IES Institute of Education Sciences

Catalyzing a New Field: Data Science Education in K-12

Technical Working Group Summary | October 26, 2021

U.S. DEPARTMENT OF EDUCATION

A Product of the National Center for Education Research

Catalyzing a New Field:
Data Science Education in K-12

Technical Working Group (TWG) Meeting

October 26, 2021

National Center for Education Research
Institute of Education Sciences
U.S. Department of Education
Washington, DC

This meeting summary was prepared by Zarek Drozda (Data Science Fellow at the National Center for Education Statistics) in collaboration with National Center for Education Research (NCER) staff Christina Chhin and Erin Higgins. The summary draws from the slide presentations, notes prepared and taken by Robin Pu Yigh (under JDC Events' contract ED-IES-D-0003), and notes taken by Corinne Alfeld (NCER), Sarah Brasiel (NCSER), and Katherine Taylor (NCSER). The views expressed in this document reflect individual and collective opinions and judgments of the presenters and participants at the meeting and are not necessarily those of the Institute of Education Sciences (IES) or the U.S. Department of Education.

Table of Contents

Technical Working Group Members

Rob Gould, Vice Chair, Department of Statistics
University of California, Los Angeles (UCLA)

Ying (Alison) Cheng, Professor of Psychology
University of Notre Dame

Hollylynne Lee, Professor of Mathematics and Statistics Education
North Carolina State University

Andee Rubin, Principal Scientist
TERC

Talitha Washington, Associate Professor of Mathematics; Director, Atlanta University Center Consortium (AUCC) Data Science Initiative
Howard University, Clark Atlanta University

Andreas Stefik, Associate Professor, Computer Science
University of Nevada, Las Vegas

Maya Israel, Associate Professor, School of Teaching and Learning
University of Florida

Ebony McGee, Professor of Education, Diversity and STEM Education
Vanderbilt University

Elisha Smith Arrillaga, Managing Director
University of Texas, Austin, Charles A. Dana Center

Evan Shieh, Senior Applied Scientist
Amazon

April Swagman, Director of Product
Insight Data Science

Lindsey Henderson, Secondary Mathematics Specialist
Utah State Board of Education

Deborah Crawford, 6-12 Math Supervisor
Frederick County Public Schools

Tina Mazzacane, Mathematics Coordinator
Virginia Department of Education

Gabriel Cartagena, Director, Secondary Mathematics
District of Columbia Public Schools

Kathi Fisler, Professor of Computer Science; Co-Director, Bootstrap
Brown University

Chad Dorsey, President and CEO
Concord Consortium

Sanne Smith, Lecturer, Stanford Graduate School of Education; Director, Education Data Science
Stanford University

Institute of Education Sciences Staff

Mark Schneider, Director
Institute of Education Sciences

Elizabeth Albro, Commissioner
National Center for Education Research

Sarah Brasiel, Program Officer
National Center for Special Education Research

Christina Chhin, Program Officer
National Center for Education Research

Erin Higgins, Program Officer
National Center for Education Research

Katina Stapleton, Program Officer
National Center for Education Research

Zarek Drozda, Data Science Fellow
National Center for Education Statistics

Executive Summary: Panel Recommendations

The National Center for Education Research (NCER) convened a Technical Working Group (TWG) panel of experts to provide recommendations on the steps IES can take to improve the state of data science education research. A full list of panelist recommendations, panelist references and resources, and the meeting's agenda is provided in the appendix.

Panelists defined data science through different lenses. Data science is often described as an inter-disciplinary field combining statistics, computer science, and domain knowledge, along with data ethics, or civic responsibility with data. A working definition was proposed to articulate the practice of data science as an iterative process of collecting data from the real-world and posing questions, exploring and modelling data, and then producing knowledge or automation. A brief discussion distinguishing unique data science processes from statistics and computer science can be found on p. 11, and the need for identifying universal learning outcomes for K-12 students on p. 10.

Concurrently, there was a clear consensus amongst panelists that perfecting definitions would unnecessarily delay tangible progress. Instead, panel members recommended that the education research community should instead focus on building different models and approaches for data science education, and then study what works through pilots and rigorous evaluations.

Panelists emphasized the urgency of investing in data science education research, given that students already interact with data on a daily basis, have pre-conceived notions of what data is and is not, and are increasingly avid consumers of digital information. Currently, students in K-12 education have limited formal learning opportunities focused on data science. To achieve effective, equitable data science education, panelists made recommendations for additional research focused on the following goals:

1. **Articulate the Developmental Pathway**: Panelists suggested more research is needed to better articulate the grades K-12 learning pathways for students. These questions include: 1) how to characterize the unique intersection of skills covered in data science versus statistics or computer science courses; 2) whether to design a separate pathway or integrate data science into existing school subjects such as math, science, or social studies; and 3) which topics should be taught when (K-5, 6-8, and 9-12 grade-levels) to be developmentally appropriate and accessible.

2. **Assess and Improve Data Science Software:** Expanding on Recommendation #1, panelists called for additional research to assess which data analysis software tools should be incorporated into instruction and when (tinker-based tools, spreadsheets, professional software, or other tools) to be developmentally appropriate and accessible. Panelists highlighted significant challenges and barriers created by existing data science tools and terminology, particularly for English learners and students with disabilities. Additional research and investments are needed to expand and ensure accessibility before data science programs and tools can be widely implemented.

3. **Build Tools for Measurement and Assessment:** Panelists highlighted how existing assessment tools do not cover the skillsets nor acumen unique to data science and data literacy. Additional research is needed to develop classroom assessment tools to support teachers and to track students' success and progress, and to ensure students may earn transferable credit for their work. Panelists also highlighted the need to develop holistic measurement frameworks, inclusive of students' data science interests, attitudes, identity, and career aspirations; social-emotional outcomes; and longitudinal labor market and career outcomes.

