Data Science in Education Administration, Policy, and Practice

Alex J. Bowers
Teachers College, Columbia University
Bowers@tc.edu

INTRODUCTION:

> *Traditional analytics can neither easily nor systematically handle the complexities of school data to address the queries school leaders have . . . To answer such questions, we need to utilize school input, process (instructional and programmatic), and output variables in one analysis. [These] data [are] stored in data warehouses (databases designed to store huge amounts of historical data employing special design structures that require additional tools to access) [including] non-normative variables . . . of multiple types: range, ordinal, discrete, and text. The goal of data mining is to use all of these measures in one analysis to develop a comprehensive and accurate model among myriad historical variables already collected that represent the stored totality of a student's school experience.* (Streifer & Schumann, 2005, p. 281–282)

The field of education data science has recently come to the fore in education research, administration, policy, and practice as education systems globally continue to produce ever increasing amounts of data. Yet, calls for applying machine learning, pattern analysis, and data visualizations to support decision-making in organizations generally, and schools specifically, are nothing new, as these calls for what has come to be known as data science have been consistent for over 50 years, especially in education. The purpose of this chapter is to overview the main components of education data science, its history and application to K–12 school administration and policy, and provide examples of how education data science can help inform and support education research, policy, and practice. Data science includes communicating with data, data analysis, and managing data, combined with deep domain knowledge, helping inform management and policy decisions through communicating and visualizing patterns, predictions, and the outcomes of decisions in ways that management and policymakers understand, value, and use. This work includes machine learning, but also a focus on open and shared data and code, ethics, and attention to issues of bias, equity, and community, as well as a focus on prediction accuracy versus model fitting. The chapter concludes with examples of techniques in the application of data science, from exploration, description, and visualization, to machine learning, classification and prediction, and analysis of unstructured data such as text mining analysis.

K–12 education systems globally collect and rely on an ever increasing quantity and variety of data collected across organizations, policy regions, and governments for not only policy compliance reporting, but also to inform decision-making and evidence-based practices and policy at each level of the system (Cohen-Vogel et al., 2015; Figlio & Loeb, 2011; Firestone & González, 2007; Halverson, 2010, 2014; Piety, 2013; Schildkamp et al., 2014; Streifer, 2004). This trend in the increasing complexity and quantity of data systems is mirrored across many industries that are leveraging big data and data analytics through data science and machine learning as a growing priority leading to many recent innovations and rapid growth of the data and information technology sectors (Demchenko et al., 2013; Jin et al., 2015; Saggi & Jain, 2018; Saide & Sheng, 2020; Singh et al., 2019). Indeed, across the education technology sector the industry has experienced rapid investment and growth over the last decade, with much of this growth focused on leveraging the data collected throughout schooling systems, providing analytics and data dashboard systems with the goal to positively influence instructional improvement (Baker, 2021; Hilbert et al., 2021; Vincent-Lancrin, 2021; Wan, 2021; Williamson, 2016).

Despite these trends, while there are individual examples of education organizations finding success with data and their data systems, multiple recent large-scale and randomized controlled studies in the United States focused on districts and regional state-level policy have shown little relationship between student and school performance in relation to data use, data system and dashboard use, and the application of evidence-based data-driven improvement cycles to school improvement (Bloom-Weltman et al., 2021; Bowers, 2021b; Faria et al., 2017; Farley-Ripple et al., 2021; Gleason et al., 2019; Grabarek & Kallemeyn, 2020; Mac Iver et al., 2019; Meyers et al., 2021; Wayman et al., 2017). Recent studies have also demonstrated generalizability issues for these data systems, showing the need for machine learning and

---

[1] This document is a preprint of this manuscript published in the *AERA Handbook of Education Policy Research*, 2nd Edition. Citation: Bowers, A.J. (2025) Data Science in Education Administration, Policy, and Practice. In L. Cohen-Vogel, J. Scott, P. Youngs (Eds.) AERA Handbook of Education Policy Research, 2nd Edition. (pp. 585-612). American Educational Research Association.

[2] Corresponding Author: Alex J. Bowers, Teachers College, Columbia University, Department of Organization and Leadership, 525 West 120th, Box 67, New York, NY 10027–6696; email Bowers@tc.edu

data systems to be retrained and analyzed with data from different school districts and contexts (Coleman, 2021; Stuit et al., 2016), rather than generalizing beyond single organizational contexts. Yet, at the student and classroom levels, research domains such as learning analytics and education data mining (EDM) have demonstrated strong relationships with student moment-by-moment cognition and learning as well as individual student and classroom achievement, such as in mathematics (Baker, 2021; Baker & Boser, 2021; Fischer et al., 2020; Koedinger et al., 2015; Roschelle et al., 2016). Learning analytics, a domain closely related to EDM (Agasisti & Bowers, 2017; Baker et al., 2021; Baker & Inventado, 2014; Feng & Law, 2021; Ifenthaler, 2021; Piety, 2019; Piety & Pea, 2018; Siemens, 2013), coincides with the growth over the last decades of the field of data science. Yet, beyond learning analytics, in education research, policy, and practice, data science has only recently come to the fore.

The purpose of the present paper is to discuss the application of the field of data science to education research and issues of school organization, leadership, policy, and practice. The intended audience for the paper is education researchers, practitioners, and policymakers interested in an overview of the history and current discussion of the application of data science in education organizations and management. In the discussion that follows, I divide the paper into ten sections. In the first section, A Definition of Data Science, I overview the rich conversation in the research and practice literature across the broad data science domain, overviewing the discussion of the work of data scientists from “learning from data” to the application of a rich set of tools that includes data management, visualization, analysis (statistical and machine learning), communication with management and stakeholders, with domain knowledge central to the work. Yet, data science is neither a novel nor new domain, and so in the second section, 50 Years of Data Analytics and Decision-Making: A Brief History, I consider what has been termed “50 years of data science” (Donoho, 2017) as it has developed from the original work of Tukey and Exploratory Data Analysis (EDA) in the 1960s and early 1970s. Importantly for education and organizational management, concurrent to the work of Tukey, organizational theorists, such as Herbert Simon (1971) were mirroring the same calls for the need to make the ever increasing vast sets of data in organizations interpretable by management in a way that helps leaders and policymakers see the patterns that matter to organizational decision-making. Indeed, for education leadership and policy researchers, the research literature from 50 years ago mirrors exactly these calls from what have become known as data scientists within education systems.

In the third section, Education Data Science and the 21st Century, I discuss the current state of the field of education data science (Agasisti & Bowers, 2017; Bowers, 2017; Piety et al., 2014) and the contemporary discussion on the need for the application of the tools of data science in education research, policy, and practice. As an example, I briefly turn then in the fourth section, Testing Management Ideas Using Data Science and Experimentation, to current discussions of the need of data science practices such as causal A/B testing to test management ideas, as the vast majority of management ideas in education and elsewhere are never subjected to experimentation to test the extent to which a prediction was accurate, and thus organizations know very little about the extent of failure or success of management’s ideas as they are put into practice. Nevertheless, training programs and professional development in education leadership, administration, and policy as yet rarely take up the issue of data science, and so in the fifth section, A Roadmap for Training in Education Data Science, I outline the current research on the skills and domains currently proposed across the education data science domain, and next steps for building capacity and training programs to provide the needed capacity building and training in data science as applied to issues of education organizations, leadership, and policy.

Throughout this discussion however, one might question the extent to which this discussion of data science is just statistics, and thus the current quantitative training and research across education may already address these issues. Hence, in the sixth section, Accuracy of Prediction Versus Model Fitting, I foreground the longstanding critique of statistics research that notes that almost all of applied statistics is focused on model fitting and reporting p-values, variance explained, and effect sizes, yet what the decision-maker wants to know is how accurate are the predictions of the model (Breiman, 2001), and thus accuracy of predictions is a core concern of data science, equal to or above model fitting. Yet, a focus on prediction, especially in machine learning, has downsides, and so in the seventh section, Machine Learning Only Learns From Data, Data from a Flawed and Inequitable System, I turn to a discussion of issues of fairness and bias in prediction and early warning systems, discussing the “4As” of algorithms in education of Accurate, Accessible, Actionable, and Accountable (Bowers, 2021b).

To address these issues, the eighth section, The Common Task Framework (CTF): Building Capacity in Data Science, presents an overview of the Common Task Framework (CTF), which has been termed the “secret sauce” of data science (Donoho, 2017) and includes (a) open large-scale real-world deidentified datasets, (b) a shared culture of shared code for shared research, (c) public and open evaluation of algorithms, and (d) neutral referees. Then, in the ninth section, Data Science as a Third Methodology in Education Research, I propose data science as a third methodology in education research to join quantitative and qualitative research, which when combined with a focus on theory, description, and prediction, has the potential to help bridge between methods domains that focus on tabular data, traditionally the domain of quantitative methods, and unstructured nontabular data such as text, images, and videos, traditionally the domain of qualitative data. And finally, in the tenth section, Conclusion and a Look to the Future, I conclude by pointing to three potential near-term benefits of the integration of

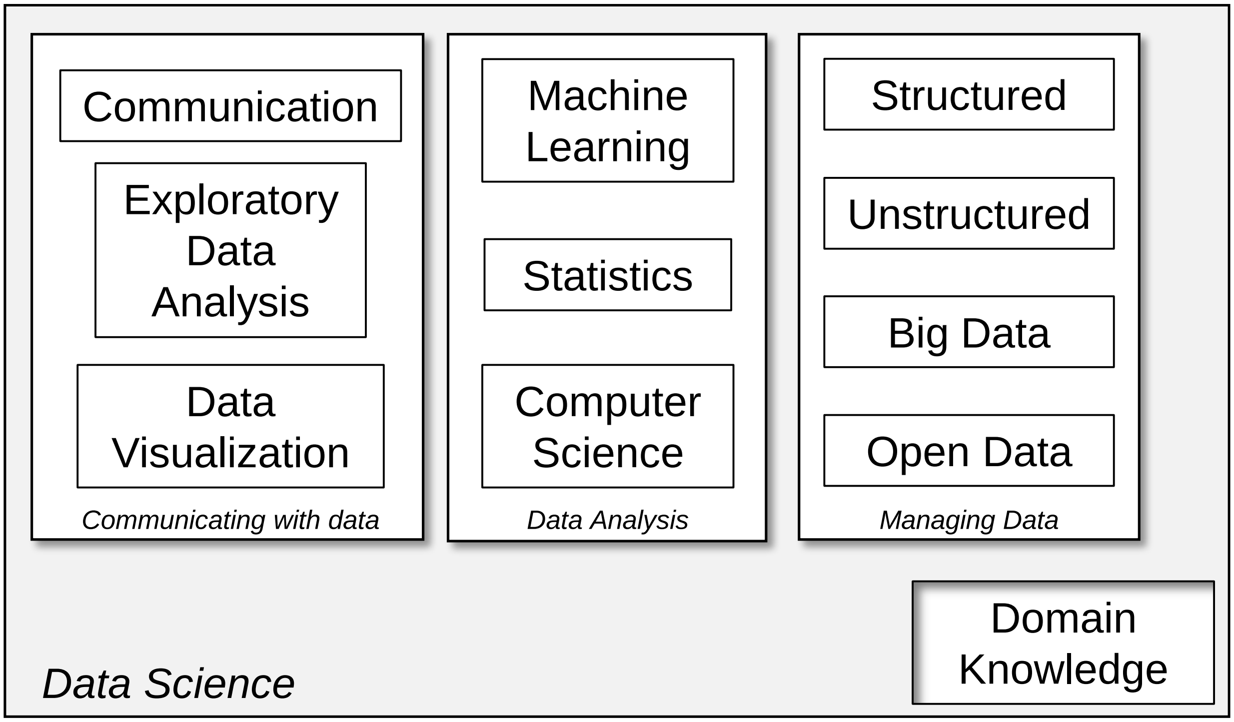

Figure 1: Conceptual diagram of the core components of data science

data science into the lexicon of education research, administration, policy, and practice.

