Transforming Curriculum and Building Capacity in K–12 Data Science Education

Travis Weiland\textsuperscript{1} Christopher Engledowl\textsuperscript{2}

\textsuperscript{1}University of Houston, Houston, Texas, United States of America, \textsuperscript{2}Independent Researcher

**Published on:** Oct 27, 2022

**DOI:** https://doi.org/10.1162/99608f92.7fea779a

**License:** Creative Commons Attribution 4.0 International License (CC-BY 4.0)
ABSTRACT
The recently released and updated *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education* (GAISE II; Bargagliotti et al., 2020) provides guidance as to how teachers can support the development of data literacy for all students in the pre-K–12 curriculum. However, to truly meet the vision of the GAISE II report and to support all students in developing data literacy for today’s societies, significant transformations need to be made to the educational system as a whole to build capacity for such development. In this article we discuss the current state of the K–12 curriculum focusing on the mathematics curriculum where statistics and data concepts are most frequently situated, presenting some challenges and exciting examples. We then discuss areas of need for capacity building that must come at all levels, including: K–12 school curriculum, K–12 teacher professional development, K–12 teacher preparation, statistics and data science education research, and policies. We also provide a set of recommendations for building capacity to develop the data literacy of all students through the teaching of data science and statistics concepts and practices in the K–12 mathematics curriculum to support democratic equity through engaged citizenship.

**Keywords:** data science education, data literacy, statistics education, teacher education, doctoral student education, educational policy, curriculum standards

Media Summary
The recently released and updated *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education* (GAISE II; Bargagliotti et al., 2020) supports the development of data literacy for all students in the pre-K–12 curriculum. To truly meet the vision of the GAISE II report, and to support all students in developing data literacy for today’s societies, significant transformations need to be made to the educational system as a whole to build capacity for such development. We discuss areas of need for capacity building that must come at all levels, including K–12 school curriculum, K–12 teacher professional development, K–12 teacher preparation, statistics and data science education research, and policies. We provide a list of recommendations for curriculum transformation as well as capacity building that needs to occur to support such curricular transformation.

*Watch the video below for a summary of this article.*

Visit the web version of this article to view interactive content.
1. Introduction

At a global level, data are being constantly and instantly collected en mass, often under the guise of benefiting society. However, such data are also collected to feed into algorithms that are used to increase profits of corporations or weaponized against people—particularly those from historically marginalized groups (O’Neil, 2016). This comes after decades of scholars calling for the need for quantitative and statistical literacies for society (Ben-Zvi & Garfield, 2004; Franklin et al., 2007; Gal, 2002; Steen, 2001; Wallman, 1993). With the rapid expansion and development of information technologies, cloud-based storage and computing, and machine learning and artificial intelligence (AI), the need for data-centric literacies has increased exponentially. In many ways, data science and literacies such as quantitative, statistical, and data literacy are all deeply connected when considering the literacies crucial for citizens in society today. It has long been the primary goal of K–12 public education to foster citizenship in democratic societies (Labaree, 1997). As a result, it is becoming increasingly important for data-centric literacies to be included in the curriculum of K–12 public schools, including concepts and practices from data science.

Data science as a field is still ill-defined partially due to its origins in academia to create new spaces (Irizarry, 2020). One definition that has been offered by Wing (2019) in this journal is “Data science is the study of extracting value from data” (para. 1). In extracting value from data, we often see data scientists and academic programs hailing from statistics or computer science backgrounds, marrying theory and methods from statistics for analyzing and interpreting data with computability and programming from computer science. A prime example of data science in the high school curriculum is through the Introduction to Data Science (IDS) curriculum developed in the Mobilize Project carried out in California, led by Rob Gould at the University of California, Los Angeles. Gould’s (2017) project team created and implemented the high school introduction to data science curriculum in 15 school districts, 45 schools, and with over 9,500 students, according to their project website (https://www.introdatascience.org/). The project involved student programming, collecting data through mobile apps, and using statistical concepts and practices to analyze and interpret the data. The Mobilize Project has been an impetus for California to allow the incorporation of data science concepts and practices into the mathematics curriculum for high school students and allowing a course in data science to count as a high school mathematics credit.