4. **Integrate Equity into Schooling and Systems:** Panelists emphasized the importance of equity in opportunities and access to high quality data science education for all learners. Data science education research should be conducted through an equity lens that critically examines what is researched and for whom the research benefits. Panelists recommended that research on data science education should include: 1) identifying best practices for ensuring equitable access to and enrollment in data science education opportunities, 2) researching and developing data science software and learning experiences that ensure all students, inclusive of students with disabilities, can access content, 3) ensuring diverse representation of stakeholders, from the researchers to the teachers to the students participating in the study, and 4) examining learning models for students to critically address bias, fairness, and justice that may be inherent in the data and in the interpretation and use of the data.

5. **Improve Implementation:** Panelists highlighted several systemic barriers to successfully implementing and scaling data science education policies and practices, including insufficient resources, lack of teacher training, and misalignment in required coursework and credentials between K-12, postsecondary education, and industry. The panel called for research to evaluate different implementation approaches to reduce these barriers and increase the scalability of data science education policies and practices. Panelists also recommended research on the development and testing of scalable models for implementing data science policies and practices.

Introduction & Motivation

The amount of data in our world has exponentially increased over the past two decades, yet the opportunities to learn about data and acquire data-related skills remain few and far between.

The recent data revolution has generated profound impacts for both the labor market and the demands of daily life. Data analysts and scientists, artificial intelligence, and big data professionals concurrently ranked as the top #1, #2, and #3 jobs demanded globally, with the share of global companies planning to adopt "big data" analytics (90%+), cybersecurity (85%+), and artificial intelligence (80%+) only increasing during the COVID era (World Economic Forum, 2020). Beyond the technology industry, job functions in manufacturing, nursing, and agriculture are becoming more data-driven. Moreover, the everyday tasks associated with consumption of news media, financial markets, and other digital information demand more sophisticated data acumen to simply decipher content. In short, everyone, including our students, face an increasingly data-driven world in which they already live.

In contrast, an estimated 75% of students have not taken a probability or statistics course by high school graduation (NAEP Math M821010, 2019). Despite increasing access to computer science courses, 70% of grade 8 students have not taken coursework for using, programming, or building computers (NAEP Technology & Engineering Literacy D811103, 2018). As the size and complexity of data increases, the foundational skills needed to draw meaning from data will likely become more complex and technologically dependent – representing an evolution even beyond these disciplines. Moreover, students pursuing careers in frontier technologies that leverage large amounts of data, including machine-learning, artificial intelligence, blockchain, and quantum computing, will require a foundation from which to build. Finally, as highlighted by a National Academies of Science Roundtable, K-12 data science education will play a crucial role in increasing diversity, equity, and inclusion in all these emerging fields.

IES has supported innovative research that leverages advanced data science methods to answer research questions[1], but has supported considerably fewer research projects that address research questions specific to data science *education*.

On October 26, 2021, the National Center for Education Research (NCER) at IES convened a technical working group (TWG) to address these gaps and seek recommendations for needed research. NCER invited practitioners in data science education and related fields, representing educators from K-12 and higher education, state and local education agencies, industry, and education research. The goal of this TWG meeting was to provide recommendations to IES on: 1) the goals for K-12 data science education research, 2) how to improve K-12 data science education practice, 3) how to ensure access to and equity in data science education, and 4) what is needed to build an evidence base and research capacity for the new field. This report summarizes key discussion themes and remarks.

[1] For example, IES is currently building a Data Science Center for Excellence, awarding prizes such as the XPrize in Digital Learning Platforms, and partnering with the National Science Foundation to fund two Artificial Intelligence Research Institutes.

Opportunities for Data Science Education

TWG panel members identified several opportunities for K-12 data science education and identified research recommendations to IES to help the field realize these opportunities, address gaps and barriers, and catalyze data science and data literacy education for all students.

Opportunity to increase relevance for students in K-12. TWG members highlighted the opportunity for data science education to easily cultivate enthusiasm among students, with some members positing that engagement should be the primary goal. Data science may have the potential to empower children and youth to make sense of the world. A TWG member noted that social-cognitive career theory is one framework that could be applied to data science education. The theory, as applied to data science education, would predict that creating student awareness in data science can lead to greater interest, persistence, engagement, and self-efficacy, all of which may lead to data science career aspirations and potentially better labor market outcomes. More work is needed to understand how best to cultivate interest and engagement in data science. TWG members highlighted how students in early elementary grades can explore topics of personal interest or social justice as they learn to apply data science skills in these domains.

Opportunity for equity and alternative pathways in STEM education*:* Data science should be considered important for all students, not an alternate track to other disciplines (e.g. calculus), nor an intimidating advanced track that discourages enrollment. Students should gain exposure in middle school or earlier to ensure this opportunity is realized and not siloed to a select group of students late in high school. TWG members noted that the field should research the implications of expanding the gold standard in K-12 mathematics education beyond calculus and consider how outcomes of data science education compare to outcomes of the traditional approach to mathematics education coursework. Outcomes of interest include college entrance exam scores, postsecondary education persistence and completion, and career attainment. Research in this area will help demonstrate whether data science education is a rigorous STEM learning pathway.

Opportunity for teaching "appropriate skepticism": It is possible that data science education can teach students how to balance two extreme viewpoints regarding data : 1) uncritical assessment of findings from data versus 2) unwillingness to accept any conclusion based on data. Learning this balance may enable students to navigate daily life, careers, and citizenship, and to build their sensemaking capability to tackle new data-related problems. Research should explore whether teachers can effectively impart this balance through examples and real-world problems as they teach data science skills. Education researchers should work toward identifying students' misconceptions about data, data management, and data analysis in order to design data science education learning experiences that address common misconceptions. TWG members raised the necessity of identifying "safeguards" to prevent misuse of data analysis, including "set aside" validation data sets, pre-stated hypotheses, and peer review for coding errors before data is communicated to the public.

Opportunity for project-based learning, community-based learning, and applied technology learning curricula: TWG members suggested educational efforts in data science can emphasize teaching students the conceptual skills and processes that support building their own analytic approaches. Program designers may be able to focus on inquiry-based instruction, and organize concepts around problems for students to engage, rather than procedures or names of calculations. Some school districts are enthusiastic about teaching data science because it may provide an alternative to procedural modes of mathematics instruction. Grappling with data can remove focus from procedures and encourages inquisitiveness. Research should be initiated to examine these potential benefits outlined by TWG members.