A Definition of Data Science

Data science is a domain of research and practice that, as Donoho (2017) stated, is concerned with "learning from data." Originating more than 50 years ago in the work of Tukey in "The Future of Data Analysis" (Tukey, 1962; see also Donoho, 2017; Singer, 2019), the field of data science brings together data collection, data modeling and analysis, and decision-making through problem understanding and problem solving (Wu, 1997). As such, data science combines managing data across multiple forms (structured and unstructured) and sizes (small, large, big), data visualization and EDA, statistics, computer science, machine learning, and communication and interpretation through domain expertise with stakeholders (Blei & Smyth, 2017; Cleveland, 2001; Donoho, 2017; Rodolfa et al., 2019; Schutt & O'Neil, 2013; Streifer & Schumann, 2005; Ullman, 2020). A recent comprehensive literature review of the definitions of data science, with a focus on data visualization and decision-making concluded that "data science is a multidisciplinary field that aims to learn new insights from real-world data through the structured application of primarily statistical and computational techniques" (Crisan et al., 2021, p. 1862), which includes four primary processes of preparation, analysis, deployment, and communication.

While many authors have worked to provide conceptual diagrams of these overlapping domains and skillsets that help to define data science (Conway, 2010; Crisan et al., 2021; Taylor, 2016; Ullman, 2020), in Figure 1 I summarize this literature as the three central domains of data science of (a) communicating with data, (b) data analysis, and (c) managing data. Importantly, a central theme throughout the data science framework literature reflected in Figure 1 is that the foundation under all data science work is deep domain expertise. As noted by Streifer and Schumann (2005) for this work to be useful "someone in the organization who possesses an understanding (domain knowledge) of both the organization and the process of data mining is needed to facilitate the interpretation of the data-mining results into language understandable and usable by others in decision-making positions" (Streifer & Schumann, 2005, p. 285). Thus, data science is fundamentally concerned with the domain in which it is applied.

Communicating with data is concerned with direct and actionable work with stakeholders that informs and supports decisions primarily through useful and informative EDA and data visualization (Bowers, 2021a; Bowers & Krumm, 2021; Crisan et al., 2021; Krumm & Bowers, 2022), building off of the long history of EDA as a methodological field (Behrens, 1997; Behrens et al., 2012; Singer, 2019; Tukey, 1977). Data analysis is concerned with modeling, inference, prediction, and discovery through the application of machine learning and statistics to data. A central emphasis is on the computer code used throughout the work, as open, shareable, and machine-readable code that can be

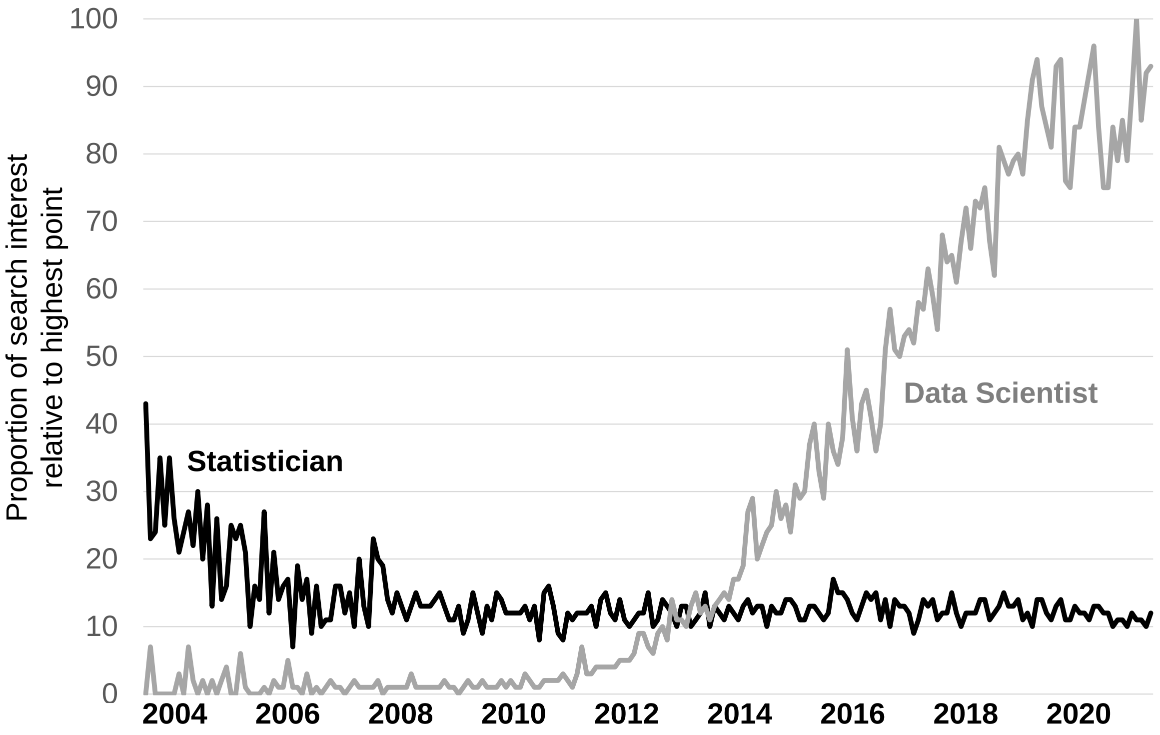

Figure 2: Data Scientist surpasses Statistician in 2013 on Google Trends. The x-axis is year from 2004 to 2021, the y-axis represents global Google search interest relative to the highest point on the chart over time. https://trends.google.com/trends/explore/TIMESERIES/1632249000?hl=en-US&tz=240&date=all&q=statistician,data+scientist&sni=3

replicated and reproduced (Stodden et al., 2013, 2016) as opposed to general forms of equations written out in research papers in nonreproduceable forms. Managing data is concerned with working with data beyond the organized and structured flat data file and relational databases to also include unstructured data, data that are estimated to be more than 95% of all of the data in an organization, that is usually heavily dominated by text data that may or may not include much else other than a source and a date for each entry (e.g., PDF files, emails, etc.), as well as extremely large datasets that may not fit into the working memory or hard drives of traditional computing systems (i.e., “big data”) (Schutt & O’Neil, 2013), and a focus on open and public datasets that encourage replication and interoperability (Cope & Kalantzis, 2016; Gandomi & Haider, 2015; Mukherjee & Kar, 2017; Tarka & Jędrych, 2020; van der Zee & Reich, 2018). Much of this work throughout the data science and data warehouse fields is organized under the term ETL for Extract Transform Load (Bowers & Krumm, 2021; Mukherjee & Kar, 2017; Vassiliadis, 2009).

Turning from the overall conceptual domains of data science, a useful framework for discussing data science that focuses on the applied skills and practices of data scientists who work across these domains includes four major components (Wickham & Grolemund, 2017) of: (a) data wrangling, including data collection, management, organization, combination, and importantly rearrangement, often employing the “tidy data” framework (Wickham, 2014), e.g. the MELT and CAST functions to quickly convert data from wide to long format (Donoho, 2017); (b) data visualization, stemming from the field of EDA (Behrens, 1997; Behrens et al., 2012; Tukey, 1977), and building upon the research literature in visual and graphical perception (Munzner, 2014; Wickham, 2010; Wilkinson, 2012); (c) modeling, including inferential statistics as well as machine learning, often referred to as data mining (Bulut & Desjardins, 2019; Singer, 2019; Tufféry, 2011); and (d) domain expertise combined with communication skills to relate the findings to stakeholders and leaders (Agasisti & Bowers, 2017; Bowers, 2017, 2021a; Bowers & Krumm, 2021; Crisan et al., 2021; Krumm & Bowers, 2022; Schutt & O’Neil, 2013).

A central issue in the literature comparing the differences between statistics and data science is that while data science includes statistics, data science also includes these other multiple aspects of “learning from data” (Donoho, 2017). Indeed, as one indication of this distinction, while “data scientist” as a search term in Google Trends (Singer, 2019) surpassed statistician in December 2013 (Yau, 2013), as plotted in Figure 2, since 2013 the term “data scientist” has continued to grow in interest. However, this growing interest in “data scientist” does not appear to be at the expense or replacement of "statistician," as interest in statistician has remained generally unchanged for more than a decade. Thus, while data science includes statistics, the four domains of data wrangling, data visualization and EDA, modeling, and communication with stakeholders provide distinctive facets to the work that is generating broad interest globally in academia and industry. And yet, while Figure 1 indicates a recent surge in interest in data science, data science as a discipline has been a topic of interest of data analysts and organizational management for over 50 years.

50 Years of Data Analytics and Decision-Making: A Brief History

As Donoho (2017) notes, the field of data science has existed for over 50 years. As the purpose of the present paper is to discuss the application of data science in education, administration, and policy, a useful place to start is also 50 years ago in the organizational and education management literature. Indeed, the work of data science of the last 50 years corresponds to the rise of large-scale datasets across industries globally, especially in considering organizational administration and policy, as ever more systems collect increasing amounts of data in varied ways. Across organizations, there is the perennial question of how to use this growing resource of information to inform decisions and policy making, especially in education (Piety & Pea, 2018). From the perspective of policy, administration, and organizational management, a central concern of this process was summarized well by Simon over 50 years ago:

> In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it (Simon, 1971, p. 40–41).

Simon then went on to discuss in the same piece that the work at the intersection of organizational management and data systems is in creating systems through research, algorithmic organization of the data, and machine learning (in 1971) to organize the information efficiently for the manager, as Simon asked "How can we design organizations, business firms, and government agencies to operate effectively in such a world? How can we arrange to conserve and effectively allocate scarce attention?" (p. 41). He answered this question with the following:

> The proper aim of a management information system is not to bring the manager all the information he needs, but to reorganize the manager's environment of information so as to reduce the amount of time he must devote to receiving it (p. 44).

Contemporaneously, in referring to information management systems specifically for education, Farmer noted in 1970:

> I believe that the greatest impact of the quantitative approach will not be in the area of problem solving, although it will have growing usefulness there. Its greatest impact will be on problem formulation: the way managers think about their problems—how they size them up, bring new insights to bear on them, relate them to other problems, communicate with other people about them, and gather information for analyzing them (p. 21).

Together, these insights from over half a century ago on data use in organizational management and administration reflect the usefulness and potential impact of data analytics and data science processes to positively influence organizational improvement. Yet, in bringing together these ideas of information systems to inform and aid in the management and administration of industry and government organizations through organizing and reformulating the data to address organizational problems, there is an assumption that the data analysis and modeling will be understood by policymakers and decision-makers. Despite Simon's and Farmer's optimistic views 50 years ago of the potential of information management systems to positively support the work of organizations and administrators in the evolving era of an information and data-rich world, contemporaries, such as Clemson (1978), noted that these systems depend specifically on modeling the data, with the point especially for education that:

> Attempting explicitly to model an educational system is difficult . . . Most attempts at modeling are further hampered by the fact that invariably mathematical techniques and programming languages are used that have technical requirements that are so exacting that the manager is excluded from meaningful participation. Two serious consequences can result. The manager may not understand the model, and, therefore, even if it were a good model, he is unlikely to use it. Further, by excluding the manager from the model-building process, the model will not be tested against the manager's own store of experience with the situation. This is tantamount to saying that the model will not reflect the political realities that are crucially important to the manager. Therefore, in terms of the manager's needs, the model will not be a good model (p. 22).

Thus, to address this issue in education of the dual needs of (a) the manager to understand the data, models, and results, while at the same time (b) deep expertise is needed within the organization to manage, transform, visualize, and analyze the information in service to the goals of management and the community. Bruno and Fox (1973) noted that "it can be concluded from these trends that school administrators need background in and exposure to the techniques of quantitative analysis in addition to traditional statistical methodologies" (p. 14). Nevertheless, in a special report commissioned at the time by the University Council for Educational Administration (UCEA) on quantitative training for education administrators, these authors note the growing need for what they termed the "education quantitative specialist" role in school districts and state systems (Bowers, 2017; Bruno & Fox, 1973) in which:

> The specialist in quantitative analysis can assist the school administrator in decision making and planning: Part of the quantitative analyst's role is to help the district to diagnose problems and to ask relevant questions concerning planning and evaluation. His information can be made available to the community, to the school board, and to political action groups, as well as to the administrator. He can outline the future implications of proposed plans. His efficacy in the administrative process, however, is contingent on the ability of the decision-maker (the general administrator) to understand and utilize the data from quantitative analysis (Bruno & Fox, 1973, p. 11).