The location of data science in the mathematics curriculum is often through the heavy overlap of statistics and data science concepts and practices in the K–12 curriculum. For example, much of the statistics content and practices in the K–12 curriculum focuses on exploratory data analysis (EDA; Tukey, 1977; see Scheaffer & Jacobbe, 2014, for historical perspective). The relationship between statistics and data science in the K–12 curriculum has been recently solidified by the American Statistical Association (ASA) with the support of the National Council for Teachers of Mathematics (NCTM) in their updated Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education (GAISE II; Bargagliotti et al., 2020) by notably adding data science as an important area of study. Though this is a significant gain for data science in the curriculum, the GAISE II
Developing data science within the mathematics curriculum has its advantages. The statistics education community has fought a long and hard battle to gain ground (Scheaffer & Jacobbe, 2014) to create what Scheaffer (2006) refers to as a “happy marriage between mathematics and statistics.” Though data science has made inroads in the computer science curriculum, the unfortunate reality is that most students in K–12 schools still do not have any experiences with computer science in their school curriculum. However, all states require mathematics to be taught every year until high school, with at least three mathematics courses—most states require four—required during high school in order to gain a diploma. Furthermore, educational researchers working at the boundaries of mathematics and statistics education (Groth, 2015) have built a strong literature base to work from on the teaching and learning of data and statistics (Langrall et al., 2017; Shaughnessy, 2007). There is great potential for data science to be incorporated into the mathematics curriculum (Franklin & Bargagliotti, 2020) and as early as elementary school (Martinez & LaLonde, 2020). All these factors make statistics and statistical/data literacy a fruitful pathway for including concepts and practices from data sciences.
in the K–12 mathematics curriculum. We would like to note here that there certainly is potential for data
science in the computer science curriculum in schools. For example, the Bootstrap curriculum (see
https://bootstrapworld.org/) integrates computer science and data science and has been implemented in a large
number of districts in the United States. Furthermore, its impact on student learning has been researched (see
https://bootstrapworld.org/impact) and is overall positive. Additionally, there is potential for development in
the science curriculum as well. However, we focus on the mathematics curriculum because it pragmatically
holds more promise for impacting all students’ learning experiences.

To make the mathematics/statistics/data science relationship work, and to successfully support all children in
developing statistical and data literacies, there are some serious changes needed to curriculum, policy, teacher
preparation, and educational research. In this article, we start by discussing the current state of curriculum and
policies around data science in the mathematics curriculum and possible fruitful action steps and directions to
overcome some of the challenges. Then, we shift our discussion to the current infrastructure for supporting data
science education in K–12 schools (or lack thereof) and discuss necessary next steps for the field to build the
capacity for the meaningful incorporation of data science concepts and practices in the K–12 education system.