Opportunity to empower students with skillsets to recognize bias and dismantle oppression perpetuated in technology: TWG members highlighted that educators may be able to teach "emancipatory data science," to produce knowledge relevant to dismantling systemic oppression. This includes teaching students to recognize potential biases in data, analysis, and reporting. One strategy for removing these biases may include designing and implementing curricula that teach students to employ an inquiry-based method in employing data science to address real-world social problems, including racial justice issues. TWG members suggested current K-12 curricula often do not teach students to consider data's social and societal implications, nor connect analysis techniques to real-world challenges. Research should be initiated to examine these strategies and their potential benefits as suggested by TWG members.

Recommendation 1: Articulate the Developmental Pathway for Data Science Education

Need to expedite progress in data science education research and development. Several TWG members stated that the data science education field, including research on data science education, has not progressed as quickly as it should, and recommended accelerating the work as K-12 data science education is at least 10 years behind, and in a "state of emergency." Students already interact with data on a daily basis, have pre-conceived notions of what data is, interact with platforms that use their data, and frequently consume data in their own time. TWG members suggested research informing data science education is also behind to expediently address these challenges.

Need to articulate data science learning progressions, including in early grades. TWG members noted that more research is needed to articulate how data science instruction should be implemented for K-12 education, including 1) whether it should be taught as a separate discipline/track or as an interdisciplinary field integrated into other school subjects and 2) how it should be taught in earlier grades. TWG members also highlighted the need to identify and examine the most appropriate time for students to learn data science concepts, the optimal ordering of content, and the trajectory that helps build the best foundation. The TWG highlighted that existing K-5 math topics (distribution, variability, informal measures of center, categorical data, and data collection) are strong foundations from which to build instruction in data science.

Need to research and evaluate strategies for integration across K-12. TWG members recommended integrating data science across academic domains and argued that students should not be able to “opt out” of exposure to data science education. TWG members suggested soliciting input for how to integrate data science into existing K-12 curriculum from researchers and educators from diverse disciplines, and suggested leveraging existing relevant materials, practices, and lessons learned from mathematics, statistics, computer science, social science, and humanities education to inform data science education. TWG members said data science education should also focus on topics that distinguish the discipline from statistics and computer science, such as data collection and production, data processing and storage, management, curation, and sharing; exploratory data analysis; prediction and automation; data visualization; data security, privacy, and ethics; data acumen; and communicating about data. TWG members also highlighted the need for additional research to identify and formalize the uniqueness of data science as a separate discipline.

Need to identify universal learning outcomes. The relationship between data science education and data literacy needs to be better articulated. TWG members suggested data literacy should be taught across school subjects from a very early age. More research is needed to guide when and how to do so. It was also noted that data science education should avoid developing curriculum with only data scientists in mind, and that it is imperative that all students develop sufficient *data literacy* to become critical data consumers and users to make informed decisions about their own lives and communities. TWG members emphasized that all students should universally learn the basics of modern data and statistics, including but not limited to: an understanding of data structures, use-cases and modern data applications, data collection and measurement, data distributions, variation, probability, research design, and validity and generalizability. More research is needed to understand which competencies are crucial for every student and when they should be taught.

Need to design bridges between K-12 and post-secondary pathways. TWG members called for a more diverse set of educational opportunities for data science, and for research to evaluate which opportunities can increase equitable access and most effective learning in data science education. These strategies could include ensuring students may receive industry-recognized credentials, have access to dual-enrollment opportunities with local universities, and local volunteer and internship opportunities facilitated with institutions of higher-education to create cohesive ecosystems of learning pathways.

Recommendation 2: Assess & Improve Data Science Software

Need to identify which software tools and tasks are developmentally appropriate for different age groups. TWG members advised that researchers should explore how to teach data science to very young children without reliance on computational tools. It is important to support students' creativity and not use tools in a way that unnecessarily constrains their representations of data. Tools should be designed to be usable, not intimidating. These tools should support students in advancing from pre-K to higher learning levels of data science competency. Data science education should also teach students how to select the appropriate tools for different tasks. TWG members noted the importance of identifying developmentally appropriate education tools for early grades, as well as impart training in tools that will prepare students for the labor market by graduation. Data science education should teach students to be flexible learners who can apply general data science principles across contexts and platforms, especially in the face of changing tools and technology over time. Further research and development is needed to fully address these recommendations.

Need for additional research to make data science tools and analysis techniques accessible for all learners beyond current regulatory mandates. TWG members shared that current accessibility standards are problematic when applied to data science. For example, standards for making graphs and images accessible require only including a verbal description. This is not an adequate solution for complex data science outputs such as scatterplots with large numbers of data points or complex violin plots, including visualizations presented in the popular press. Software should be designed to generate charts and graphs that are accessible and do not require descriptive captions. Unfortunately, there is little research on how to make tools more accessible for students with disabilities. Existing accessibility software patches typically only cover one type of disability and fail to hold across a variety of learning and sensory differences. Multiple TWG members expressed concern that there are few, if any, solutions for students with visual impairments given the highly visual nature of core data science skills. Finally, TWG members highlighted that syntax inherent to many modern data analysis tools can be inaccessible and esoteric for all learners, including learners with disabilities, and that training in one tool may translate poorly if at all to others.

Need to create plain language for data science terminology to serve all students. TWG members agreed that data science theory, methods, and analysis techniques should be accessible to all students. TWG members recommended conducting research on how best to include and serve students with disabilities when imparting new concepts, especially with respect to the vocabulary used to refer to certain procedures during instruction. TWG members highlighted that unnecessarily cumbersome terminology ("ANOVA" instead of "compare these things" or "pair-wise t-tests") can create barriers for students with disabilities and students at-large. TWG members recommended building on what is already known about accessibility to change terminology to be more intuitive and consistent across textbooks and programming languages to improve accessibility and equity.

