267.
- The equivalent of nearly three grade levels in **Colorado**, from 297 to 270.
- The equivalent of three grade levels in **Maryland**, from 296 to 263.

FIGURE 4: NAEP math grade 8 performance, 2011 - 2022, by content area



There are large racial gaps in student outcomes when it comes to data literacy. These gaps exist across all grade levels but are at times most egregious at the middle and high school levels. For instance, fourth-grade Black students scored 28 points lower than their white peers in the Data Analysis, Statistics, and Probability section of the NAEP math exam. Put differently, it would take an additional three years of academic progress for fourth-grade Black students to perform at the level of their white peers in data literacy. Eighth-grade Black students scored 34 points lower than their white peers, or are over three years behind. As another example, eighth-grade Hispanic/Latino students scored 42 points lower than their Asian peers in eighth grade, the equivalent of four grade levels.

FIGURE 5: NAEP grade 8 performance in Data Analysis, Statistics, and Probability, 2011-2022, by student race/ethnicity

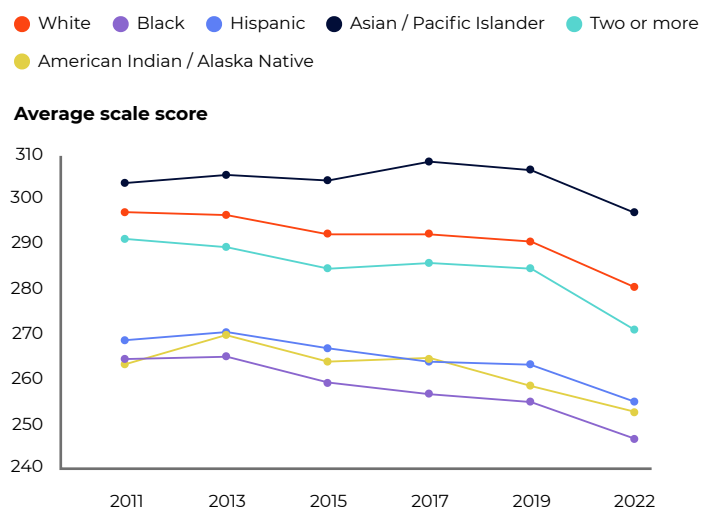
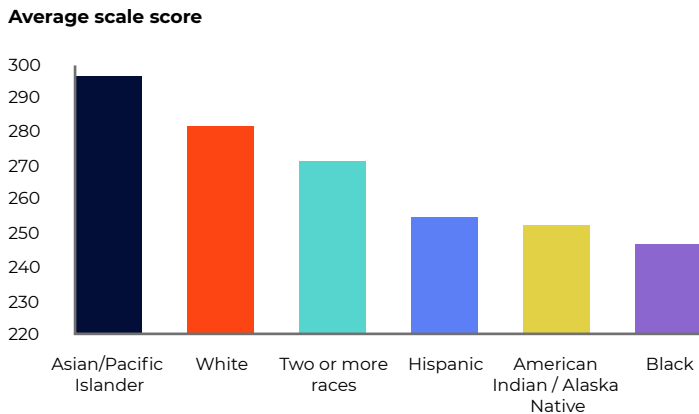


FIGURE 6: NAEP grade 8 performance in Data Analysis, Statistics, and Probability, by student race/ethnicity



These gaps appear along socio-economic lines as well. On the latest fourth-grade exam, low-income students scored 23 points lower than their higher-income peers in Data Analysis, Statistics, and Probability. In eighth grade, low-income students scored 29 points lower than their higher-income peers.

Student performance also differs significantly at the state level. For instance, eighth-grade students in Washington D.C. are nearly three years behind their peers in Minnesota. Washington D.C. scored the lowest on the 2022 eighth-grade exam in Data Analysis, Statistics, and Probability with an average score of 251. Students in Minnesota scored the highest state average at 279, a 28-point score difference that is the equivalent of nearly three grade levels.

A Possible Cause? Data-Related Instruction Has Declined in Recent Years

Generally, math teachers report a lower emphasis on data literacy compared to the previous NAEP results. The national share of eighth-grade math teachers reporting “moderate” or “heavy” emphasis on data analysis dropped two percentage points from 83% to 81% between 2019 and 2022, and for fourth-grade math teachers it dropped five percentage points from 81% to 76%.