2. Curriculum and Policy Background

To help frame our discussion, we are drawing upon Remillard and Heck’s (2014) empirically derived
framework for the curriculum enactment process (see p. 709 for a visual model). Our goals in education
systems are often focused on student outcomes. However, education systems are complex and there are many
factors that contribute to a) the curriculum that students directly experience and b) the outcomes the curriculum
influences. The curriculum that students experience in the classroom is generally referred to as the enacted
curriculum, which is most directly impacted by the intended curriculum classroom teachers plan for their
students, and the instructional materials (i.e., textbooks, worksheets, digital tools, etc.) used in the classroom
(Remillard & Heck, 2014; Stein et al., 2007). There are many other contributing factors as well, such as
teachers’ knowledge of the content, pedagogy, and pedagogical content (Ball et al., 2008; Hill et al., 2008;
Mishra & Koehler, 2006), teachers’ beliefs and attitudes (Philipp, 2007), teachers’ identities (Aguirre et al.,
2013; Langer-Osuna & Esmonde, 2017), and teachers’ perceived obligations (Chazan et al., 2016), just to name
a few. There are also policy-level factors often centered in what is called the official curriculum, which starts
with a set of curricular aims and objectives that often come in the form of curriculum standards. These then
influence the official designated curriculum and the content of consequential assessments like high-stakes
standardized tests that increasingly drive what is taught in classrooms in the United States (Au, 2007;
Remillard & Heck, 2014; Wilson, 2007). We lay this out to say that changing the school curriculum is a very
challenging, multipronged process because our education systems are exceedingly complex and, in the context
of the United States, also very decentralized (Schmidt & McKnight, 2012). In other words, there is no golden
lever of change to pull that will magically get data science and statistics into the curriculum in meaningful
ways everywhere. To say it simply: we have a lot of work to do.
Starting with a look at official curriculum, the Common Core State Standards in Mathematics (CCSSM; National Governors Association Center for Best Practices [NGA Center] & Council of Chief State School Officers [CCSSO], 2010) has had a significant influence on the K–12 mathematics curriculum in the United States since 2009. It calls for a significant increase in the scope and rigor of the statistics taught in grades 6–12 mathematics, which has created an entryway for more statistics, and in turn, data science to be taught in secondary schools. At the same time, the amount of statistics concepts called for in the elementary grade levels (K–5) was significantly reduced, particularly around probability content, which has been lamented even by some who were involved in the CCSSM standards writing process (Confrey, 2010). This shift has left some to wonder whether students will be prepared to engage in the increased depth and breadth of the middle grades’ statistics and probability standards. Since 2009 there have been political shifts that have led to policy shifts, which have collectively led to states abandoning the CCSSM standards in favor of creating their own standards (Orrill, 2016). Additionally, many states conduct a standards review process every 10 years, so many states have done so since initially adopting the CCSSM.

At the same time, a handful of states never adopted the CCSSM, meaning there is variability in the opportunities students have to learn statistics. For example, Weiland and Sundrani (2022) found that as of 2021, only 21 states followed the CCSSM statistics standards to fidelity in their K–8 standards documents. Although Weiland and Sundrani reported that many states maintained very similar wording to the CCSSM, this change still represents a significant amount of variability in the statistics that students have opportunities to learn. Furthermore, in the experience of the authors of the present article, many high school teachers often report to us they are not teaching many of the statistics standards in their curriculum because those standards are often not assessed in standardized assessment, and as the saying goes, ‘what is tested is what gets taught’—a point that is also reflected in Remillard and Heck’s (2014) framework of the factors that influence the enacted curriculum.

This trend has only become more common and more deeply entrenched as many states have moved toward putting significant weight on the results of standardized assessments (Wilson, 2007)—a trend no doubt heavily influenced by federal policies such as the No Child Left Behind Act of 2001 (H.R.1 - 107th Congress, 2002) and Race to the Top Act of 2011 (H.R.1532 - 112th Congress, 2011). For example, a number of states have begun making personnel decisions based upon the use of value-added models (VAMs), which have been highly criticized for their questionable validity evidence (American Statistical Association, 2014; Goldhaber, 2015; Goldring et al., 2015), thus leaving teachers feeling increased pressure to have their students perform well on the assessments, which can have negative impacts on the enacted curriculum. For instance, only about 30%–40% of K–12 teachers in the United States report that their mathematics instruction is promoted by existing state and district accountability policies and 20%–25% reported policies that inhibited their instruction; the remaining proportion were neutral (Banilower et al., 2018).
The situation in the official K–12 curriculum (Remillard & Heck, 2014) may seem somewhat grim based on our description, but there are some bright spots on the horizon and potential areas for the field to take action and advocate to transform the curriculum for future generations. One area of potential promise for the future of data science education in K–12 schools is the currently proposed mathematics framework for the state of California (California Department of Education [CDE] et al., 2022). In the California mathematics framework, there is an entire chapter on how data science should be integrated into the K–8 curriculum by combining recommendations from both CCSSM and the GAISE II report.