Recommendation 3: Build Tools for Measurement & Assessment

Need to establish standards and assessments for basic data literacy and for data science competencies, as well as measures of student affect and engagement. TWG members highlighted gaps in existing assessment tools, which focus too narrowly on either statistics or computing, and suggested a need for assessments that cover unique data science competencies. TWG members highlighted how existing assessment tools, including tests used in Advanced Placement Statistics, may incentivize teachers to focus on hand-written procedures and procedural knowledge in place of using real-world data or technology to build adaptable and transferable skills. This concern has also been highlighted in survey research with teachers. Assessments should not focus on minutiae to the detriment of key learning objectives that engage students. Student interest, attitudes, engagement, self-efficacy, and career aspirations are important outcomes to measure, as they relate to student learning and achievement. New assessment tools and standards should also build on existing research from related fields, including mathematics, science, computer science, and other K-12 STEM research. Finally, as new assessment tools are being developed, the TWG recommended that they employ modern techniques (including artificial intelligence) in their implementation when feasible, while also balancing and evaluating concerns for fairness and bias.

Need to ensure K-12 data science assessments holistically evaluate progress and prepare students for modern careers. Core data science skills include interpreting computational information, contextual variables and potential confounds, communication, and presentation. The TWG underscored that "data science is a social career," and noted the necessary integration of what is known about social-emotional learning into data science. Data science education should teach students about the broad range of fields and careers that employ data science, including engineering and social sciences. TWG members also suggested industry standards should only guide content in high school or later, and that attempting to replicate data science as performed in industry may not be effective for elementary-level data science education.

Need to support increasing teaching capacity and to develop methods to evaluate instructional quality. TWG members called for evidence about which strategies increase data science teacher capacity, including through credentialing and performance assessment. Administrators should not assume that all mathematics teachers can teach any quantitative subject equally well, and should work to align teaching assignments with teachers' skills and preferences. TWG members suggested examining whether aligning data science content or units with educators' pre-existing skills across different school subjects would support learning standards in other disciplines. It was suggested that the field should develop approaches for monitoring and evaluating data science instruction, in particular to prevent courses from becoming traditional probability and statistics courses.

Need to establish and communicate postsecondary metrics. TWG members called for development of better measures of college and career pathways in data science education. It is important for students to know how data science education can support transitions to higher education and career options in STEM. TWG members discussed the need to better understand the importance of cultivating early student and family interest in data science. Diverse post-secondary options may help parents and students understand the direct value of data science. Data science education and its overlapping pedagogies (student collaboration, project-based learning, community-based learning) can also be linked to other ongoing priorities in STEM education, and evidence should be generated so as to describe those linkages. Research should assess how outcomes of data science education compare to outcomes of the traditional approach to mathematics education. Outcomes of interest include college entrance exam scores, post-secondary education, and careers.

Recommendation 4: Integrate Equity Goals into Data Science Education Research

Need to ensure equity is a core consideration in all data science education efforts. The panel recommended that opportunities for rigorous data science education should be made available to all students, including students from underserved and disadvantaged communities. Research is needed to explore, develop, and test effective strategies and approaches for closing opportunity and achievement gaps in data science education. Moreover, equity should be embedded throughout all topics in data science education research (goals, practice, access, and research capacity). Finally, the articulation of goals, frameworks, and learning outcomes for data science should work to align with higher education programming so that disparities in opportunities and access for underserved and minoritized students are not unintentionally (re)created.

Need to recognize and address barriers to participating in data science. TWG members highlighted that a large majority of data scientists are White and male; only a small percentage are female (Boston Consulting Group, 2020), Latinx, or Black (General Assembly, 2017). Moreover, TWG members highlighted that related academic fields (including statistics and computer science) have historically been designed by and for privileged groups. All stakeholders should pay close attention to patterns of enrollment in data science-related coursework at all levels, in addition to course offerings, after-school resources, and should measure student affect (awareness, interest, persistence, career aspirations) to assess and address barriers to equity and access to data science education. Several TWG members highlighted that students may decide early on that quantitative fields are not for them, suggesting K-12 may be a critical intervention point for determining who does and who does not access the field. More evidence is needed to better describe when and why students make these choices. TWG members recommended that K-12 leaders should then build on evidence to design data science programs to engage the greatest number of students as possible.

Need to determine how best to cultivate interest in data science. Potential approaches to cultivating student engagement include pointing out the role of data science in students' daily lives, such as video game design, and encouraging students to use data science methods to answer questions that interest them. Many students have expressed enthusiasm for using data science to address local community challenges. Certifications and dual enrollment opportunities may encourage students to pursue data science. One TWG member described survey research that suggested that some teachers are hesitant to introduce relevant, rich data-based experiences with computers, primarily because existing standardized tests (APs and state assessments) focus instead on graphing calculators. The TWG member suggested that these norms may incentivize teachers to focus more on antiquated technology and less on what excites kids, limiting on-ramps to data science education.

Need to demonstrate the success of individuals from historically underrepresented and marginalized groups in data science and related fields, and to ensure diverse representation in data science teaching. The TWG panel noted that students should learn about a broad range of data science careers, highlighting the diversity of successful data scientists, including data scientists from diverse racial and ethnic backgrounds, and data scientists with disabilities. Demonstrating diversity and success in careers builds on broader evidence for increasing participation STEM. In addition, diversifying the teacher workforce may also be an important step in helping to reduce potential biases and increase interest and engagement of diverse groups of learners. For example, teachers can present data that illustrate excellence and success among underrepresented groups and be wary of data that may provide a skewed perspective of certain groups of individuals (e.g., data on incarceration rates). Data science education can incorporate lessons that recognize influential data scientists whose work has promoted social justice, such as W.E.B. DuBois and Ida B. Wells. The TWG panel also noted that researchers and educators must recognize and leverage the intersectionality of students' various identities, as they may represent multiple underrepresented populations. Finally, TWG members noted that data science education research should not just be conducted with "convenience" samples. The panel recommended conducting more research in education settings that have limited resources and more challenges with access and opportunities to better understand the conditions under which data science education can be successfully implemented.