They go on to then provide a job posting from the early 1970s for the Dallas, Texas, school district for the role of the educational systems analyst, a job posting that required skills in organizing and managing data, computing, statistics, visualization, communication, and expertise in the management of education systems. This list of needed skills in the 1970s mirrors those noted above as the core components of the work of data science in the 21st century. Thus, this issue of the need for education organizations to have a job role that includes data management, data analysis, and data communication, with deep domain expertise, in which the analyst is able to help directly inform planning, decision-making, and policymaking, is neither novel, nor recent. For half a century, this intersection of organizational management and decision-making with data and information systems has been seen as a core component of building more effective decision-making, yet there is a strong need in the organization for data analytics personnel who focus on communication and engagement between data system designers and modelling, and decision-makers and stakeholders. In effect, this literature called for the need for people who can both talk to people and who are also people who can talk to machines.

Education Data Science and the 21st Century

More recently, this intersection between stakeholders, management, policymakers, and the data scientists has come to be seen as a core concern of data science as a field. This is reflected in the description of the work of data science by Schutt and O'Neil (2013):

> A data scientist is someone who knows how to extract meaning from and interpret data, which requires both tools and methods from statistics and machine learning, as well as being human . . . Once she gets the data into shape, a crucial part is exploratory data analysis, which combines visualization and data sense . . . She'll find patterns, build models, and algorithms . . . She may design experiments and is a critical part of data driven decision making. She'll communicate with team members, engineers, and leadership in clear language and with data visualizations so that even if her colleagues are not immersed in the data themselves, they will understand the implications (p. 16).

In this description, Schutt and O'Neil (2013) focused on the work of data science as work that is first human-centered, in which the data scientist makes "data sense," who then communicates, so that stakeholders will understand the implications. Thus, data science works beyond modeling data, sitting at the intersection of research and application in practice, working to inform decisions with evidence and information from the data systems deployed throughout the organization.

In education research and practice, the term "data mining" preceded data science (Baker & Yacef, 2009; Piety, 2019; Singer, 2019), such as Streifer and Schumann's (2005) application of machine learning to predict student achievement to inform school leadership decision-making, noting that:

> Data mining is more than statistics: It is a process of problem identification, data gathering and manipulation, statistical/prediction modeling, and output display leading to deployment or decision making. Fundamental to this process is the ability to sort through and utilize the vast amounts of data already collected on students by districts and stored in data warehouses (p. 283).

One of the earliest uses of the phrase "educational data sciences" to refer to the field of applying the field of data science to issues in education was by Piety et al. (2014) in which they noted at the time there were four overlapping communities of research and practice that included learning analytics through personalization, learning analytics through EDM, academic/institutional analytics, and systemic/instructional improvement. In working to define the work of educational data science, these authors noted that:

> Considering these four communities that appear to make up the Educational Data Sciences, we see a number of important features emerge across them. These following five features inform our description of this nascent field.
> 1) Rapid evolution indicative of a broad sociotechnical movement.
> 2) Boundary issues indicating commonality.
> 3) Disruption in evidentiary practices.
> 4) Questions about visualization, interpretation, and culture.

5) Ethics, privacy, and information architecture (p. 196).

While these features of education data science are important to consider as the tools and techniques of big data analytics and machine learning have made their way into education research methods (Cope & Kalantzis, 2016; Fischer et al., 2020; Hilbert et al., 2021; Pea & Jacks, 2014; Piety, 2019; Piety & Pea, 2018; Singer, 2019; van der Zee & Reich, 2018; Williamson, 2017), Piety et al.'s (2014) summary was understandably focused on learning analytics and the application of machine learning to the recent rapid growth of large sets of structured and unstructured data across the education enterprise from K–12 through post-secondary and into careers.

Building on this work, Agasisti and Bowers (2017) worked to bridge between this learning-analytics and machine-learning focus to the over 50 years of organization-level data analytics. This research focused on the applied work of organizational managers at the intersection of intensive data use and organizational improvement as a human-centered activity grounded in the relationships between management and data analysts in education administration (Agasisti & Bowers, 2017). Mirroring the authors above from across the last 50 years of research and practice, these authors noted in considering the traditional separation of data analysis and decision-making in the consideration of the application of data science to education:

> Despite the rapid growth of attention towards the role of data and quantitative information for exploring and analysing educational patterns and results, there is still a relevant separation between decision-makers (principals and middle managers at the institution level, politicians at the governmental level) and data analysts and researchers. In a certain sense, the former actors are aware of the great potential that resides in data, but consider their expertise as technically inadequate to use the analyses. The data analysts, on the other side, are satisfied with their empirical (academic) work, and do not enter the practical life and reality of school management and improvement. The way we see as potentially innovative is to promote the diffusion, within schools, of a new professional profile, that of the educational data scientist, who owns the technical skills to collect, analyse and use quantitative data, and at the same time the managerial and communication skills to interact with decision-makers and managers at the school level to individuate good ways of using the information in the practical way of improving practices and initiatives (p. 190).

In this definition of an educational data scientist, this "new professional profile" reflected Schutt and O'Neil's (2013) general definition of "data scientist" through a communication and relationship focus, mirroring Bowers (2017), who argued that for data scientists in education organizations specifically "a central role of the quantitative methods work of practitioner-scholars is to focus on translating data analysis into actionable information for evidence-based improvement cycles" (p. 88). Thus, the work of a data scientist in an organization includes skills in data management, organization, analysis, machine learning, and visualization, but also includes communication and collaboration with management around decision-making with data. Echoing Tukey's exploratory data analysis points above, this work aligns with recent calls across both education research and psychology and sociology in general, for an increased role of descriptive analysis (Loeb et al., 2017; Yarkoni, 2022).

This data science work differs from machine learning engineering (Wan et al., 2021) in that a machine learning engineer may focus the majority of their time on applying machine learning algorithms to the data at hand (testing, training, iterating, validating) with a focus on the machine learning algorithms themselves. For example, in learning analytics and EDM there is a central focus on machine learning and "discovery with models" (Baker & Yacef, 2009) in which researchers apply machine learning algorithms broadly in education settings (Baker, 2013; Baker & Boser, 2021; Baker & Koedinger, 2018; Fischer et al., 2020; Gašević et al., 2015; Ifenthaler, 2021; Ifenthaler & Yau, 2020; Piety, 2019; Piety & Pea, 2018; Siemens, 2013). As noted by Ifenthaler (2021) "learning analytics draw on an eclectic set of methodologies and data to provide summative, real-time and predictive insights for improving learning, teaching, and organisational efficiency and decision making" (p.168). Thus, this summative aspect of learning analytics differs from the more formative aspect of education data science as integral to evidence-based improvement cycles through ongoing collaboration with the administration and management (Bowers & Krumm, 2021).

In education, the work of data scientists focuses thus not only on the application of machine learning and data visualization, but also communication, relationship building, and interfacing with management to focus on the questions at hand for current decision-making (Bowers & Krumm, 2021; Hilbert et al., 2021; Krumm & Bowers, 2022; Krumm et al., 2018; Piety, 2019), and building comprehensibility around models and findings (Michael, 2017). Rather than providing summative metrics, in this recent literature, the work of data science is to use the tools from big data analytics, exploratory and visual data analysis, and machine learning to provide decision support systems through formative, iterative, and responsive data analytic expertise in the decision-making process in schools, working to build clarity and understanding about data intensive tasks, outcomes, and next steps for organizational improvement within and between data analysts, stakeholders, and management (Bowers, 2021a; Bowers & Krumm, 2021; Krumm & Bowers, 2022). Nevertheless, while the application of data science tools to issues of organizational management and policy continues to grow, while just about every organizational leader would note that they are data driven, decision-making across organizations historically has lacked an empirical focus that tests management decisions against actual data.

Testing Management Ideas Using Data Science and Experimentation

Indeed, this type of formative data science process, of using data and evidence to interrogate and test organizational and management ideas, is a core operational procedure across a growing set of industries. Reports from across many technology sector industries indicate that internal organizational estimates are that the vast majority of business ideas when tested rigorously either have a neutral impact or fail drastically with some estimates as high as 90% of all business ideas having negative or unexpected outcomes, such as decreasing customer engagement (rather than increasing) and decreasing revenue (Colson et al., 2021; Kohavi & Longbotham, 2016; Thomke, 2020). In arguing for the use of causal experimentation using A/B testing, also known as randomized controlled experiments, Kohavi and Longbotham (2016) stated that "most who have run controlled experiments in customer-facing websites and applications have experienced this humbling reality: we are poor at assessing the value of ideas" (p. 3). As noted by Thomke (2020), this work of data scientists of:

> …business experimentation is of absolute importance to a company's ability to compete. Experimentation helps us begin to answer the kinds of questions that all organizations confront: How do we know what products to make, what customer experiences to offer, and what information we need to make those decisions? How can we begin to innovate if we don't know what customers want and are willing to pay for? How do we direct our organizations' resources wisely? How can we distinguish between cause and effect? How can we reduce uncertainty in our decision-making? (p. 6–7)

The issue however is that historically a pervasive feeling throughout organizations is usually that management's ideas don't need to be tested because they will be successful because they come from management. As noted by Colson et al. (2021):

> However, the dismal failure rates appear to be known only to the few companies that regularly conduct experiments to scientifically measure the impact of their ideas. Most companies do not appear to employ such a practice and seem to have the impression that all or most of their ideas are or will be successful. Planners, strategists, and functional leaders rarely convey any doubts about their ideas. To the contrary, they set expectations on the predicted impact of their ideas and plan for them as if they are certain. They attach revenue goals and even their own bonuses to those predictions. But, how much do they really know about the outcomes of those ideas? If they don't have an experimentation practice, they likely know very little about the impact their roadmap is actually having. Without experimentation, companies either don't measure the outcomes of their ideas at all or use flimsy methods to assess their impacts. In some situations, ideas are acted upon so fluidly that they are not recognized as something that merits measurement (section 2).

Thus, there is a growing concern across organizational research that with the ever increasing amount of data available to an organization, no longer can decision-making rest upon a good argument about an idea from management with the assumption that the idea will succeed (Benoliel & Berkovich, 2020), but rather many ideas should be tested against the data to support decisions that actually will lead to the intended outcome, as most ideas will fail or have the opposite or unintended consequences given the nonlinear and complex organizational system (Gans et al., 2019; Ghosh et al., 2020; Thomke, 2020). In effect, causal experimentation using randomized controlled experiments and A/B testing (Athey & Imbens, 2015; Kohavi & Longbotham, 2016) is needed. Indeed, recent research tracking causal A/B testing use by startup companies suggested that while few startups adopt A/B testing, those companies using A/B experimental testing of their ideas improved performance 30–100% faster over one year (Koning et al., 2022).

In education, recent examples at scale using causal A/B testing and large datasets across many organizations have shown for example in the Massive Open Online Course (MOOC) domain that specific interventions can show consistent small treatment effects (Kizilcec et al., 2020; Savi et al., 2017), and new platforms continue to be developed to test ideas, such as the timing of feedback on class assignments (Fyfe et al., 2021). Another recent example of this type of testing is the examination of the straightforward idea of schools giving attendance awards to students for perfect attendance, demonstrating that these types of awards actually have the opposite of the intended effect and decrease subsequent student attendance (Robinson et al., 2021). Yet, embedding this causal A/B experimentation work as a consistent and pervasive practice in organizations, and education organizations specifically, requires the training of data scientists with the tools to test and communicate the results to help support decision-making (Fischer et al., 2020), thus indicating a strong need for roadmaps to plan and then deliver training in education data science.

A Roadmap for Training in Education Data Science

In applying data science in education and reflecting on the work of Bruno and Fox (1973), Bowers (2017) noted that to address the dual issues of organizing the information and communicating models and outcomes to decision and policy makers, that school management preparation programs have at least two different types of practitioners to train. First, school administration programs should “train school leaders around knowing what concepts and tools are available, when and where to apply them, and what the limitations may be” (Bowers, 2017, p. 81). Second, in considering the needs of organization-level data analytics (Agasisti & Bowers, 2017), there is a strong need across schooling organizations for training in data science that applies to education. Recently, this education data science research and practice domain has been proposed as Education Leadership Data Analytics (ELDA) (Bowers et al., 2019) in which:

> Education Leadership Data Analytics (ELDA) practitioners work collaboratively with schooling system leaders and teachers to analyze, pattern, and visualize previously unknown patterns and information from the vast sets of data collected by schooling organizations, and then integrate findings in easy to understand language and digital tools into collaborative and community building evidence-based improvement cycles with stakeholders (p. 8).