Furthermore, for high school, the new CA mathematics framework provides guidance for curriculum in two pathways: “(1) experiences and expertise in data science common for all high school students, and (2) experiences and expertise for a high school pathway with a data science focus” (CDE et al., 2022, p. 52). This represents the first concerted effort by a state to include concepts and practices of data science into their K–12 mathematics standards. Additionally, there is a significant push in the framework to explicitly address issues of equity and critical citizenship in the teaching of statistics, which has been called for by scholars in statistics education for many years (Frankenstein, 1994; Lesser, 2007; Nicholson et al., 2018; Weiland, 2017, 2019). The framework also explicitly calls for the use of technology in the teaching of statistics, with the most noticeable addition being the call for teaching programming relevant for data wrangling and analysis in the high school grades for those choosing to focus on a data science pathway.

Despite these improvements, the new California framework is by no means perfect. The framework is anemic on simulation and probability in middle grades—particularly the notion of simulating sampling distributions, which is explicitly called for in GAISE II. Also, similar to critiques of the CCSSM, there is little discussion of probability or chance at the elementary grades. Additionally, though there are explicit mentions of incorporating computational thinking, which is relatively absent from the GAISE II report, there is still little guidance as to how to do this. Questions also remain about data science education in general, which need to be considered. For example, the field itself has not coalesced around a common definition or set of primary concepts, practices, or skillsets. This makes it challenging to create a curriculum—official, written, intended, or enacted—as there is no agreed-upon set of learning objectives for data science. The California framework writers have taken a first step toward attempting to create such a set of learning objectives; however, this is new territory and there is not much prior scholarship to rely upon in making such decisions. At the same time, there is an impetus to capitalize on the growing demand for such curriculum, as the California framework writers point out:

In total, over 70 individual high schools and 15 districts offered a data science mathematics or elective course in California during the 2019–2020 school year that counted for A–G credit (University of California data). That compares to just 34 high schools and 6 districts two years before in 2017–2018. This rapid increase in course offerings is likely an indication of both high interest in and importance of data science content throughout the curriculum. (CDE et al., 2022, p. 80)
This change does not come without growing pains as there has been significant pushback to the frameworks—particularly their focus on equity and de-tracking of mathematics (Fortin, 2021), which has reinfamed old tensions in the so-called ‘math wars’ (Schoenfeld, 2004).

Unfortunately, the pushback in California looks like it may collapse the whole endeavor before it is ever approved. Several other states have not gone as far as California in the scope of their incorporation of data science, but they have gotten much farther along in the process of being enacted. Oregon is a good example of a state revising its curriculum standard K–12 to better incorporate the consideration of data for all students. The 2021 revision of Oregon’s mathematics heavily mirrors suggestions from the GAISE II framework in terms of considering the investigative process throughout grades K–12 (Oregon Department of Education, 2021). Though these revisions are significant and go far beyond what has been implemented in other states recently (see Weiland & Sundrani, 2022), they still do not go beyond common statistical practice. For example, there is no mention of data-handling practices such as sorting or filtering and, though they mention using technology, there is no mention of programming. Additionally, a major change they made was to eliminate the use of simulation, which is a commonly used practice in data science. However, students would likely be well prepared with a strong statistical literacy to tackle data science courses during their postsecondary education if they choose or are able to.