Need for accessibility in professional development for teachers, classroom tools, and content. Data science curricula and tools should be accessible to learners with disabilities, and be culturally and linguistically appropriate for all learners. The TWG panel noted that design approaches that ensure all students, inclusive of students with disabilities, can access content , along with culturally responsive and sustaining pedagogies, should be integrated throughout the design and development of data science curricula. Professional development efforts must instruct teachers on how to be universally inclusive. School administrators and policy makers should ensure data science efforts are responsive to communities' needs. Curriculum and program developers must consider accessibility for all learners, not just a subset of underrepresented people. Research that informs how to meet these goals in the context of data science education is limited.

Need to avoid research practices and program designs that perpetuate inequity, and take proactive steps to ensure representation in research. The TWG panel emphasized the importance of including diverse participants, both students and teachers, and specifically students with disabilities or other learning differences, in data science education research. This intentional focus on diverse learners will help address critical questions of what works for whom, under what conditions, and how data science can best respond to students' needs. Data science stakeholders should leverage lessons learned about promoting equity in other fields as appropriate. Course-related gatekeepers and prerequisites can discourage students from learning data science. TWG members underscored that systemic changes to increase representation in education systems require conscious effort to implement inclusive design. Integrating data science into core curricula across school subjects can improve equity and increase access and opportunities for all learners.

Recommendation 5: Test & Improve Implementation Approaches

Need to facilitate partnerships among multiple education stakeholders in all phases of developing K-12 data science education. Involving postsecondary educators and admissions professionals in curriculum design and associated research and development is necessary for ensuring K-12 data science education prepares students for opportunities in higher education, both in data science degree programs and other fields of choice. TWG members noted that recognition and support from higher education will be crucial for the scaling of K-12 data science education. In addition, collaborations between researchers and other education stakeholders (e.g., teachers, guidance counselors, school administrators, parents) will help ensure that the data science education programs, practices, and policies being examined are relevant to and address the needs of all learners. To this end, TWG members explicitly called for more funding opportunities to support research-practice partnerships with community-based organizations as strategic partners, along with identifying the most effective approaches to stakeholder engagement and partnership development. The panel noted that researchers must invest time and effort to develop trusting partnerships between researchers and communities, especially for historically marginalized communities.

TWG members also highlighted the need to work with stakeholders and funding sources across the education system, including career and technical education (CTE), the Regional Education Laboratories (RELs), and historically black colleges and universities (HBCUs) and other minority serving institutions (MSIs). By creating research partnerships with other stakeholders, the field will be able to access additional resources and create an expansive research community that reaches and addresses the needs of underserved and underrepresented communities.

TWG members envisioned that the ultimate goal of research-practice partnerships is to build cohesive data science educational pathways informed by all stakeholders. This includes examining course pre-requisites, transfer credit between institutions, and higher-education admissions requirements, and researching how those requirements affect enrollment and longitudinal labor market outcomes. Survey research to understand how employers may revise degree requirements following rigorous K-12 interventions may also inform the design of research projects. TWG members also highlighted that pathway barriers to equity may discourage students from pursuing coursework in data science or other new school subject areas, and comparing pathway designs through research will be crucial to address this concern.

Need to support teachers through targeted and intensive professional development: TWG members highlighted the interdisciplinary nature of teaching data science, which may require teachers to learn new content and approaches. TWG members noted that research shows teachers can grow into new areas, but that they be intimidated by learning a great deal of new material at once. Members also cautioned against labeling data science as “the new statistics,” which could result in traditional statistics instruction being offered but re-labeling it as “data science.” The panel noted that supporting teachers to successfully implement effective data science will require both an individual mindset shift, along with potential shifts in school policy and practice. The TWG highlighted the potential for teachers to help other teachers to implement new content, the need to rely on peer relationships for scaling programs (especially across school subjects within the same school or district), and for additional research to evaluate and improve these implementation practices to addresses significant capacity constraints.

Need to approach teachers as assets who can contribute to development of data science education across disciplines. The TWG panel noted that existing international statistics education research findings and resources, as well as recent data science education research, can be leveraged to develop teacher training and K-12 data science curricula. TWG members noted the importance of teacher learning communities to support teachers as they develop data science education skills, along with ensuring students’ ability to transfer learned data-related skills from one class to another through curricular coordination. Relegating data science to only one discipline or another may compromise quality, inhibit teachers from multiple perspectives attempting to impart content, and create equity barriers for students by segmenting content. Teachers should be supported in multiple school subjects to ensure wide exposure for all students.

Conclusions

IES looks forward to supporting the research opportunities that TWG members identified during the meeting, including supporting the field in establishing goal posts, identifying effective curricula and teacher training needs, exploring approaches ensuring equitable access to data science education, changing the perception of who can access data science opportunities, and engaging community partners in data science education research. We hope the recommendations provided by the TWG panel will further catalyze collaboration between stakeholders and serve as a useful reference for the field moving forward, as we prepare all our students to succeed in the age of data.

Appendix: Meeting Agenda

11:00-11:30: Welcome, Introductions, Overview of the Day

Mark Schneider, Commissioner, IES
Elizabeth Albro, Commissioner, NCER
Zarek Drozda, Data Science Fellow, NCES

11:30-12:30 Topic 1: Goals for K-12 Data Science Education Research

What research is needed to inform K-12 data science and data literacy education? To understand what research is needed, we first need to articulate the vision for data science education in K-12 given existing theory, research, school implementation considerations, career and industry practice, and the current (or projected) post-secondary landscape.

Discussion Questions:

- *What is K-12 data science? How do we define it, what does it entail, and what are its outcomes?*
- *What research has been conducted, or needs to be conducted, to establish the learning trajectory for data science in K-12? How early can it start, and in what form?*
- *At what grade-levels, school subjects, and post-secondary pathways does data science education belong?*
- *How do we measure student success in data science? What data science measurement tools already exist, and what measures need to be created and validated? What labor market outcomes could be used?*
- *What additional factors (academic acceleration, equity, etc.) should we track and measure? How?*

12:30-1:00: Break (Lunch, East Coast)

1:00-2:00: Topic 2: Improving Practice for Data Science Education

While acknowledging the field of K-12 data science education is nascent, identifying "what works" has both a short-term urgency and is a long-term goal for ensuring continuous improvement. What theory or empirical research exists to identify promising practices for teaching and learning, and what additional research is needed across curricula, professional development, and technology to build evidence-based practices?