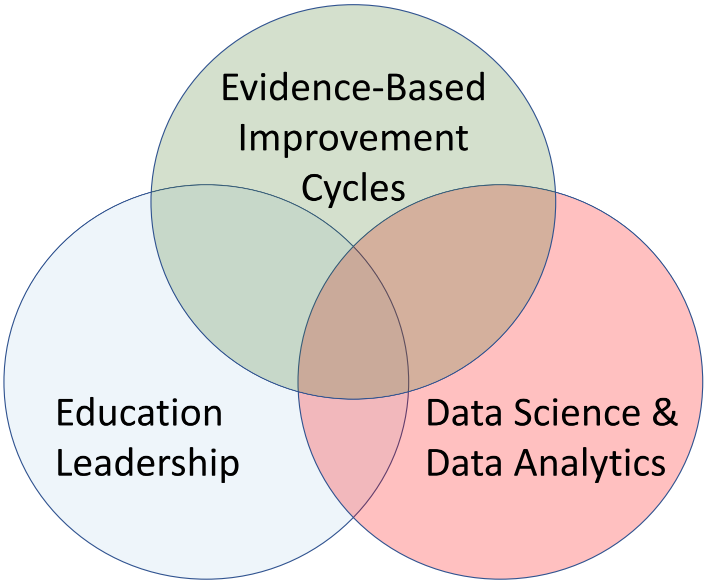

Figure 3: Education Leadership Data Analytics (ELDA) Venn Diagram (reprinted from page 9 Bowers et al. 2019).

Thus, this definition of the work of ELDA practitioners places the application of data science in schools within a larger framework of the current research at the intersection of education leadership, data driven decision-making and evidence-based improvement cycles, and data science and data analytics, as depicted in Figure 3 (reprinted from page 9 of Bowers et al. 2019).

Integrating the responses of over 100 education practitioner and researcher participants in a design-based project around the skills and practices of the emerging domain of education data science and ELDA, Bowers et al. (2019) also proposed a roadmap for training ELDA practitioners in the skills and competencies identified, including four levels of training for the practicing administrator, quantitative analyst, research specialist, and educational data scientist, each with overlapping needs and competencies.

The ELDA training roadmap described in Figure 4 from Bowers et al. (2019, p. 33) attempts to capture the many and varied overlapping and related issues needed for ELDA practitioners, which include the core components of education data science previously discussed, and combines these with many of the types of training already available for education practitioners in colleges of education and school practitioner certification programs, including issues of school leadership, learning analytics, statistics, policy, and assessment, evaluation, and psychometrics (Bowers et al., 2019). This roadmap mirrors the broader data science training literature that focuses on overall data science technical skills while also focusing on moving away from toy datasets to working on real applied data from actual organizations in a domain to train practitioners through doing the work, perhaps in residencies (Rodolfa et al., 2019). As the ELDA domain has only recently been proposed, questions remain as to the extent to which ELDA is separate or synonymous with the work of education data science (Bowers, 2021a; Bowers et al., 2019; Bowers & Krumm, 2021; Krumm & Bowers, 2022), similar to issues of learning analytics and data science (Piety & Pea, 2018), as the application of data science methods within the broader dialogue and collaboration in organizations around the use of evidence for decision-making (Piety, 2019) is the key component of both current conceptions of ELDA and education data science. Throughout this chapter, I use the terms interchangeably.

A recent example of this work at the intersection of data science models, application to practice, and decision-making, is demonstrated in Bowers and Krumm (2021). In this study, we partnered with a charter management organization to pattern analyze and visualize the logfile clickstream data from hundreds of students in Algebra I across their organization, visualizing the patterns of how many times students took and retook teacher-made summative assessments across the many strands of Algebra I. Importantly, however, the goal was not an analysis and then reporting the finding, leaving the application and next steps up to the organization, but rather the interactions between the data analysts and the Charter Management Organization (CMO) leaders around the data analysis and visualization included the following:

> . . . the leader then turns back to the analyst, and generates ideas for restructuring the figure, to ask new questions that are of interest to the school in an effort to find solutions for specific students on specific curriculum modules and assessments. The visualization helps the practitioner rework the problem of practice, and turn the question back to the data analysts. This

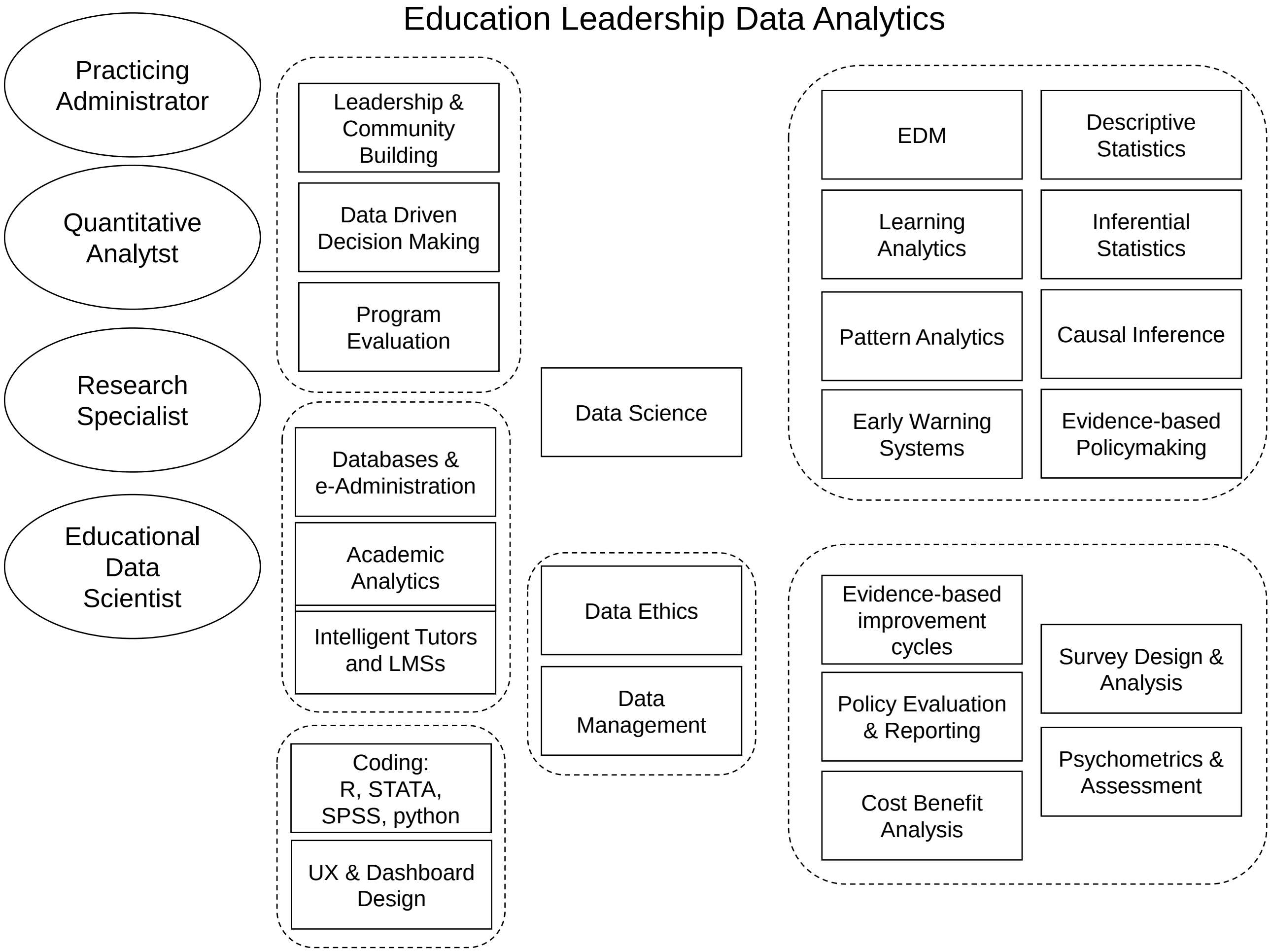

Figure 4: Core competencies and skills for Education Leadership Data Analytics (ELDA) (reprinted from page 33 Bowers et al. (2019). LMS = learning management system, UX = user experience, EDM = education data mining.

> move points to an important takeaway from the partnership approach attempted in this project: joint data interpretation meetings matter . . . In this meeting with the support of analysts, practitioners were helped in making sense of the visualization and scaffolded toward principled takeaways. Importantly, for the analysts, many of the subtleties in the data patterns could not have been disentangled absent the practitioners . . . Importantly, participating researchers did not draw conclusions from the visualization and report these to the CMO leaders. Rather, the CMO leaders drew their own inferences and conclusions and communicated these back to the researchers in a way that promoted joint sensemaking (p. 640).

Note throughout this example the interplay between managers and data analysts as a core component of data science in education organizations, that the work of data science is joint work between organizational stakeholders and analysts, as both work to become clearer on this joint work over time, in which the clarity of the task and application of the analysis becomes crisper (Bowers, 2021a; Crisan & Munzner, 2019) through collaborative data-intensive improvement projects (Krumm & Bowers, 2022; Krumm et al., 2018).

Accuracy of Prediction Versus Modeling Fitting

> *The goal is not interpretability, but accurate information* (Breiman, 2001, p. 210).

Throughout the research on the application of data science to issues of organizational improvement in education, a core focus of the work is on machine learning and prediction. Indeed, in the application of machine learning and data science throughout the broader technology industry, prediction accuracy is the primary concern, secondary is interpretation of the model, if interpretation is done at all. For instance, many companies focus exclusively on predicting customer engagement, retention, and sales, as these are key performance indicators (KPIs). Recent advancements in machine learning image and text recognition similarly focus on model prediction over model interpretation, with stunning success in defined areas over the last decade (Jordan & Mitchell, 2015; Raghu & Schmidt, 2020; Singh et al., 2019), such as in image recognition and large natural language processing (NLP) transformer models, such as Google's BERT, and Open AI's GPT series of models, each including hundreds of millions or billions of parameters in a neural network architecture, but with much work remaining on interpreting exactly how the models make accurate predictions (Kjell et al., 2020; Singh & Mahmood, 2021).

Concurrently in education, as the EDM domain is chiefly concerned with applying machine learning and data mining to student learning issues, predictive learning analytics also generally mirrors this focus on prediction accuracy over model interpretation (Du et al., 2021; Feng & Law, 2021; Herodotou et al., 2020), although with the argument that learning analytics also provides a deep grounding in theory and interpretation of these predictions (Baker et al., 2021; Baker & Koedinger, 2018; Baker & Yacef, 2009; Knight & Shum, 2018; Lee et al., 2020; Namoun & Alshanqiti, 2021). Outside of learning analytics, a goal of the broader data mining of education data is also on prediction, for example Streifer and Schumann (2005) noted:

> That the benefits of data mining are the ability to (a) handle very complex data environments in one analysis and (b) systematically apply the results obtained from one cohort to predict the outcome from totally different cohorts with minimal or no data manipulation (p.283).

Nevertheless, in general, across social science research domains such as education, the central focus of research (both theory and applied), is on model interpretation, not on prediction accuracy (Shmueli, 2010; Singer, 2019). For example, in the domain of K–12 at-risk dropout prediction and early warning systems, Bowers et al. (2013) reviewed over 100 dropout flags that were used in research, policy, and practice, using the data in the papers to calculate predictor accuracy and found that (a) prediction accuracy of the flags was haphazardly reported, (b) none of the dropout predictor studies had reported accuracy, mostly focusing instead on reporting p-values and model fit, and that (c) the vast majority of the dropout flags were little better than a random 50–50 guess (Bowers, 2021b; Bowers et al., 2013; Bowers & Zhou, 2019). Rather, across the dropout research, studies focused instead on reporting p-value significance in regression models, occasionally reporting effect sizes, and then interpreting how the related variables helped to explain a theory of student dropout. But it is not model fit about which practitioners and policymakers wish to know more and hear specific details. Instead, decision-makers want to know accurate predictions of outcomes, given the data collected. In many ways, theory testing and a focus on model fit is utterly irrelevant to practice, leading to the long-running complaints of educators, administrators, and policymakers of the irrelevance of theory and model fitting to practice.

As noted by Parker et al. (2017) this issue in education research of focusing on theory and interpretation to the exclusion of (or completely ignoring) prediction accuracy is problematic as:

> The prediction-focused approach is relatively unusual in the analysis of longitudinal data in the social sciences . . . in exploring the literature we found a severe paucity of papers that reported the predictive accuracy of their models on unseen or test data . . . This is concerning, as the goal of science is not merely to explain phenomena of interest but also to be able to reliably predict their occurrence . . . This lack of focus on accuracy may be due to a common, though mistaken, belief that explanation implies prediction. Shmueli (2010: 289) suggests that in fields like psychology and education, statistical models are used 'almost exclusively for causal explanation' typically under the mistaken belief that 'explanatory power . . . inherently possess predictive power'. Thus . . . there has been a plethora of new explanatory models of educational attainment but little evidence has been presented that any of these models predict future events accurately (p. 97).