Virginia and Ohio have taken a different approach to incorporating data science into their K–12 curriculum. Though both states have a content strand focused on data in their K–12 standards, their biggest moves to incorporating data science into their curricula is in the form of alternative advanced high school mathematics courses. Virginia adopted standards in April of 2022 for a one-semester high school data science course, which looks quite promising with explicit language about the incorporation of open source technology tools and data science problem-solving structures (Virginia Department of Education, 2022). Ohio has proposed a Data Science Foundations course that takes recommendations from the GAISE II framework and also incorporates computer science standards and practices. The course is designed as a possible alternative to Algebra II (Ohio Department of Education, 2021). The proposed framework for the course and standards explicitly intersects statistics and computer science, making for a very promising course. However, all available documentation shows this work is in the draft state, so we will have to wait and see if it becomes a reality. Both of these examples dive deep into data science; however, as elective courses they will likely only be taken by a very small proportion of high school students.

Finally, Georgia has embraced the computer science approach to incorporating data science into its grades 6–12 curriculum. These courses include Data Science I (Georgia Department of Education, 2021a), which “sits at the nexus of mathematical and computational thinking” (p. 1) and Data Science II (Georgia Department of Education, 2021b), which includes “a deeper dive into how statistical analyses can be performed using technical tools and computing skills to solve real world problems” (p. 1). This is similar to Virginia and Ohio, though as electives these courses will likely only be taken by a very small proportion of students. Furthermore,
because they are designated as computer science courses, they may see even less enrollment than those offered as mathematics courses because computer science itself is considered an elective strand and some schools do not have any computer science courses.

3. Curriculum and Policy Recommendations

Now that we have given some background on the curriculum and policy space, we propose some future directions for data science education. First off, we would like to say that transforming education requires partnerships across disciplines and stakeholders. This includes data scientists, statisticians, computer scientists, computer science educators, data science educators, statistics educators, mathematics educators, educational researchers, policymakers, teachers, parents, and students, to name a few. Just as data science is an interdisciplinary space, so is the work that will have to be done in educational settings. We make the following recommendations to stakeholders, drawing connections to Remillard and Heck’s (2014) curriculum enactment process framework, in order to address what we see as some of the most important and immediate needs for curricular transformation:

Official Curriculum

1. Establish a set of agreed-upon key big ideas that are crucial for data science education and data literacy in K–12 education.
2. Engage in state-level curriculum standards review and/or revision processes to enhance students’ opportunities to learn data science in K–12 mathematics coursework.

Operational Curriculum

3. Collaborate with stakeholders to integrate data science into the broader K–12 curriculum, such as through science, social studies, and computer science.
4. Develop data science curriculum materials to integrate into the K–12 mathematics curriculum based on research on the teaching and learning of data concepts and practices.
5. Incorporate content and pedagogy into teacher education to support K–12 teachers and preservice teachers in integrating key learning objectives or big ideas of data science into the K–12 curriculum.

Entire Curriculum Process

6. Conduct research related to recommendations 1–5 on the teaching and learning of data science in K–12 settings.

Initial efforts have begun for the first three recommendations above; however, there needs to be much more transdisciplinary work done toward these needs with large-scale collaborations between various stakeholders. Although the California framework can serve as a proof of concept, significant work is necessary to make it an attainable goal at a larger scale. This is where we see recommendations 4–6 above becoming exceedingly
important. However, for recommendations 4–6 to become a reality, work is needed to build capacity in data science education.

4. Building Capacity

To realize curriculum and policy recommendations 4–6 above requires research to practice partnerships between educational researchers, data science experts, teacher educators, and school personnel (i.e., teachers, curriculum coaches, curriculum specialists, and administrators). One of the biggest and most immediate capacity issues we see in realizing data science education in K–12 educational settings is preparing teachers to teach data science curriculum. Teacher educators and teacher-preparation programs are still scrambling to meet statistics content demands put forth in the CCSSM over a decade ago. For example, Banilower et al. (2018) reported that only 31% of high school and 40% of middle school teachers in the United States felt very well prepared to teach statistics and probability. In a survey of secondary school mathematics teachers, Lovett and Lee (2017) found many teachers did not feel confident to teach many of the statistics concepts in the high school curriculum. The lack of confidence was most pronounced with concepts around simulation-based inference, which is also a common practice in data science.