Discussion Questions:

- *What research has informed curriculum development to date? How can existing research in statistics, computer science, or math education inform this work?*
- *What research is needed to extend evidence-based practices in data science education across grades K-12?*
- *What resources are needed for schools to successfully implement data science education in grades K-12?*
- *What promising models of pedagogy, teacher training, and professional development exist for data science education in grades K-12? What additional research is needed to inform effective instruction?*

- *How should research on data science education be disseminated to education stakeholders?*

2:00-3:00: Topic 3: Ensuring Access and Equity for Data Science

To make data science truly accessible to everyone, we will need research to ensure students from historically underserved or underrepresented communities, students with disabilities, English-language learners, and students in rural areas can pursue advancement in data science.

Discussion Questions:
- *How can data science education be implemented in grades K-12 to ensure equitable access and opportunities for all students? What additional considerations and resources are needed to ensure equitable access and opportunities?*
- *How can research on data science education address the instructional needs of students from historically underserved or underrepresented communities, students with disabilities, English language learners, and students in rural areas?*
- *What recommendations do you have for IES to catalyze research in these critical areas? How can IES facilitate outreach to the data science education field to ensure these topics are given sufficient attention?*

3:00-3:30 - Break (Lunch, West Coast)

3:30-4:30: Topic 4: Building and Scaling the Evidence Base

Data Science is a fast-evolving field. How do we ensure effective research capacity for growing the evidence base quickly, and for ensuring the field's work is relevant to education stakeholders in 10 or 20 years?

Discussion Questions:
- *What should a school or district leader consider when piloting or implementing a data science education program? How can their implementation efforts contribute to the evidence base?*
- *What capacity (personnel, talent, resources) exists for carrying out this research at the local or state education levels? How can education researchers best support and assist education practitioners?*
- *Who should be involved in research partnerships (government, industry, private sector) to keep data science education research relevant, especially for in 10 or 20 years?*
- *How can IES help facilitate a well-supported improvement ecosystem, that is responsive to technological change?*

4:30-4:55: Lightning Round: Reflections from the Day
- *Based on the discussions from the day, each participant will provide one suggestion to IES on what they see as the greatest research need for K-12 data science education.*

4:55-5:00: Housekeeping Reminders

Appendix: Panelist References & Resources

In preparation for the Technical Working Group meeting, panel members provided either presentations or written comments in response to at least one topic in the agenda. This section gathers research papers, publications, and frameworks that were highlighted by panel members as references for the field to begin building the evidence base and to assist future research. Inclusion of these references does not imply endorsement from the Institute of Education Sciences (IES), and instead are provided as context for the reader.

State Policy References and Examples

Georgia Data Science I and Data Science II Course Frameworks, *Georgia Department of Education*, 2021.

Oregon Math Standards Review and Revision, *Oregon Department of Education*, January 2021.

Ohio Data Science Foundation Course Description, *Ohio Department of Education,* 2020.

Introduction to Data Science Micro-Credential, *Utah State Board of Education*, 2021.

Virginia Mathematics Pathways Initiative, *Virginia Department of Education,* 2021.

Frameworks & Guidance

Bargagliotti, A., Arnold, P., & Franklin, C. (2021). GAISE II (Guidelines for Assessment and Instruction in Statistics Education): Bringing Data into Classrooms. *Mathematics Teacher: Learning and Teaching PK-12*, 114(6), 424–435.

Charles A. Dana Center at The University of Texas at Austin. (2021). Data Science Course Framework. Austin, TX. https://www.utdanacenter.org/sites/default/files/2021-06/Data_Science_Course_Framework_2021_Final.pdf

Data Education in Schools (2020). National Progression Award (NPA) in Data Science Guide for Educators. https://dataschools.education/resource/guide-for-educators-teaching-the-npa-in-data-science/

DeLyser, L. A., & Born, C. (2021). The Future of Problem Solving with Data and Intelligence: Increasing Artificial Intelligence and Data Science Education Across the US. New York, NY: CSforALL. https://www.csforall.org/projects_and_programs/ai-and-data-science-education/

International Data Science in Schools Project (2020). Curriculum Frameworks for Introductory Data Science. http://idssp.org/files/IDSSP_Frameworks_1.0.pdf

Parent Institute for Quality Education (2020). National Best Practices Model in Parent Engagement, 2020. https://www.piqe.org/national-best-practices/

Research Publications

Adams, Caralee (2020). What the Research Says About the Best Ways to Engage Parents, *Hechinger Report*. https://hechingerreport.org/what-the-research-says-about-the-best-way-to-engage-parents/

Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: what is Involved and what is the role of the computer science education community?. ACM Inroads, 2(1), 48-54. https://doi.org/10.1145/1929887.1929905

Ben-Zvi, D., Makar, K., & Garfield, J. (2017). International Handbook of Research In Statistics Education. Springer.

Boaler, J., Cordero, M., & Dieckmann, J. (2019). Pursuing gender equity in mathematics competitions: A case of mathematical freedom. *Mathematics Association of America MAA Focus, 39*(1), 18-20.

Bowers, A. J., Bang, A., Pan, Y., & Graves, K. E. (2019). Education Leadership Data Analytics (ELDA): A White Paper Report on the 2018 ELDA Summit. *Columbia University, Academic Commons*. https://doi.org/10.7916/d8-31a0-pt97

Cox, Megan and Friedman, Kerry (2019). Engaging Families for Math Success, *Institute for Education Sciences*, *Regional Educational Laboratory.* https://ies.ed.gov/ncee/edlabs/regions/appalachia/blogs/blog14_engaging-families-for-math-success.asp

Diekman, A.B., Brown, E.R., Johnston, A.M., & Clark, E.K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science, 21*(8), 1051-1057.

Diekman, A.B., Clark, E.K., Johnston, A.M., Brown, E.R., & Steinberg, M. (2011). Malleability in communal goals and beliefs influences attraction to STEM careers: Evidence for a goal congruity perspective. *Journal of Personality and Social Psychology, 101*, 902–918.