Thus, while it is useful and desired to focus on interpretation and explanation in academic research in attempting to explain models and advance theory, for school decision-makers, such as in the at-risk dropout literature, they need to know who is going to drop out, which requires an accurate prediction from the data. As noted by Singer (2019):

> Researchers in education—and especially those in allied social science fields such as economics, sociology and political science—began privileging questions that lead to causal inferences to the near total exclusion of descriptive and predictive questions . . . I have found myself worrying that many researchers have forgotten that descriptive and predictive questions have an important place in the world (p. 576).

This issue relates directly to the definition of education data science being at the intersection of providing actionable information for decision support systems and data driven decision-making by not only school leaders, but the educators throughout the system. As Marshall (1997) noted, school "staff need to have the knowledge base to interpret data and events and to predict with confidence conditions, behavior, and performance," (p. 152) while "the goal of a quality school is to develop a level of stability which allows for predicting outcomes with a high level of certainty" (p. 152–153). A focus on theory and models then to the exclusion of prediction rightly seems baffling to the practitioner, contributing to the perception (or actual) uselessness of research applied to practice.

This difference between model interpretation and prediction is a long-standing debate that sets data science apart from statistics (Donoho, 2017; Singer, 2019), and in many ways explains how data science and machine learning have become so dominant across many technology industries while statistics remains at a consistent level as shown previously in Figure 2. Described as the "two cultures" in statistics of the generative versus predictive, Breiman (2001) estimates that 98% of all statistics is the generative type, with only 2% of statisticians (including himself and most of industry) within the predictive culture. As others have pointed out (Donoho, 2017; Singer, 2019), Breiman's (2001) abstract (and I am by far not the first to quote it) is not only damning for the industry of academic statistics, but also provides the core of his argument in that:

> There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools (p. 199).

My thesis throughout this chapter on the application of data science in education is that the tools of data science apply very well to use on education data to solve problems in education organizations in the support of decisions, especially around accurate prediction. Mirroring Breiman (2001), accuracy of algorithmic prediction thus is of primary concern to the practitioner, with theory and interpretation of the model secondary, or even irrelevant to the question at hand if the model is not accurate.

Returning to Parker et al. (2017) and their application of Breiman's (2001) points to algorithmic prediction in education:

> The goal of analysis, then, according to Breiman, is to model this picture in order to be able to a) accurately predict future/unseen outcomes and b) provide insight into the processes that transform the predictors into the outcome. The dominant paradigm to meet these goals is the data-modelling approach in which clearly defined theories are transformed into statistical models that are applied to the data. The aim is to use a given statistical model to simulate the mechanisms by which nature transforms 'x's into 'y's, with the adequacy of results tested by goodness-of-fit tests and checking of assumptions. Alternatively, an algorithm/machine learning approach treats natural functions as both incredibly complex and largely unknown and thus aims to simply build models that most accurately predict future events irrespective of the processes that give rise to the phenomena of interest . . . the difference in approach comes from a primary focus on explanation versus prediction (p. 96–97).

Thus, the vast majority of quantitative work in academic statistics-focused research is the generative data-modelling approach, focused on explaining how x's turn into y's with goodness-of-fit tests as the evidence, an approach that has received quite scathing critiques in this literature of its irrelevance to practice and business decision-making (Breiman, 2001; Donoho, 2017; Shmueli, 2010). Yet, to state that theory and statistical model fitting and interpretation are irrelevant to decision-making in education organizations is a bold claim. However, equally bold is to claim the opposite: that model prediction accuracy is irrelevant to decision-making in education organizations. But page through any quantitative social science or education research journal, and it would be quite hard to find a consistently large fraction of the articles devoted to prediction accuracy on unseen data in comparison to model fitting. My point is that the argument from the "two cultures" perspective is that this is exactly where we find ourselves with the 98% versus 2% split with the vast majority of attention across the education literature focusing on theory, statistical model fitting, p-values, and interpretation, with little to no attention given (outside of domains such as learning analytics) to the extent to which one can accurately predict outcomes given new data. Yet, issues of prediction, prediction accuracy, and integrating these analytic findings into decision-making and evidence-based improvement cycles is at the heart of the use of data science in education.

These parallel arguments from the organizational decision-making literature noted above in comparison to the "two cultures" perspective are striking. As Parker at al. (2017) stated, the "algorithm/machine learning approach treats natural functions as both incredibly complex and largely unknown" (p. 97), which mirrors the same problem in the recent organizational decision-making literature also discussed previously. As noted by Colson et al. (2021):

> One force that is actively working against us is the way we reason about our companies. We like to reason about our businesses as if they are simple, predictable systems. We build models of their component parts and manage them as if they are levers we can pull in order to predictably manage the business to a desired state (section 3).

This explanation also describes the vast number of quantitative education research papers that model linear relationships, such as in regression frameworks, with independent variables and covariates modeled as if the relationships are linearly additive (a grossly over-simplified assumption of real world organizations) in which each variable is a lever we can pull, independent of the other variables because they are controlled as independent variables in a regression framework. Yet, organizations are complex systems that are difficult to predict without experimentation, and thus according to this literature, most management ideas fail or give unexpected results, and so data science through A/B testing and experimentation is a powerful innovation to test many ideas given the growing amounts of data across organizations (Koning et al., 2022; Savi et al., 2017; Thomke, 2020).

I posit that one way to interpret this literature together is that as administration and management programs teach aspiring organizational leaders and policymakers the data-modeling approach (Bowers, 2017), as it is all the instructors know how to do, then it stands to reason that these aspiring managers in training programs pick up the idea that a model explanation equates to accuracy of prediction, or that just having a good theory for how the model fits the data is the point, as that is all that is represented in the literature they read in their training. This may then understandably lead to managers themselves (a) having a good idea in their subsequent organization, then (b) justifying it with their own theory, and finally (c) moving forward toward assumed success, with little interest in testing or questioning their model, or checking to see if the idea ever worked.

This sentiment is captured well by Watts (2017) in asking "should social science be more solution oriented?":

> One reason for the relative rarity of use-inspired basic social science is that . . . real-world social problems are typically messy and multifaceted, thereby greatly complicating the task of evaluating progress or even defining the problem to be solved in the first place. Health, education, inequality, cultural norms, economic policies, and physical environments all interact in complicated ways to produce particular individual and group outcomes. Attempts to understand or influence these outcomes in the real world therefore often result in a difficult choice between focusing on such a small part of the problem that one misses the larger picture, and drowning in complexity. Exacerbating this difficulty is the reality that not everyone cares as much about standards of evidence as social scientists do. Why invest in a multi-year research project when one can simply follow one's instincts, be inspired by a best-selling book, or pay a consultant to deliver an answer in a matter of months? (p. 3)

Why indeed? If models and theory are all that matter, and one is unconcerned with prediction accuracy on unseen or new data, then any one theory or model is as good as another, the decision-maker uses backward causality to justify any problems and definitely any successes, and so it goes in education as it goes in so many organizations across industries.

In the end, a core component and innovation of the application of data science to education organizational problems is a focus on prediction accuracy, providing expertise in data analysis and testing ideas, which feeds directly into evidence-based improvement cycles in collaboration with management. The argument here however is not that prediction accuracy is the only issue to focus on in data science, but that a closer balance between data-modeling and prediction accuracy is healthy for a domain, much healthier than the 98% to 2% split noted by Breiman (2001). Thus, I am not arguing that the majority of research should be prediction focused, but only for a reasonable balance. One reason to not focus exclusively on prediction accuracy over all others, is that, as I discuss in the next examples from multiple industries over the last decades, that path leads to biased algorithms that learn from the historic data, a history plagued with bias and segregation, providing back exactly the same problems, as the machines are quite good at learning from the training data provided.

Machine Learning Only Learns from Data, Data from a Flawed and Inequitable System

While prediction is a key aspect of data science, prediction based on machine learning without understanding of both the models and the bias of the data used to train the models is a core critique and a well-known downside across the data science, big data, and machine learning fields, including education (Baker & Hawn, 2021; Bowers, 2021b; Ifenthaler & Tracey, 2016; O'Neil, 2016; Wachter & Mittelstadt, 2019; Willis et al., 2016). Machine learning algorithms are only able to learn patterns from the data provided, and thus providing past education data, such as from the United States schooling system, which is a system heavily critiqued for its issues with systematic bias and segregation (Douglass Horsford, 2017; Hawn Nelson et al., 2020; Lomotey & Lowery, 2013; Noguera & Alicea, 2021; Palardy et al., 2015; Reardon, 2019), will lead to the algorithm learning and providing back exactly the same biases. As noted by Benjamin (2019) in discussing this issue in healthcare, as well as other industries:

> Data used to train automated systems are typically historic and, in the context of health care, this history entails segregated hospital facilities, racist medical curricula, and unequal insurance structures, among other factors. Yet many industries and organizations well beyond health care are incorporating automated tools, from education and banking to policing and housing, with the promise that algorithmic decisions are less biased than their human counterpart. But human decisions comprise the data and shape the design of algorithms, now hidden by the promise of neutrality and with the power to unjustly discriminate at a much larger scale than biased individuals (p. 422).

Given these issues across many industries, including education, recent research has begun to focus on algorithmic bias and fairness detection across machine learning domains, especially in education (Baker & Hawn, 2021; Corbett-Davies & Goel, 2018; Gardner et al., 2019; Hakimi et al., 2021; Hu & Rangwala, 2020; Khosravi et al., 2022; Lee et al., 2019; Loukina et al., 2019; Madaio et al., 2021; Paquette et al., 2020; Singh et al., 2019; Soland et al., 2020; Vaughan & Wallach, 2020; Williamson & Eynon, 2020). Across much of this literature, a primary tension is between the issue that algorithmic bias can be detected and corrected through more machine learning, versus the point that the implicit biases of the designers and coders of these systems unavoidably will make their way into the systems, as the design and code are designed and coded by humans. Thus, concurrently a growing set of literature also encourages focusing on working in collaboration with the leaders, communities, and stakeholders whom the system is designed to serve (Bowers, 2021a, 2021b; Bowers et al., 2019; Bowers & Krumm, 2021; Hawn Nelson et al., 2020; Krumm & Bowers, 2022; Lee, 2018). For example, Hawn-Nelson et al. (2020) provide a tool kit and exemplars from multiple school districts of data analysts and community stakeholders working together to build data systems that serve the community, as they state:

> We call for re-users of administrative data to center racial equity in data practice. We call for the inclusion of community voices and power sharing at every stage of design, use, and implementation. We call for relationship building among those represented in the data and those using the data. Without a deliberate effort to address structural racism, institutional racism, and unrecognized bias, data integration will inevitably reproduce and exacerbate existing harm (p. 3).

In this way, no longer is it the job of the data analysts and researchers to invite the communities to the table, as it is not the researchers' table. Rather, this literature calls for data science to work in collaboration and in concert with the communities that the systems serve, working to address the issues that the stakeholders designate and build action around addressing.

Applying the 4As to Early Warning Systems

Recently, to address these multiple issues, published by the Organisation for Economic Co-operation and Development (OECD), Bowers (2021b) proposed the 4 A's education algorithms framework of Accurate, Accessible, Actionable, and Accountable. Accurate in that at-risk and early warning predictor systems no longer focus on p-values and model fitting, but rather demonstrate that education early warning predictors are actually accurate in comparison to other current predictors in the same domain using signal detection theory (Bowers & Zhou, 2019). Accessible in that, for any algorithm that makes a recommendation or a decision on a student, while the data should be private and restricted, the code must be open and auditable. Just as any parent can ask a teacher how a decision was made about their student, or a parent can inspect the school budget or hire accountants to audit the budget, so too must algorithms be open to inspection, as the taxpayer paid for it and thus has the right to audit the code that makes such decisions or recommendations (Agasisti & Bowers, 2017). Proprietary and hidden code in education is unethical. This recommendation follows the guidance from the open science movement (Bezuidenhout et al., 2020; Molloy, 2011; Stodden et al., 2016), and open education science and research (Bowers et al., 2019; van der Zee & Reich, 2018), which also requires open code and algorithms to promote comparison, confirmation, replication, and extension of findings and prediction using the FAIR standards of Findable, Accessible, Interoperable, and Reusable (Bahim et al., 2020; Bowers & Choi, 2023, Bowers et al., 2022; Wilkinson et al., 2016). Actionable in that "actionable early warning indicators rely on predictors that are recent or real-time, malleable, and under the influence of stakeholders, in opposition to predictors that students, teachers, administrators, and family and community members have no control over" (Bowers, 2021b, p. 185), such as demographics, SES, and geographic location (Paquette et al., 2020). And Accountable, in that education algorithms are made in collaboration with the stakeholders whom the system serves, such as in following the recommendations of Hawn-Nelson et al. (2020) discussed above.