The call for further support of mathematics teachers to teach statistics concepts and practices is not new, but decades old (Franklin, 2000). Recently, there has been a significant push, starting with the Statistical Education of Teachers (SET) report (Franklin et al., 2015) published by the ASA. The SET report, building from the Mathematical Education of Teachers II (Conference Board of the Mathematical Sciences, 2012), outlines how the standards for mathematical practice in the CCSSM are viewed from a statistical lens, and the statistical concepts and practices mathematics teachers need to have an understanding of, as well as how to teach them, at each grade band (i.e., elementary, middle, and high school). Shortly after the SET report was published, the largest professional organization for mathematics teacher educators—the Association of Mathematics Teacher Educators (AMTE)—published their Standards for Preparing Teachers of Mathematics (SPTM; 2017), which argues that “Well-prepared beginning teachers of mathematics possess robust knowledge of mathematical and statistical concepts that underlie what they encounter in teaching. They engage in appropriate mathematical and statistical practices and support their students in doing the same” (p. 6). Though these documents created space and momentum, there is a significant amount of work that needs to be done in terms of educational research and teacher preparation to meet the goals of these policies. For example, the authors are unaware of a single teacher preparation program for secondary teachers that meets the current recommendations for coursework laid out in the SET (Franklin et al., 2015). We have focused on statistics teacher education thus far, since we have found neither research nor policy documents on the preparation of teachers to teach data science.

Though we have been discussing statistics content, much of the statistics content taught at the K–12 level is exploratory data analysis, which is typically used in data science. Furthermore, data science delves further into the use of technology than much of the statistics content typically taught in K–12 schools. Both authors still commonly see teachers having students do statistical calculations and visualizations by hand or on TI-84
calculators at the secondary level, which is due, in part, to the College Board’s outdated policy that limits students to using a graphing calculator on the Advanced Placement Statistics test (College Board, 2022). To realize data science education where technology is used for most work, such as data wrangling and visualizations, significant work needs to be done to develop teacher technological pedagogical content knowledge (Mishra & Koehler, 2006) in regards to using technology like CODAP (The Concord Consortium, 2020), R (R Core Team, 2021), or Python (Van Rossum & Drake, 1995) to teach data science concepts and practices. This work includes computational thinking as well, which is still fairly novel in mathematics education, let alone mathematics teacher education. Finally, it is useful to note that a) only 5% of high school mathematics teachers in the United States report feeling very well prepared to teach computer science and b) only 27% of high school computer science teachers in the United States feel very well prepared to teach data analysis topics (Banilower et al., 2018). Although we have argued in this article about building capacity from a statistics education perspective, the point is not lost that these statistics show an important gap in the preparation of mathematics and computer science teachers that could also be addressed to help create more substantial overlap to further support building capacity in data science education. So, how will we prepare teachers to take on such tasks?