D'ignazio, C., & Klein, L. F. (2020). Data feminism. MIT press. https://doi.org/10.7551/mitpress/11805.001.0001

Education Trust (2019). Five Things to Advance Equity in Access to and Success in Advanced Coursework. https://edtrust.org/resource/5-things-to-advance-equity-in-access-to-and-success-in-advanced-coursework/

Evans, C.D. & Diekman, A.B. (2009). On motivated role selection: Gender beliefs, distant goals, and career interest. *Psychology of Women Quarterly, 33*(2), 235-249.

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. Technology Innovations in Statistics Education, 12 (1).

Fidai, A., Capraro, M. M., & Capraro, R. M. (2020). "Scratch"-ing computational thinking with Arduino: A meta-analysis. Thinking Skills and Creativity, 38, 100726. https://doi.org/10.1016/j.tsc.2020.100726

Forbes, S. (2014). The Coming of Age of Statistics Education in New Zealand, and its Influence Internationally. Journal of Statistics Education, 22(2), 2. https://doi.org/10.1080/10691898.2014.11889699

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372-396. https://doi.org/10.1111/j.1751-5823.2007.00029.x

Geller, W., Cratty, D., & Knowles, J. (2020). *Education data done right: Lessons from the trenches of applied data science.* Leanpub. https://leanpub.com/eddatadoneright

Gould, Robert. "Statistics and the Modern Student," *International Statistical Review,*78(2), 297-315. 3 August, 2010. https://doi.org/10.1111/j.1751-5823.2010.00117.x

Hazzan, O. & Mike, K. (2020). Ten Challenges of Data Science Education. Communications of the ACM. https://cacm.acm.org/blogs/blog-cacm/246219-ten-challenges-of-data-science-education/fulltext

Heinzman, Erica (2021). Math Is No Longer a Four Letter Word: A Mixed Methods Study of Two Non-Traditional Fourth Year Mathematics Classes. University of California, San Diego. https://escholarship.org/uc/item/25d4s7vq

S. Hidi and K. A. Renninger. The four-phase model of interest development. Educational Psychologist, 41(2):111–127, 2006.

Hicks, S.C. & Irizarry, R.A. (2018). A Guide to Teaching Data Science. *The American Statistician, 72*(4), 382-391. DOI: 10.1080/00031305.2017.1356747

Israel, M. (2020). Countering pedagogy for the" privileged" in CS for all: a universal design for learning theoretical and practical lens for addressing the needs of students with disabilities in K-8 CS education. In *Proceedings of the 15th Workshop on Primary and Secondary Computing Education*.

Irizarry, R. A. (2020). The Role of Academia in Data Science Education. *Harvard Data Science Review, 2*(1). https://doi.org/10.1162/99608f92.dd363929

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. Educational Studies in Mathematics. https://doi.org/10.1007/s10649-013-9529-8

Kesar, S. (2017). Closing the STEM gap: Why STEM classes and careers still lack girls and what we can do about it. Microsoft Philanthropies.

Krishnamurthi, Shriram and Fisler, Kathi (2019). Programming Paradigms and Beyond. *The Cambridge Handbook of Computing Education Research.*

LaMar, T., & Boaler, J. (2021). The importance and emergence of K-12 data science. *Phi Delta Kappan*, *103*(1), 49-53.

Lee, V. R., & Wilkerson, M. (2018). Data use by middle and secondary students in the digital age: A status report and future prospects. *Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12*. Washington, D.C. http://sites.nationalacademies.org/DBASSE/BOSE/DBASSE_071087

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A Call for a Humanistic Stance Toward K-12 Data Science Education. EDUCATIONAL RESEARCHER.

Lehrer, R., & Romberg, T. (1996). Exploring Children's Data Modeling. Cognition and Instruction, 14(1), 69–108. https://doi.org/10.1207/s1532690xci1401_3

Lehrer, R., & Schauble, L. (2000). Inventing Data Structures for Representational Purposes: Elementary Grade Students' Classification Models. Mathematical Thinking and Learning, 2(1), 51–74. https://doi.org/10.1207/S15327833MTL0202_3

Lehrer, R., & Schauble, L. (2006). Cultivating Model-Based Reasoning in Science Education. In R. K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences (pp. 371–387). Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511816833.023

Lent, R. W., Brown, S.D., & Hackett, G. (2002). Social cognitive career theory. In D. Brown and Associates (Eds.), Career Choice and Development (4th ed., pp. 255–311).

Luse, A., & Hammer, B. (2017). High School Introductory Programming on Raspberry Pi Made from Scratch. International Journal of People-Oriented Programming (IJPOP), 6(2), 33-45. https://doi.org/10.4018/IJPOP.2017070103

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12?. Computers in Human Behavior, 41, 51-61. https://doi.org/10.1016/j.chb.2014.09.012

Manz, E. (2018). Designing for and analyzing productive uncertainty in science investigations. Proceedings of International Conference of the Learning Sciences, ICLS, 1(2018-June), 288–295.

Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), International Handbook of Research in Statistics Education (pp. 261–294). Cham, Switzerland: Springer.

McFarland DA, Khanna S, Domingue BW, Pardos ZA (2021). Education Data Science: Past, Present, Future. AERA Open. doi:10.1177/2332858421105205

Mike, K., Hazan, T., & Hazzan, O. (2020, November). Equalizing Data Science Curriculum for Computer Science Pupils. In *Koli Calling'20: Proceedings of the 20th Koli Calling International Conference on Computing Education Research* (pp. 1-5).

Moudgalya, S. K., Yadav, A., Sands, P., Vogel, S., & Zamansky, M. (2021, June). Teacher Views on Computational Thinking as a Pathway to Computer Science. In Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1 (pp. 262-268). https://doi.org/10.1145/3430665.3456334

Monroe-White, T. & McGee, E. O. (under review). Emancipatory Data Science: A Liberatory Framework for Mitigating Data Harms and Fostering Social Transformation. In *Proceedings of the 2021 on Computers and People Research Conference* (pp. 23-30). https://doi.org/10.1145/3458026.3462161

Odden, Tor Ole B. Odden; Russ, S Rosemary (2019, January). Defining sensemaking: Bringing clarity to a fragmented theoretical construct. Science Education, 103(1).