In relation to accuracy, an example of this work is Knowles (2015). In this study, the author analyzed entire student cohorts in Wisconsin using multiple machine learning methods to generate at-risk dropout predictors, but importantly compared the predictors using signal detection theory and Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) accuracy analysis (Bowers et al., 2013; Bowers & Zhou, 2019) to multiple other predictors already demonstrated in the literature. In the end, Knowles (2015) showed that none of the machine learning predictors were more accurate than either the Chicago on-track indicator (student ninth-grade failure in either mathematics or reading and student is behind in credits; Allensworth et al., 2018; Bowers, 2021b) or predictors such as pattern analysis of longitudinal noncumulative grade point averages (Bowers & Sprott, 2012; Brookhart et al., 2016). Additionally, as a policy document on the at-risk predictors in Wisconsin, Knowles's (2015) study also provided the full code in the open access R statistical language, in addition to providing the full details on the accuracy of the predictors, thus providing accuracy and accessibility around actionable predictors that then could potentially lead to increased accountability to the communities that the system serves, as the code and predictors are available for auditing, discussion, and critique. This example then exemplifies the potential of the 4As for algorithms in education, however given the above discussion that the vast majority of quantitative training in education is focused on model fitting and theory testing, I next turn to a discussion of the Common Task Framework (CTF) in data science as a framework to build capacity and innovation across this domain.

The Common Task Framework (CTF): Building Capacity in Data Science

In the machine learning engineering domain the Common Task Framework (CTF) has become a foundation of the industry, in which for a specific research or technical issue there is a large public dataset, open and shared code, an agreed upon task and outcome, and a means to algorithmically evaluate the outcome, with a neutral and public referee (Watts, 2017). Donoho (2017) has postulated that for data science, the CTF is the "secret sauce" that creates and incentivizes gradual, verifiable, and replicable gains toward an agreed-on goal in a research domain. On the history of CTF, Lieberman (2010) noted:

> Charles Wayne, starting a new speech-recognition program at DARPA in 1986, adopted the idea of quantitative comparison of alternative algorithms on a fixed task. In the context of this program, the formal quantitative competition was not only among algorithms but among research groups, with all the sites involved in the project sharing a predefined automatic evaluation metric and a body of material for training and testing. At first, DARPA's key motivation for this "common task method" seems mainly to have been to persuade skeptics that the new program would successfully avoid glamour and deceit . . . But it soon became clear that this approach had a number of key advantages as a method of research management. It lowered barriers to entry, by providing well-defined tasks and the resources needed to undertake them; it created a research community with shared goals and assumptions; and, perhaps most important, it offered proof of gradual progress that could be used to justify stable funding . . . over several decades, even when no "killer app" had yet emerged (p. 4).

Thus, machine learning has made rapid progress over the last decades, especially in domains such as NLP and image recognition with access to large public labeled training datasets, public code, and a means to publicly test and compare results against an agreed upon evaluation metric (De Cnudde et al., 2019; Donoho, 2017; Raghu & Schmidt, 2020), with authors in the computational data visualization and social science domains also nominating CTF as an attractive domain to adopt and pursue (Baker & Boser, 2021; Crisan et al., 2021; Watts, 2017) in the application of data science to research discovery, innovation, and organizational improvement.

Similarly, CTF is beginning to take shape within education data science, and especially through learning analytics. In the above example of at-risk predictor accuracy such as in the high school dropout research, there already are some of the components of CTF, including a common and open evaluation metric such as ROC AUC accuracy calculations and an increasing focus on publishing open and accessible code, such as in the R software packages (Bowers, 2021b; Bowers & Zhou, 2019; Coleman, 2021; Knowles, 2015). With open and accessible code and a common evaluation metric, at-risk as well as other indicators of student persistence and challenge in school can be replicated and compared. However, while this dropout flag and at-risk early warning system research domain has these two components of CTF, it lacks large open and public datasets, as well as neutral referees and a common and extensible data and code sharing framework and community.

Expanding the usefulness and opportunities of CTF is a central recommendation in learning analytics, educational data mining, and learning engineering (Baker & Boser, 2021). Learning analytics has seen some recent success in this area, such as in large scale datasets and competitions around MOOCs (Gardner et al., 2018), and large collections of learning analytics datasets (Liu et al., 2017). However, while there has been some progress, that progress has been uneven and sporadic, with much work remaining throughout the domain (Baker & Boser, 2021).

Looking forward toward the application of CTF in education data science to issues of education organizations and policy, bringing this literature together I encourage the domain to continue to build toward:

Open large-scale real-world deidentified datasets: These datasets should represent not only real-world student data, but also the types of data formats and structures that current K–12 and higher education data systems actually use, so that code generated on this data can be quickly deployed in current education data warehouse systems (Bowers, 2021c; Bowers et al., 2022; Khan, 2021; Pratt, 2021). Additionally, these datasets should also follow a common set of data standards, such as the U.S. Department of Education Common Data Standards (CEDS) (Hall & Huennekens, 2016), which aid in interoperability, data sharing, and data extraction, transformation, and loading (ETL). Yet, this call for data sharing and collaboration is not new, as for example in 2002 and 2004 the National Research Council and National Academy of Sciences listed data sharing as a core recommendation for improving education research to "enable reanalyses, replications, and further investigation with available data" Towne et al., 2004, p. 5). Such datasets could benefit as well from not being a sample but the entire census of full school systems, K–12 or PK–20 or beyond into careers, to aid in shared analysis across a large number of research and application questions through open data. While there are, of course, issues of data privacy, deidentification, and data confidentiality, these issues are not unique to education data, and many of the recent innovations in the medical, biological, and legal domains could be applied to these types of datasets (Bowers et al, 2022; Engelhardt et al., 2022). Additionally, potentially fruitful large datasets, while samples, do exist, such as the large nationally generalizable datasets from the U.S. National Center of Education Statistics as well as global datasets, such as through the OECD (Chudgar & Luschei, 2016; Hernández-Torrano &

Courtney, 2021; Urick, 2018). I encourage these datasets to be included within an emerging education data science Common Task Framework as these datasets represent a deep set of survey and assessment data over many decades that have not been a focus of a sustained and combined research effort around specific and actionable research and application prediction tasks.

A Shared Culture of Shared Code for Shared Research: The Common Task Framework requires a shared culture of sharing open access code, not only in running data visualizations, models, tests, and machine learning algorithms, but also in ETL. While there are few centralized opportunities for researchers in education to share open code around defined research issues, there are a small but growing set of examples (Baker & Boser, 2021). Cultivating and encouraging a culture of both shared data and interoperable and open code (Bahim et al., 2020; Engelhardt et al., 2022; van der Zee & Reich, 2018) helps address issues of replication (Watts, 2017) while also building capacity across the domain, lowering barriers of entry, and building a culture of shared and open responsibility and accountability around the predictions and results identified.

Public and Open Evaluation of Algorithms: As has been noted in the literature on the application of data science and machine learning to the social sciences (Watts, 2017), desired outcomes in many machine learning tasks to date are quantifiable and can be algorithmically scored, such as higher engagement in online recommendations, is a picture a picture of a cat, what is the next word in the sentence, or what is the answer to a specific set of questions. In contrast, many outcomes in the social sciences can be radically more nuanced and complex. That said, returning to the example of at-risk high school dropout status, a student either drops out or not, is absent or not, completes their intended major in college or not, clicks on the correct answer in an online tutor or not, and so on. These are good starting points for open and public evaluation of the results of studies that use open and shared code and datasets, as there are standard means to quantify the evaluation and compare outcomes, such as ROC AUC, precision-recall, and the like (Bowers & Zhou, 2019). Thus, rather than an ever-continuing stream of difficult to compare studies that focus on model fitting, p-values, and unitless effect size measures, prediction accuracy could be evaluated and gradually improved on using CTF, while the mechanisms for how the prediction takes place, and mitigating the biases around those predictions are studied and understood.

Neutral Referees: A central innovation of CTF is that predictions can be submitted on a regular basis and algorithmically judged by a neutral third party. This vital work is important, and should itself be a collaborative effort, perhaps centralized within conferences, or perhaps by communities with whom the algorithm is being designed, as ceding this work to for-profit companies, journals, or even nonprofit journals is a conflict of interest for the journals, as they also promote (and create revenue from) the papers that include the top algorithms.

Data Science as a Third Methodology in Education Research

Theory and the literature review have traditionally been at the center of the intersection of quantitative and qualitative research methods in education, with data science now joining as an additional methodology in education research that builds on the expanding definitions of the variety of data sources that can be analyzed, research questions, and analysis methods (Agasisti & Bowers, 2017; Bowers, 2017; Piety & Pea, 2018; Singer, 2019). As described in Figure 5, while inferential quantitative methods aim to generalize to the population, and qualitative methods may generalize to theory, data science provides a means to bring both tabular data (traditionally the domain of quantitative methods) together with unstructured data, such as text, (traditionally the domain of qualitative methods) to examine new types of research questions. Data science pattern analytic techniques that bring together both tabular structured data and unstructured text data provide a means to "make visible data that have heretofore gone unseen, unnoticed, and therefore unactionable" (Bienkowski et al., 2012, p. ix).

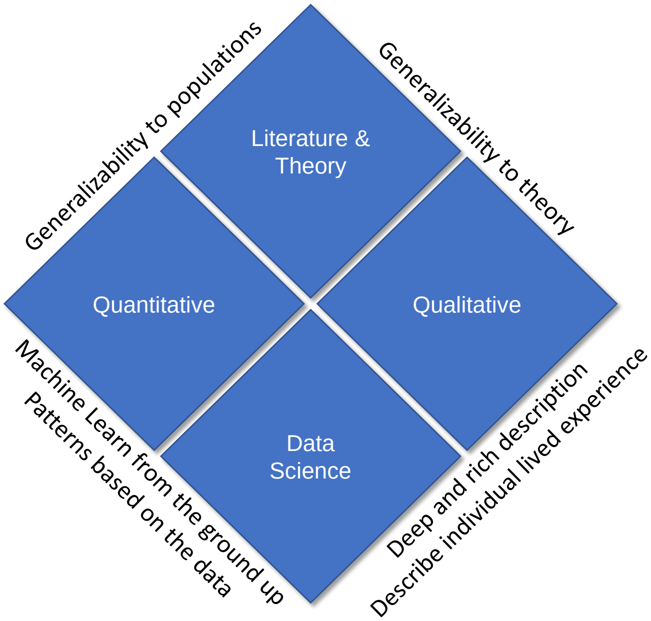

Figure 5: Data Science as a Central Method in Education Research

Thus, the work of education data science is to bring together many different data sources that may include data from logfile clickstreams, dashboards, devices, geographic data, large data matrices or databases, classic tabular structured data in individual flat files, as well as unstructured data, such as text, images, audio files, and video.

Spatial-Temporal Data

For example, in a study by Coleman et al. (2021), 80 participants in an educator professional development co-design collaborative event were each provided a location tracking device (with consent) in a large collaborative meeting space, placed into teams of about seven or eight participants, with a set of design-based collaborative data visualization tasks to discuss and generate, and then the location data for each person on each team were tracked over an entire day of a workshop (Coleman et al., 2021). This longitudinal dataset represented a very large dataset as the location of each participant in the collaboration space was recorded every few seconds. The location data were then merged with data such as who was on which team, and then team cohesion over the workshop day was analyzed over time to provide longitudinal visualizations and analysis of the temporal–spatial data to map how teams interacted within and among each other over time on a specified professional development task.