Capacity building in terms of research and teacher preparation is a key component of improving the statistical education of citizens (e.g., da Ponte & Noll, 2018), and we argue the same is true for data science education. We pair educational research and teacher preparation together as they are often both done by higher education faculty who teach in teacher education programs at postsecondary institutions and also research the education of teachers, which is a synergistic relationship. There have been several notable efforts to support the preparation of mathematics teachers to teach statistics. For example, projects led by Dr. Hollylynne Lee at the Friday Institute at North Carolina State University have produced MOOCs (massive open online courses) for teachers to learn the pedagogy for teaching data investigations and statistical inference in grades 6–12. Furthermore, investigators Hollylynne Lee, Rick Hudson, Stephanie Casey, and William Finzer, in their ESTEEM (Enhancing Statistics Teacher Education with E-Modules) project, produced modules that mathematics teacher educators (MTEs) and statistics teacher educators (STEs) can use in their university-level courses to support the preparation and professional development of teachers. Most other efforts have been quite localized, such as in the form of a single course or professional development experience impacting a small group of teachers (Batanero et al., 2011; Langrall et al., 2017; Peters, 2013). That is not to say these efforts are not important. As a field, statistics educators have learned much about the preparation of teachers from such studies. However, there is a dire need to prepare and support mathematics teachers in teaching concepts and practices from statistics and data science and at large scale (Weiland et al., in press). One project of note working toward preparing teachers to teach data science is the InSTEP project led by Hollylynne Lee, which is still in its development but holds the potential to be highly impactful. The goal of the project is to develop an online learning platform for teachers to learn how to carry out data investigations and incorporate them into their classroom instruction.
The issue of preparing and supporting teachers is exacerbated by the dearth of educational researchers and teacher educators with expertise in statistics education or data science education to support teacher preparation programs and teacher professional development. Furthermore, the SET (Franklin et al., 2015) and SPTM (AMTE, 2017) have only recently solidified space for STEs. There is no such space for data science teacher education. Because of this we focus our attention on statistics education and teacher preparation, which we see as currently the most relevant and the space most likely to take on data science teacher education. Many statistics educators come from either statistics or mathematics education (Zieffler et al., 2018). Both of the authors fall into the latter category. Typically, those coming out of these fields have to go above and beyond to become statistics educators because statistics programs do not often focus on educational issues or theories of learning, and mathematics education programs do not typically include much focus on statistical practice outside of quantitative research methods. Furthermore, statistics education researchers and STEs come from an even wider set of backgrounds than those commonly described above, including statistics, mathematics, mathematics education, educational psychology, and learning sciences. Such diversity is beneficial for bringing new ideas to the field, but it also points to a systemic issue, which is that there are only a handful of programs that specifically prepare statistics educators. In other words, one of the issues we are calling for universities to take on is capacity building in terms of creating programs and support structures to prepare and graduate enough statistics educators that every large public teacher preparation institution is able to have at least one statistics education faculty member. These programs could emerge as concentrations out of statistics or mathematics education programs.

There are only three examples of doctoral programs in statistics education in the United States, namely North Carolina State University’s PhD program in mathematics and statistics education, which emerged out of the field of mathematics education; University of Minnesota’s PhD program in Quantitative Methods in Education, which offers an area of emphasis of statistics and statistics education, emerging out of the fields of educational psychology and statistics; and Montana State University’s PhD program in Statistics, which offers a concentration in statistics education. This is not to say there are not individuals investigating and exploring issues relevant to data science education, there are just no current larger systems or structures in place to clearly support such work in flourishing. Since there are no formal programs with a focus in data science education, there may be a significant gap in the curriculum to prepare educational researchers and teacher educators who have the computability background in working with tools commonly used in data science such as R (base R or via an interface such as RStudio) or Python, or the background to use tools such as R Shiny, Jupyter Notebooks, or bookdown, which may support the teaching of data science—though this is an area that needs study. Beyond just the technological knowledge to use such tools from data science, there is at best a dearth of literature on the Technological Pedagogical Content Knowledge (TPCK; Mishra & Koehler, 2006) necessary for data science education. Answering such questions is important for data science teacher education, and it requires scholars with interdisciplinary expertise that is not currently the focus of any PhD programs to support the development of university teacher educators and education researchers. This also means that only a
handful of people bridge expertise to support creating curriculum for teaching data science concepts and practices to K–12 students or preservice/in-service teachers in those settings. Intertwined with this is the need for researchers with interdisciplinary expertise to study how teachers and students learn concepts and practices of data science to inform the development of effective curriculum. One solution could be the creation of interdisciplinary teams of curriculum writers, mathematics and statistics educational researchers, MTEs, STEs, and data scientists to collaborate on how to tackle these issues. However, creating such teams still requires boundary crossers (Groth, 2015)—and data science crosses many boundaries. Furthermore, academic institutions do not always support such boundary crossing.