Some states experienced a particularly sharp decline. For instance, between 2019 and 2022, the share of fourth-grade math teachers reported “moderate” or “heavy” emphasis on data analysis:

- Dropped 11 percentage points in **Maryland**, from 74% to 63%.
- Dropped 12 percentage points in **Iowa** from 70% in 2019 to 58%.

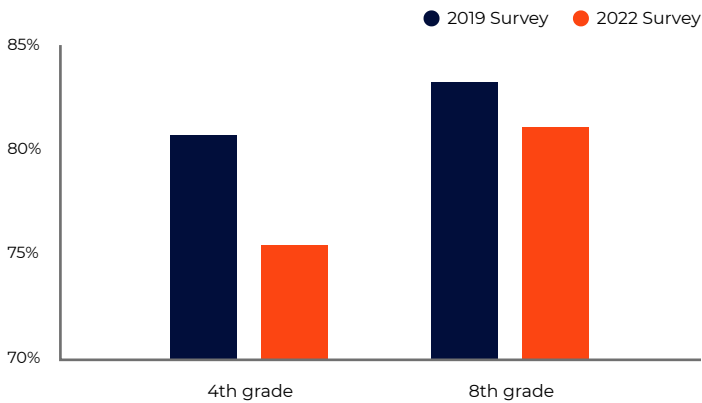
Other states with reported declines in fourth-grade data-related instruction larger than five percentage points include Alaska, Arizona, Arkansas, California, Colorado, District of Columbia, Georgia, Hawaii, Idaho, Louisiana, Maine, Maryland, Massachusetts, Minnesota, Mississippi, Nebraska, New Jersey, South Dakota, Vermont, Virginia, West Virginia, Wisconsin, and Wyoming.

While middle school math teachers generally report greater emphasis on data-related instruction than elementary teachers on an absolute basis, there were notable declines over time for the middle school level as well. For instance, in **Massachusetts**, the share of eighth-grade math teachers reporting “moderate” or “heavy” emphasis on data analysis dropped 14 percentage points from 85% in 2019 to 71% in 2022.

It is important to remember that self-reported data come with limitations. For instance, what qualifies as “moderate” or “heavy” emphasis on data analysis can vary from state to state, or even year to year. The content underlying “data analysis” will also vary drastically given the quickly changing landscape in data education – today, it could mean anything from basic “mean, median, mode” to the interpretation of machine-learning algorithms. That is why the NAEP exam needs to [better incorporate modern terms and raise the bar](#) for proficiency in data science, given the complex technological metamorphosis taking root in society today.

Nonetheless, it is concerning that **every state except Alabama reported a decline or stagnant trend in data-related instruction**. In Alabama specifically, the share of teachers reporting “moderate” or “heavy” emphasis on data analysis jumped 7 percentage points for eighth-grade teachers from 86% to 93%, and it increased one percentage point for fourth-grade teachers from 86% to 87%. That increase is likely in part due to Alabama’s new mathematics standards released in 2019, which open with the following line from NCTM’s Catalyzing Change report: “Today, mathematics is at the heart of most innovations in the ‘information economy,’ which is increasingly driven by data.” The new standards revamp the traditional mathematics sequence, creating 1) “Geometry with Data Analysis,” 2) “Algebra I with Probability,” and 3) “Algebra II with Statistics” courses that emphasize data in their names and infuse technology, mathematical modeling, and a greater focus on statistics throughout.

FIGURE 7: Change in share of fourth- and eighth-grade math teachers reporting emphasis in data analysis between 2019 and 2022



Another Possible Cause? Technology Instruction in Classrooms Trails Access.

Access to technology – including laptop programs, better internet access, and classroom learning tools – has increased significantly over the past decade. The instructional use of that technology, however, may be creating headwinds for students to learn about modern data theory and practice. In short, we *purchase* technology in education frequently, but barely focus on *training students* to leverage current or future technology effectively.

Accompanying NAEP achievement scores, participating schools and districts also administer a rich set of questionnaires that survey demographics, school conditions, classroom practices, course access, and even student attitudes.

The below frequency table shows data from questions structured on a spectrum from “Never or hardly ever” to “Every day or almost every day.” Sadly, the most recent data indicates that the utilization of modern data analysis tools is far too low for their frequency in modern workplaces.

FIGURE 8: Students’ reported use of data analysis tools in math classrooms ¹

I never or hardly ever:	Grade 8	Grade 12	Year (most recent data)
Use statistical programs for math class	80%	82%	2015
Use spreadsheets for math assignments	72%	83%	2015
Use database programs	No data	86%	2015
Use computers to research a math topic	82%	No data	2015

¹The wording of the NAEP survey questionnaire items in this table is slightly adapted to reflect math classroom emphasis.