Parry, A., O'Hanlon, M., Bowley, M., & Hogan, M. (2020). Investigating the impact that the Raspberry Pi online learning project has on teachers' self efficacy in teaching computing. Cambridge Computing Education Research Symposium 2020, UK. https://www.raspberrypi.org/app/uploads/2020/04/Investigating-the-impact-that-the-Raspberry-Pi-online-learning-project-has-on-teachers%E2%80%99-self-efficacy-in-teaching-computing.pdf

Parent Organization Network (2021), Teacher-Family Partnerships Empower Math Students. https://static1.squarespace.com/static/5e7ce800a65f6b4998972009/t/615f2c73d9dadc2ebcd8976d/1633627257997/2021_PON_Teacher-Family+Partnerships+Empower+Math+Students.pdf

Philip, T.M., Rocha, J., and Olivares-Pasillas, M. C. (2017). Supporting teachers of color as they negotiate classroom pedagogies of race: A case study of a teacher's struggle with "friendly-fire" racism. Teacher Education Quarterly, 44(1), 59-79.

Philip, T.M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data literate about race: A case of data visualizations in the classroom as a site of racial-ideological micro-contestations. Cognition and Instruction, 34(4), 361-388.

Rafalski, Timothy; Uesbeck, P. Merlin; Panks-Meloney, Cristina; Daleiden, Patrick; Allee, William ; Mcnamara, Amelia; and Stefik, Andreas (2019). A randomized controlled trial on the wild wild west of scientific computing with student learners. In

Proceedings of the 2019 ACM Conference on International Computing Education Research, ICER ’19, pages 239–247, New York, NY, USA. ACM.

Ramirez, Pablo (2015). Latino Parent Engagement: Struggle, Hope, and Resistance, *Latino Civil Rights in Education, Routledge Press.* https://www.taylorfrancis.com/chapters/edit/10.4324/9781315672526-22/latino-parent-engagement-struggle-hope-resistance-pablo-ram%C3%ADrez

Ray, M. J., Israel, M., Lee, C. E., & Do, V. (2018, February). A cross-case analysis of instructional strategies to support participation of K-8 Students with disabilities in CS for All. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (pp. 900-905).

Rodríguez-Martínez, J. A., González-Calero, J. A., & Sáez-López, J. M. (2020). Computational thinking and mathematics using Scratch: an experiment with sixth-grade students. Interactive Learning Environments, 28(3), 316-327. https://doi.org/10.1080/10494820.2019.1612448

Rowe, T. (2021, June). “Bridging the Connectivity Divide in K–12 School Districts.” *EdTech Magazine (Focus on K-12)*. Retrieved from https://edtechmagazine.com/k12/article/2021/06/bridging-connectivity-divide-k-12-school-districts

Rubin, A. (2020). Learning to reason with data: how did we get here and what do we know?. *Journal of the Learning Sciences*, *29*(1), 154-164. https://doi.org/10.1080/10508406.2019.1705665

Rubin, A., (2021). What to consider when we consider data*, Teaching Statistics* 43, S23– S33. https://doi.org/10.1111/test.12275

Schanzer, Emmanuel; Fisler, Kathi; Krishnamurthi, Shriram; Pfenning, Nancy; Politz, Joe; Lerner, Ben; Poole, Jen; and Denny, Flannery (2022). Integrated Data Science for Secondary Schools: Design and Assessment of a Curriculum. ACM Technical Symposium on Computer Science Education.

Schanzer, Emmanuel; Fisler, Kathi; Krishnamurthi (2018). Creativity, Customization, and Ownership: Game Design in Bootstrap: Algebra. ACM Technical Symposium on Computer Science Education.

Stefik, A., Ladner, R. E., Allee, W., & Mealin, S. (2019, February). Computer science principles for teachers of blind and visually impaired students. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 766-772).

Stefik, Andreas; and Siebert, Susanna (2013). An Empirical Investigation into Programming Language Syntax. ACM Transactions on Computing Education 13, 4, Article 19, 40 pages.

Torbey et al (2020). Algebra I Before High school as a Gatekeeper to Computer Science Participation. SIGCSE '20: Proceedings of the 51st ACM Technical Symposium on Computer Science Education, 839–844, https://doi.org/10.1145/3328778.3366877

Vogel, S., Hoadley, C., Castillo, A. R., & Ascenzi-Moreno, L. (2020). Languages, literacies and literate programming: can we use the latest theories on how bilingual people learn to help us teach computational literacies? *Computer Science Education*, *30*(4), 420-443.

Vogel, S., Santo, R., & Ching, D. (2017, March). Visions of computer science education: Unpacking arguments for and projected impacts of CS4All initiatives. In proceedings of the 2017 ACM SIGCSE technical symposium on computer science education (pp. 609-614). https://doi.org/10.1145/3017680.3017755

West, S.M., Whittaker, M. and Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. Retrieved from https://ainowinstitute.org/discriminatingsystems.pdf

Wing, J. M. (2006). Computational thinking. Communications of the ACM, 49(3), 33-35. https://doi.org/10.1145/1118178.1118215

Wise, A. F. (2020). Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences*, *29*(1), 165-181. https://doi.org/10.1080/10508406.2019.1705678

Whitaker, D., Foti, S., & Jacobbe, T (2015). "The Levels of Conceptual Understanding in Statistics (LOCUS) Project: Results of the Pilot Study." Numeracy , v.8 , 2015, Article 3

Yadav, A., Zhou, N., Mayfield, C., Hambrusch, S., & Korb, J. T. (2011, March). Introducing computational thinking in education courses. In Proceedings of the 42nd ACM technical symposium on Computer science education (pp. 465-470). https://doi.org/10.1145/1953163.1953297

Yonezawa, Susan; Renner, Nan; Sweet, Monica; Simon, Beth; Baxter, Diane (2015). CS-CAVE: Districts' roles in broadening CS access. ACM Inroads, Vol 6 (4).