A second example involves data collected from a collaborative interactive science museum exhibit in which middle school students interacted with physical blocks that represent different concepts in creating electrical circuits on a physical table with a virtual overlay of a game space to capture virtual exotic fish using the blocks to create virtual circuits within the game, with the purpose of students learning the fundamentals of electrical circuit design through constructivist cooperative play (Jorion et al., 2020). Here the data represented the per second location data of where each block was on the surface of the two-dimensional table with the virtual display over many play sessions with many blocks in which one to up to four total students could stand at the table and participate by moving the physical blocks on the virtual tabletop. By analyzing this dataset, not only could individual players be detected, but the analysis demonstrated that when the table hosted the maximum four players, together the students collaboratively constructed many more circuits and persisted in the educational game much longer than when there were fewer students playing, which was an intentional part of the constructivist collaborative design of the game (Tissenbaum, 2020; Tissenbaum et al., 2017). These two examples demonstrate the usefulness of data science types of analysis for these nonstandard types of data that do not necessarily have participants in rows and variables in columns (tabular data), but instead include spatial and temporal data that can be analyzed and visualized in new and interesting ways to approach research questions from new angles.

Text as Data

One type of data that has received quite a lot of attention in data science and machine learning analysis methods over the last two decades is text data (Grimmer et al., 2022; Srivastava & Sahami, 2009). Traditionally the domain of qualitative research, text data mining methods have grown substantially over the last few years and are now used across a variety of industries. For example, in learning analytics, the number of papers using text data mining methods is up over tenfold in the last decade (Ferreira-Mello et al., 2019), and include techniques such as clustering of text into topics, NLP to detect student sentiment or correct/incorrect answers in essays, organizing documents to aid in information retrieval, and automatically summarizing text.

Moving beyond summarization word frequency counts, text mining as a research method provides a means to cluster and organize the text across a corpus in a variety of ways that begins to get at the meaning, latent topics, and latent semantics of words, phrases, and documents. For example, sentiment analysis aims to detect emotion and the sentiment of words used (Liu & Zhang, 2012), such as inferring that a sentence is happy or sad based on the word correlation frequencies with lists of keywords that indicate these sentiments or that has been machine learned to detect this sentiment with high accuracy (Soleymani et al., 2017). A well-known example in education and learning analytics is Coh-Metrix (Graesser et al., 2011), which automatically detects the level of cohesion in texts aligned to theories of language level and discourse across grade levels, including word concreteness, syntactic simplicity, referential cohesion, causal cohesion, and narrativity, and has been shown to be useful in classifying and organizing student reflections and discussion board forum posts, for instance from online undergraduate courses (Kovanović et al., 2016, 2018).

Another form of text data mining is classifying documents based on latent topics identified through word correlation frequencies, commonly referred to as correlated topic models (CTM; Blei & Lafferty, 2007, 2009; Blei et al., 2003; Jordan & Mitchell, 2015; Steyvers & Griffiths, 2007). In these models, each document is considered a “bag of words” in that a document consists of a set of terms (words) regardless of the order in the sentences. Then, through these models using latent Dirichlet allocation (LDA):

> LDA assumes that the words of each document arise from a mixture of topics, where each topic is a multinomial over a fixed word vocabulary. The topics are shared by all documents in the collection, but the topic proportions vary stochastically across documents, as they are randomly drawn from a Dirichlet distribution (Blei & Lafferty, 2007, p. 18).

Following on this work, there are now many open access coding packages, such as in the R statistical language, that apply different varieties of topic modeling in this way (Chang, 2015; Grün & Hornik, 2011; Silge & Robinson, 2017; Steyvers & Griffiths, 2007). Briefly, the process proceeds first with cleaning a corpus, that is, first removing numbers, any embedded code, and then stemming and removing words such as all words with three or fewer letters to remove stop words and basic articles of speech (such as “a,” “the,” “is,” etc.). Then a term-by-document matrix is created in which the frequency of every term in the corpus is listed for each document. The topic model then creates two new matrices, terms by topics, and topics by documents, in which the latent topics are identified through the “story” told by the list of the highest occurring terms (similar in some ways to how one tells the "story" of a factor through individual item factor loadings in factor analysis), and each document has a probability to then be in any of the identified topics, usually classifying the document into the highest probability topic (Blei, 2012; Blei & Lafferty, 2007, 2009).

One of the earliest examples of text data mining using topic modeling in education finance and policy research was by Bowers and Chen (2015) in which we used an automated text mining correlated topic model to analyze the full text of n = 1,210 individual capital facility finance school district bond proposals in Michigan over 16 years, as in states such as Michigan, districts may propose to build, renovate, or purchase capital facilities using debt paid back through increased local property taxes passed through a bond referendum. The bond ballot proposals clustered into nine distinct topics related to the bond ballot request, such as transportation (with high frequency terms such as bus, busing, transportation), athletics (football, gym, track), building new schools, or renovating schools, and the like, with bond proposals such as athletics passing much less often (Bowers & Chen, 2015).

A second example uses topic modeling to cluster school improvement plans and implementation report tasks from all schools in the state of Washington from 2011–2012 to 2015–2016, analyzing about 24,000 individual tasks noted across 2,873 school improvement plans (Sun et al., 2019). Using LDA, the authors identified a 20-topic model and merged relevant panel data such as test scores and average student absenteeism. The model identified multiple improvement plan topics, such as "teacher team activities focused on reviewing data, planning, aligning standards, developing interventions" and "administering common assessments and disaggregating data to differentiate interventions," each of which was negatively correlated with high student absence rates (interestingly this second topic was one of the lowest topic proportions across the dataset for the state). These same topics as well as others such as setting goals for recognizing and monitoring teachers' and students' growth were positively correlated with average school-level mathematics standardized test scores. Thus, automated text data mining can be a useful means to empirically describe not only which documents cluster together based on word correlation frequencies, but also to provide the proportions of documents that cluster into each of the topics, along with topic probabilities, each of which describes a rich set of information that is unavailable to researchers without these types of text pattern analytics.

Text data mining also has been shown to be useful in organizing and clustering literature by latent topics, such as Wang et al. (2017) who analyzed the full text of all 1,539 articles published in the journal Educational Administration Quarterly (EAQ), from its inception in 1965 to 2014 (Wang et al., 2017). EAQ is especially of interest as it sits at the center of the education leadership research citation network in English globally as the most prominent journal in the domain when analyzing all citations across articles in the 40 top journals in school leadership research (Wang & Bowers, 2016). Following Blei and Lafferty's (2007) correlated topic model text mining analysis of the journal Science and the growing domains of the digital humanities (Berry, 2012; Jockers & Underwood, 2015; Vanhoutte, 2016) and computational social science (Alvarez, 2016; Lazer et al., 2009), Wang et al. (2017) applied text mining to the entire corpus of all articles in EAQ, identifying 19 topics across the 50 years, with the top two most prevalent topics being the "epistemology of educational leadership" (12.7% of articles) and "research reviews and reflections" (10.7%). The authors then plotted the rise and fall of topics over the 50 years, and found that while many topics grew over the years, such as "teaching and instructional leadership," and "inequities and social justice," others died off, including "epistemology of educational leadership," visualizing the rise and fall of the central topics of concern to the field over time. Interestingly, the topic model automatically identified special issues, such as a spike in interest in the topic of "trust" in school leadership for the year in which there was a special issue, as the algorithm was not given the information that there had been a special issue, but because topic modelling relies on word correlation frequencies, special issues of journals stand out as similar words are used across multiple articles. Thus, text mining topic models provide a useful means to cluster and organize a corpus of text empirically and provide a new and interesting set of tools and lenses to analyze patterns in large volumes of text that were previously inaccessible.

A final example of the usefulness of automated text mining is in working to empirically group discussion group participants in classrooms, professional development workshops, and online or large-scale courses through clustering students together into discussion groups based on the word correlation frequencies patterned using correlated topic modeling of papers or surveys from the start of class or a survey before or after a workshop event (Bowers, 2021c; Bowers et al., 2021; Gegenheimer, 2021; Kang & Bowers, 2021). For example, in Bowers et al. (2021), for a four-week online noncredit professional development course for teachers and school administrators globally on using evidence and data to inform decision-making and instructional improvement in schools, we used topic modeling of participants' reflection papers in the first week of the course on successes and challenges with evidence use in their school to cluster participants into subsequent online discussion board groups. As the central work of the course was in networking and interacting through the discussion boards, participants were grouped into discussion board groups of 10 based on the word-correlation frequencies from the automated topic modeling, and so thus for the rest of the course, no matter if there were a hundred or more participants, discussion board groups of 10 would only interact with their discussion board group and literally shared a common language in the group based on the word correlation frequency clustering (Bowers et al., 2021). Thus, the intention of this use of text mining in workshops and courses (Bowers, 2021c; Gegenheimer, 2021; Kang & Bowers, 2021) is to use the actual words written by participants to empirically and informatively group them for interactions rather than rely on happenstance or randomization for grouping participants. In this way, event orchestrators and instructors can informatively group participants either by clustering them together using similarity of language or ensuring that each group has a diversity across the clusters identified by the topic model.

Conclusion and a Look to the Future

In conclusion, data science in education research, policy, and practice is an emerging and exciting domain with a wide and ever-growing list of applications. While data science has existed for more than 50 years, the core concerns of data science of description, patterning, and prediction have been at the forefront of management, administration, and organizational theory in education and K–12 schooling for the same 50 years. While definitions of data science continue to evolve, at its core:

> Data science is more than the combination of statistics and computer science—it requires training in how to weave statistical and computational techniques into a larger framework, problem by problem, and to address discipline-specific questions (Blei & Smyth, 2017, p. 8691)

It is these discipline-specific questions that are of core concern to policymakers and practitioners, and the techniques and application of data science across education organizations has the potential to provide a much deeper and nuanced view into the patterns and predictive accuracy of management and policy decisions.

Mirroring the calls noted throughout the previous discussion for an increased focus on the emerging domain of education data science in education policy and practice, related social sciences with strong quantitative methods and traditions, such as geography and GIS (geographic information systems), are making similar contemporary calls that echo and mirror the same arguments outlined throughout the present paper. For example, in their recent article titled “Geographic Data Science” Singleton and Arribas-Bel (2021) worked to define the term geographic data science and stated:

> there needs to be intensified critical engagement of Data Science by geographers . . . The long interdisciplinary tradition that exists within Geography makes it particularly well positioned to facilitate such engagement. At the same time, further interaction with Data Science will bring new methodological tools that can help Geography, and the Geographical Analysis community, to remain relevant in an increasingly data-driven and digital world . . . To realize such a vision and to foster interaction, we propose the term of Geographic Data Science (p. 62).

Thus, the argument by these authors for a deeper engagement for GIS and data science mirrors the same argument I’ve made throughout this chapter for a deeper engagement in education, organization, and leadership and quantitative and statistics research with the tools and techniques of data science to “remain relevant in an increasingly data-driven and digital world” (Singleton & Arribas-Bel, 2021, p. 62). This argument of course is not unique to these two fields. The call for applying data science methods is a contemporary concern across a wide range of domains (Donoho, 2017; Neff et al., 2017; Provost & Fawcett, 2013; Ullman, 2020); many fields are at similar points as education research in these discussions of applying data science to issues of research and practice across each domain.

Looking toward the future, the inclusion of data science as the third methodology in education research, as I’ve proposed here, is an exciting prospect. Integrating a deeper set of tools from data science will provide the potential to build exciting new techniques, innovations, and discoveries across the traditional qualitative and quantitative (statistics) methods. In this chapter, I have briefly noted three potential future areas. First, for instance, education research has a deep history and strong findings in psychometrics and survey validity studies, from designing and testing individual assessment items at the student level (Furr, 2021; Haertel, 2013; Popham, 2016), to validating factors, to hierarchical and structural models that delve deeply into the complex relationships in classrooms, schools, and communities (Raudenbush & Bryk, 2002), to cross-nation comparisons using the well-known international tests such as PISA and TIMSS, and surveys such as the Teaching And Learning International Survey (TALIS), among many others (Chudgar & Luschei, 2016; Urick, 2018). Recently, for example, computational psychometrics has emerged as an innovative framework to address new issues in testing and assessment through using data science tools, as digital assessment systems create ever larger volumes and variety of data (Hao & Mislevy, 2021), mirroring many of the innovations emerging from learning analytics (Baker et al., 2021; Du et al., 2021; Ifenthaler, 2021; Kizilcec & Davis, 2022). This work brings together large expansive datasets, including assessments, clickstream logfiles, student writing, latent factors, combined with deep and complex interrelated and nested hierarchical data structures. As education systems globally continue to collect an increasing volume and variety of data at ever increasing scales, the application of data science tools across the traditional education research and policy assessment, psychometrics, and survey domains is an exciting area to watch for the emergence of next-stage methods, innovations, and actionable findings for research, policy, and practice.