Unfortunately, NAEP has not released most of this data publicly for more recent years.

We have seen recent improvement in some categories, though it is hard to understand how these responses are affected by using technology to deliver content and homework assignments (remote learning) vs. using technology to teach about technology (i.e. training students in data science or A.I. analysis methods and tools). If anything, these results indicate that both students and teachers are comfortable using computer technology and do so regularly, but may not be *linking content and technology together* in instructional goals.

In particular, students use computers frequently for help with assignments – 28% do so daily, and 73% at least once a month or higher. Results in the table below suggest most students are using computers frequently. Yet they do not appear to use their devices beyond Googling.

“Data exploded over 20 years ago. The thing is, we’re just now catching up to it in education.”

–Deb Crawford, Supervisor for Mathematics & World Language with Frederick County Public Schools in Virginia

FIGURE 9: Reported use of computers and digital devices in math classrooms ²

I never or hardly ever:

	Grade 8	Grade 12	Year (most recent data)
Use computers or digital devices to teach math	9%	18%	2022/2019
Use computers or digital devices for math at school	10%	No data	2017
Use computers or digital devices for math assignments	15%	No data	2017
Use computers or digital devices for online resources to help with math assignments	12%	18%	2019

²The survey items may occur on different Likert scales, but the wording of the most extreme survey response is standardized for consistency.

Students' Motivation for Math Declines in High School

Students not only fall further behind in data literacy performance and expectations as they progress through grade levels, but they also experience declining interest in math more generally – especially once they experience high school mathematics. For instance, 73% of eighth graders believe “it is important to do well in math” compared to 58% of 12th graders. This disconnect between what students learn about math and what they find meaningful in high school may be attributed to an outdated math curriculum that has not kept pace with the modern information age.

For instance, less than half of high school seniors strongly agree that “math will help them in the future” or “math will help them even when not in school.” Yet, data literacy is necessary for everyday activities and decision-making, from evaluating data and claims in news reports to weighing the risks and benefits of personal choices. However, few districts are emphasizing data literacy as an important 21st-century skill and integrating it into math curricula or other subject areas. Thus, it shouldn't be surprising that students' dislike for math rises as they

progress through school and struggle to see the application of math to the real world.

Teachers receive general training in technology, but not for future skill sets in content areas.

Another concern is the lack of robust training for teachers on how to meaningfully incorporate data literacy into instruction. For instance, while 89% of eighth-grade teachers report proficiency or recent training in integrating computers into instruction, less than half of these teachers report learning about how to use a computer or digital device to teach math.

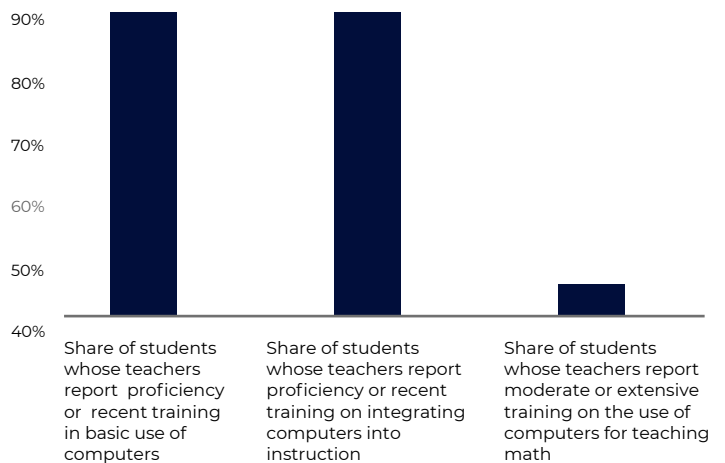
Put differently, most middle school teachers receive basic training on computer use but lack specific training on how to modernize math instruction by leveraging digital and computer technology. This gap limits opportunities for students to engage with math problems through modern spreadsheet tools and computer database programs and reinforces traditional methods of calculator, paper, and pencil.