Because data science education is an emergent field that has no data science educators or researchers to call its own, and because much of the efforts associated with K–12 environments are situated around statistics content being taught within mathematics courses, statistics education can provide a substantial and meaningful support to capacity-building efforts in data science education. Thus, another capacity-building issue is that large public institutions involved in mathematics teacher education need to prioritize hiring statistics educators. This could involve creating specific tenure lines dedicated to statistics educators, or considering ways of creating joint appointments between departments to lessen the cost burden on a single department. For example, a joint appointment between a mathematics department that prepares secondary education mathematics teachers and a department of curriculum and instruction. Departments also need to prioritize creating or updating their current courses and course requirements so that preservice teachers at every grade have experiences in learning statistics consistent with calls from related professional organizations (AMTE, 2017; Conference Board of the Mathematical Sciences, 2012; Franklin et al., 2015). As a note, we do not mention hiring data science educators because data science education is such a nascent area at this point that one would be hard pressed to find such a person, hence our recommendation for building capacity in doctoral education.

5. Building Capacity Recommendations

To meet the vision of the GAISE II report and thus support all students in developing data literacy, significant transformations need to be made to the educational system as a whole to build capacity for curricular changes. Therefore, we make the following recommendations to support our prior six recommendations:

1. Universities need to explore the creation of academic programs to prepare data science and statistics teacher educators and education researchers.
2. Universities need to put funding behind creating positions and prioritizing the hiring of data science teacher educators and statistics teacher educators.
3. Teacher preparation programs need to modify program requirements to create opportunities for their teacher candidates to learn concepts and practices of statistics and data science outlined in the GAISE II report (Bargagliotti et al., 2020), SET report (Franklin et al., 2015), SPTM (AMTE, 2017), and recent educational research.
4. State, district, and school leaders and curriculum specialists must create opportunities for current teachers to engage in professional learning focused on supporting their development of data science and statistics concepts and practices outlined in the GAISE II report (Bargagliotti et al., 2020), SET report (Franklin et al., 2015), SPTM (AMTE, 2017) and recent educational research.

5. States and districts need to allocate money for high-speed internet access and to create one-to-one computer programs at all public schools so that all students have a computer to use at school and home to meet the technological demands for learning up-to-date skills for being a citizen in society today.

These recommendations require large financial and time commitments, particularly in higher education to begin to transform the system. Governmental organizations and agencies could also support such transformations by providing robust support to such transformational efforts. For example, as recommendations for changes in state-level standards and curriculum decisions are made, institutions of higher education should begin working with state legislators to secure legislative funding, and with Department of Education officials, to support program development and strategic hiring to advance the six recommendations we have made for curricular transformation. Moreover, when these opportunities arise and are combined with increased demands from schools to provide data science options as part of the mathematics curriculum (CDE et al., 2022), it provides some impetus for higher education institutions to create tenure track positions in data science education as well as to build new courses and programs to prepare teachers for the new demands. We acknowledge that these recommendations will require significant commitment from relevant stakeholders, but to meet these ambitious goals, the time to make meaningful advances is now. The world is changing rapidly, and educational transformation is needed to support students in developing the literacies they need for today’s world—not that of decades past.

---

**Disclosure Statement**

This work was conducted without the support of any funding agencies. The authors have no potential conflicts of interest in this manuscript.

---

**References**


Irizarry, R. (2020). The role of academia in data science education. *Harvard Data Science Review, 2*(1). [https://doi.org/10.1162/99608f92.dd363929](https://doi.org/10.1162/99608f92.dd363929)


Zieffler, A., Garfield, J., & Fry, E. (2018). What is statistics education. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 37–70). Springer. [https://doi.org/10.1007/978-3-319-66195-7_2](https://doi.org/10.1007/978-3-319-66195-7_2)

©2022 Travis Weiland and Christopher Engledowl. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.