FIGURE 10: Change in students' affective disposition towards math between eighth grade and twelfth grade ³

I take math because:	Grade 8 (strongly agree or agree)	Grade 12 (strongly agree or agree)	Year (most recent data)
Math meets graduation requirements	Not Asked	90%	2015
Math will help me in the future	Not Asked	45%	2019
Math will help me when not in school	55%	45% ↓	2019
It is important to do well in math	73%	58% ↓	2019
Math is a favorite subject	47%	44% ↓	2015
I enjoy doing math	40%	32% ↓	2019
I'm interested in what I learn in math	40%	31% ↓	2019
I look forward to my math class	36%	23% ↓	2019

³“Agree or strongly agree” was the survey response for most questions. Otherwise, students answered an equivalent phrase of “quite a bit like me” or “exactly like me.”

FIGURE 11: Middle school teachers' reported experiences in basic computer training vs. math-related computer training



I have taken training or am already proficient in:

	Grade 8	Year (most recent data)
Basic computer use in past two years	89%	2022
Integrating computers into instruction	86%	2022
Use of computers for teaching math	13%	2015

Background

Data science skills are becoming increasingly critical – not just for experts and specialists, but for the general public as well. They will be key to building back after a number of global challenges ranging from the COVID-19 pandemic to economic disruptors both nationally and on a global scale.

“Enhancing data science education will have significant benefits for our nation’s citizenry and economy. Like reading, writing, and arithmetic, basic data literacy is vital for informed democratic participation. We also see employers increasingly demanding data science expertise,” wrote 10 U.S. Senators, including [Sen. Alex Padilla](#) (D-Calif.), in a [2021 letter from Congress](#) to the directors of the Institute of Education Sciences and the National Science Foundation. “According to the Bureau of Labor Statistics, employment of computer and information research scientists is expected to grow 15% between 2019 to 2029, far exceeding the average growth rate of other occupations.”

Employers know the value and vitalness of data science in their future growth.

A 2019 McKinsey Analytics [report](#) stated that 47% of organizations believe that data analytics has fundamentally or significantly transformed how their industries compete.

In fact, in order to be a successful company in the 21st-century, businesses have to use data to their advantage. Therefore, the demand for data literacy skills is poised to remain high as companies increasingly prioritize data-driven decision-making.

Given these circumstances and changes in the labor market, shifts in the nation’s K-12 curriculum must be evaluated as appropriate next steps that are necessary to prepare young students for future roles and careers in a transitioning economy where each day is more data dependent.

“Enhancing data science education will have significant benefits for our nation’s citizenry and economy. Like reading, writing, and arithmetic, basic data literacy is vital for informed democratic participation.”

–10 U.S. Senators, including Sen. Alex Padilla (D-Calif.), in a 2021 letter from Congress to the directors of the Institute of Education Sciences and the National Science Foundation.



A Once-in-a-Century Opportunity

Data science is not a solution for systemic inequities in the K-12 education system. However, in some states, the achievement gaps when it comes to data analysis are particularly large. In Virginia, low-income fourth- and eighth-grade students are three years behind their higher-income peers in Data Analysis, Statistics, and Probability.

What's more, data science education can help increase access to meaningful opportunities and ultimately better jobs. This can be done by supporting all students and encouraging them to develop the skills that will help them meet the needs of the modern economy and to engage in their communities.

Ensuring that all students graduate high school with a basic command of data is crucial, especially for students of color and students in poverty. In order to do this, schools need to update their standards, courses, and curricula, and commit to adopting data science education.

In short, society's most vulnerable students are being left behind in the advance toward improved data literacy and data science skills for every student, which is becoming increasingly linked to future success.

There is enormous potential to change this, however.

"Adults are the ones who have the power to help every kid see themselves as a mathematician. I think data science has a real opportunity to re-engage kids, to make them active learners, and get them excited about math," said Lindsey Henderson, Secondary Mathematics Specialist with the Utah State Board of Education. "It's about changing the ecosystem for STEM to make diversity possible."

The good news is that there is still time and opportunity to capture these students in the push forward with data science education — if we move quickly.

How Data Science Education Outcomes Can Be Improved

While the data does not explain why there's such variation in outcomes and teacher practices, a few things are clear: building a strong path for data science now to fill emerging and evolving jobs in the workforce requires initiative, innovation, collaboration, and a firm commitment to young learners.

In some cases, it is the young learners themselves seeking and embracing changes in the curricula aimed at preparing them for life after school.

"I was taking a bunch of calculus classes and I wasn't having as much fun," said Emilio Jaime, a former Phineas Banning High School student in Los Angeles who went on to study computer science and later media studies at the University of California, Berkeley. He was introduced to data science in his senior year. "It was totally different from the math classes I'd taken before."

The appeal of it was a more hands-on, and real-world applicable experience with coding, statistics, and analysis. "It was a really cool collaborative learning experience. It felt like no one was wrong as long as you could support your answer with the data. It wasn't just $2+2=4$," he said.

As a result: "It made some of us more confident," Jaime said, noting his high school was not as economically advantaged as some and that, while he was tracked into higher-level math courses early on, his data science class welcomed peers without such a strong core in mathematics. Like mathematician Steven Strogatz said [in an interview](#) with *Freakonomics* co-author Steve Levitt,

"You can't go wrong by being more data savvy."

Jaime graduated from UC Berkeley with 5 data science courses to his name, a passion for media technology, and is now working with Dreamworks, the animation company. When students find reasonable relevance and a clear motivation to learn challenging content, they will succeed relative to when justifications are abstract or lacking.

Even more important, however, is supporting educators. Many educators say America must invest in a strong future by prioritizing data science education in ways that foster data literacy in all K-12 students, regardless of their pathway to achieving these skills. This can take many forms and be done in many locations:

- By making data science courses and content relevant within math curricula
- By building data literacy connections within subjects as diverse as social studies or English
- By creating career & technical education course options

There are many implementation models through which individual schools, school districts, or entire states can integrate data science into K-12 education. Several can be accessed via the DS4E website [here](#). To be sure, much more investment is needed to pilot different models and better understand what works when it comes to data science education.

Conclusion

Focus on the future is key when it comes to promoting, supporting, and adopting efforts to move fundamental data literacy skills and comprehensive data science education forward.

Interpreting and working with data is now central to even the most commonplace activities. The demand for these skills only grows as we are forced to evaluate the information we consume, determine the quality and reliability of data used to justify claims, and weigh the risks and benefits of personal choices.

Schools that equip students with valuable data science skills prepare them to become leaders who will transform our nation's potential – in technology, business, thought leadership, and across sectors – far into the future.

Furthermore, equitable access to data science education presents an opportunity to open doors to higher education, higher-paying careers, and a more engaged community for countless children across diverse communities. The result is a well-rounded, highly-effective, and robust workforce capable of meeting any challenge.

The strength of our society today is built on decisions people in past generations made about what was important to learn for the future. Similarly, the strength of our innovations and advancements a generation or two from now will be a reflection of the priorities we set today.

Methodology

[All NAEP data is from 2022, 2019, 2017, 2015, and 2011.](#) The data was downloaded in October and November of 2022. All survey data is self-reported by teacher respondents, which may account for certain variances. The self-report data from 2022 was taken specifically from the teacher survey questionnaire for the fourth- and eighth-grade math exams. However, some data was not available at the state level across both grades due to sampling issues.

Survey responses used a Likert scale (e.g., “Little or No Emphasis, Moderate Emphasis, Heavy Emphasis”). This report combined the top Likert scale items (e.g., “moderate” and “heavy”) and summed the two data points together.

We analyzed student performance on the NAEP mathematics assessment in fourth and eighth grade. NAEP math scores range from 0-500 for grades four and eight. The math exam assesses students in five broad areas of mathematical content, including the content area of **Data Analysis, Statistics, and Probability**. This content area measures students' understanding of data representation, characteristics of data sets, experiments and samples, and probability.

“We need more students learning demanding math and science skills. Without that foundation, we will never have a large, diverse, and strong STEM workforce—a precondition for the U.S. economy to prosper. Simply telling students who have not truly mastered STEM skills that they are ‘A students’ who have finished a rigorous math and science curriculum is not the way to produce that workforce.”

–Mark Schneider, director of the U.S. Department of Education's Institute of Education Sciences (IES).



About the Author

Zarek Drozda is the Director of Data Science 4 Everyone. Zarek helped launch DS4E in 2019, co-organizing a coalition of now 1000+ education leaders to advance data science and data literacy education in K-12 schools. He also served as a Data Science Fellow for the U.S. Department of Education's Institute of Education Sciences (IES), where he led research on data science, artificial intelligence, blockchain, and other emerging technology education. Prior to Federal service, Zarek helped build a social impact incubator at the University of Chicago (the Center for RISC) with economist and *Freakonomics* co-author Steven Levitt. Zarek earned a Bachelor's degree in Economics from the University of Chicago, and loves using data to tackle complex social problems.

You can reach Zarek at zarekd20@uchicago.edu.