A second area to highlight in the application of data science as a third method in education research is in deeper engagement with nontabular data, such as text, images, video, and qualitative types of data. As noted above, in the domain of text data, there have been many advances recently in analyzing student essays as well as other forms of text data mining that help bring together data science tools in education research. Across schooling systems, nontabular and more free-form data abounds, from the mountains of student essays and papers, to the large range of documents that schools, districts, and governments create, to moment-by-moment formal and informal interactions, to how teachers and students move, where they look, and what they say in classrooms, among many others. As a goal in qualitative research is to surface and privilege the voices of participants from these many different types of data (Denzin & Lincoln, 2000), so too must it be a goal of the application of data science to these types of data. Integrating theory, qualitative methods, and data science to give greater access to both small and large-scale sets of these types of data has the potential to provide a wealth of new insights and techniques that can further elevate the stories and lived experiences of students and families within school systems, and surface currently unknown and unidentified patterns across the system. This type of work should also help to address the fourth "A" of Accountability in the 4As framework for education algorithms (Bowers, 2021b) as researchers work in collaboration with communities and stakeholders, working to elevate and surface the voices and needs of individual communities for which these education systems are built and serve.

And finally to conclude, a third area to highlight in looking to the future of the integration of data science into the methodological lexicon of education research, policy, and practice, is to return to first the "A" in the 4As of education algorithms (Bowers, 2021b): Accuracy. As noted at the beginning of this paper and throughout the application of data science to organizational practice, a core focus of the work of data science and machine learning is to make accurate predictions, testing the outcomes of models against new, naïve, or holdout group data that the model has never seen (Breiman, 2001; Donoho, 2017; Singer, 2019). Combine with this the salient points that managers, leaders, administrators, and policymakers are most concerned with the accuracy of predictions of models and research, and mostly have little interest in theory testing and model fit. Throughout this paper I have worked to make the case in education research for the elevation of accuracy of predictions as a core focus of education research and policy, on par with theory testing and model fitting. The percentage of articles in education research outside of learning analytics that report accuracy is exceedingly small, as evidenced by examining any education research journal. Throughout education research, the focus has been, and is currently on p-values, variance explained, and effect sizes with much consternation about the "replication crisis" in psychology and thus also education research (Wagenmakers et al., 2021; Wicherts et al., 2016; Yarkoni, 2022). Yet, for the vast majority of quantitative articles, authors fit a model to all of their data, report the model fit and p-values in a paper or report, and call it a day, and the vast industry of education sees yet another irrelevant study that does not bridge theory to practice. But imagine instead that accuracy, as argued throughout data science, is just as important, or more so, than model fit, especially when it comes to the application of research to the practice of organizing, leading, and running schools and schooling systems. Perhaps knowing the accuracy of the predictions that a model from research can make, in which those accuracy calculations are generated from data that the model had never seen, either from a hold-out sample or future data (train on last year's data, test on this year's data), would be more helpful to practitioners than the vast amount of text devoted in education research and policy to discussing the extent that p-values are significant, or that unitless effect sizes are relevant to practice. As a data science practice that is well-known and well-tested, this is an innovation that can be implemented immediately, and could have far reaching impacts on research, policy, and practice. Combining this accuracy focus within the larger CTF outlined in this chapter and throughout the data science literature along with open public large-scale real-world deidentified datasets, a shared culture of shared code for shared research, and public and open evaluation of algorithms, has the potential to lead to not only new strides in research methods and findings, but just as importantly application to practice and policy that matters to stakeholders across education systems.

Suggested Citation:

Bowers, A.J. (2025) Data Science in Education Administration, Policy, and Practice. In L. Cohen-Vogel, J. Scott, P. Youngs (Eds.) *AERA Handbook of Education Policy Research*, 2nd Edition. (pp. 585-612). American Educational Research Association.

References

Agasisti, T., & Bowers, A. J. (2017). Data analytics and decision-making in education: Towards the educational data scientist as a key actor in schools and higher education institutions. In G. Johnes, J. Johnes, T. Agasisti, & L. López-Torres (Eds.), *Handbook on the Economics of Education* (pp. 184–210). Edward Elgar Publishing. https://doi.org/10.7916/D8PR95T2

Allensworth, E. M., Nagaoka, J., & Johnson, D. W. (2018). High school graduation and college readiness indicator systems: What we know, what we need to know. UChicago Consortium of School Research. https://consortium.uchicago.edu/publications/high-school-graduation-and-college-readiness-indicator-systems-what-we-know-what-we

Alvarez, R. M. (2016). *Computational social science.* Cambridge University Press.

Athey, S., & Imbens, G. W. (2015). Machine learning methods for estimating heterogeneous causal effects. *stat*, 1050(5), 1–26. https://gsb-faculty.stanford.edu/guido-w-imbens/files/2022/04/3350.pdf

Bahim, C., Casorrán-Amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., Russell, K., & Stall, S. (2020). The FAIR Data Maturity Model: An approach to harmonise FAIR assessments. Data Science Journal, 19(1), 41. http://doi.org/10.5334/dsj-2020-041

Baker, R. S. (2013). Learning, schooling, and data analytics. In M. Murphy, S. Redding, & J. Twyman (Eds.), *Handbook on* *innovations in learning.* (pp. 179–190). Information Age Publishing.

Baker, R. S. (2021). Artificial intelligence in education: Bringing it all together. In *OECD Digital Education Outlook 2021: Pushing the Frontiers with AI, Blockchain, and Robots*. (pp. 43-54) OECD Publishing.

Baker, R. S., & Boser, U. (2021). High-leverage opportunities for learning engineering. Penn Center for Learning Analytics. https://www.upenn.edu/learninganalytics/Learning_Engineering_recommendations.pdf

Baker, R. S., Gašević, D., & Karumbaiah, S. (2021). Four paradigms in learning analytics: Why paradigm convergence matters. Computers and Education: *Artificial Intelligence*, 2, 100021. https://doi.org/https://doi.org/10.1016/j.caeai.2021.100021

Baker, R. S., & Hawn, A. (2021 , March 1). Algorithmic bias in education. EdArXiv Preprints. https://doi.org/10.35542/osf.io/pbmvz

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning Analytics* (pp. 61–75). Springer. http://link.springer.com/chapter/10.1007%2F978-1-4614-3305-7_4

Baker, R. S., & Koedinger, K. R. (2018). Towards demonstrating balue of learning analytics for K–12 education. In D. Niemi, R. D. Pea, & B. Saxberg (Eds.*), Learning Analytics in Education* (pp. 49–62). Information Age Publishing.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–16. http://www.educationaldatamining.org/JEDM/images/articles/vol1/issue1/JEDMVol1Issue1_BakerYacef.pdf

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2), 131–160. https://doi.org/10.1037/1082-989X.2.2.131

Behrens, J. T., DiCerbo, K. E., Yel, N., & Levy, R. (2012). Exploratory data analysis. In I. Weiner, J. A. Schinka, & W. F. Velicer (Eds.) *Handbook of Psychology*, Second Edition (Vol. 2, Foundations of Research Issues, 1). https://doi.org/https://doi.org/10.1002/9781118133880.hop202002

Benjamin, R. (2019). Assessing risk, automating racism. *Science*, 366(6464), 421–422. https://doi.org/10.1126/science.aaz3873

Benoliel, P., & Berkovich, I. (2020). Learning from intelligent failure: an organizational resource for school improvement. *Journal of Educational Administration*, 59(4), 402–421. https://doi.org/10.1108/JEA-07-2020-0155

Berry, D. M. (2012). Introduction: Understanding the digital humanities. In D. M. Berry (Ed.*), Understanding Digital Humanities* (pp. 1–20). Palgrave Macmillan UK. https://doi.org/10.1057/9780230371934_1

Bezuidenhout, L., Quick, R., & Shanahan, H. (2020). Ethics when you least expect it: A modular approach to short course data ethics instruction. *Science and Engineering Ethics*, 26(4), 2189–2213. https://doi.org/10.1007/s11948-020-00197-2

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED611199.pdf

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. http://www.cs.princeton.edu/~blei/papers/Blei2012.pdf

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 1(1), 17–35. https://doi.org/10.2307/4537420

Blei, D. M., & Lafferty, J. D. (2009). Topic models. Text mining: classification, clustering, and applications, 10, 71–86. http://www.seas.harvard.edu/courses/cs281/papers/blei-lafferty-2009.pdf

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. http://dl.acm.org/citation.cfm?id=944937

Blei, D. M., & Smyth, P. (2017). Science and data science. Proceedings of the National Academy of Sciences, 114(33), 8689–8692. https://doi.org/10.1073/pnas.1702076114

Bloom-Weltman, J., Honey, R., Meholick, S., & Fettro, M. (2021). Profile of state data capacity in 2018: Statewide longitudinal data systems (SLDS) survey descriptive statistics. Institute of Education Sciences, Stats in Brief, NCES 2021–126. https://nces.ed.gov/pubs2021/2021126.pdf

Bowers, A. J. (2017). Quantitative research methods training in education leadership and administration preparation programs as disciplined inquiry for building school improvement capacity. *Journal of Research on Leadership Education*, 12(1), 72–96. https://doi.org/10.1177/1942775116659462

Bowers, A. J. (2021a). Dashboards, data use, and decision-making—A data collaborative workshop bringing together educators and data scientists. In A. J. Bowers (Ed.), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists* (pp. 1–36). Teachers College, Columbia University. https://doi.org/10.7916/d8-jj2g-e225

Bowers, A. J. (2021b). Early warning systems and indicators of dropping out of upper secondary school: The emerging role of digital technologies. In *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*. OECD Publishing. https://doi.org/10.1787/589b283f-en

Bowers, A. J. (2021c). Planning, organizing, and orchestrating the education data collaborative workshop. In A. J. Bowers (Ed.), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists* (pp. 37–67). Teachers College, Columbia University. https://doi.org/10.7916/d8-jj2g-e225

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. Teachers College, Columbia Unviersity. https://doi.org/10.7916/d8-31a0-pt97

Bowers, A. J., & Chen, J. (2015). Ask and ye shall receive? Automated text mining of Michigan capital facility finance bond election proposals to identify which topics are associated with bond passage and voter turnout. *Journal of Education Finance*, 41(2), 164–196.

Bowers, A. J., & Choi, Y. (2023). Building School Data Equity, Infrastructure, and Capacity Through FAIR Data Standards: Findable, Accessible, Interoperable, and Reusable. *Educational Researcher*, 52(7), 450–458. https://doi.org/10.3102/0013189X231181103

Bowers, A. J., Choi, Y., Shi, H., Sun, F., Sun, H., Williams, J., & Weeks, S. (2022). Mapping 16 equity indicators to the New York City schools public datasets: Access, opportunity, and outcomes. Black Education Research Collective. Teachers College, Columbia University. https://doi.org/10.7916/a143-aq05

Bowers, A. J., & Krumm, A. E. (2021). Supporting the initial work of evidence-based improvement cycles through a data-intensive partnership. *Information and Learning Sciences*, 112(9/10), 629–650. https://doi.org/10.1108/ILS-09-2020-0212

Bowers, A. J., Pekcan, B., & Pan, Y. (2021). Grouping online education leadership professional development discussion board participants using automated text data mining Annual meeting of the University Council for Educational Administration, Columbus OH.

Bowers, A. J., & Sprott, R. (2012). Examining the multiple trajectories associated with dropping out of high school: A growth mixture model analysis. *Journal of Educational Research*, 105(3), 176–195. https://doi.org/10.1080/00220671.2011.552075

Bowers, A. J., Sprott, R., & Taff, S. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. *The High School Journal*, 96(2), 77–100. http://muse.jhu.edu/journals/high_school_journal/v096/96.2.bowers.html

Bowers, A.J., Zhao, Y., & Ho, E. (2022). Towards hierarchical cluster analysis heatmaps as visual data analysis of entire student cohort longitudinal trajectories and outcomes from grade 9 through college. *The High School Journal*, 106(1), 5–36. https://doi.org/10.1353/hsj.2022.a906700

Bowers, A. J., & Zhou, X. (2019). Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes. *Journal of Education for Students Placed at Risk* (JESPAR), 24(1), 20–46. https://doi.org/10.1080/10824669.2018.1523734

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), 199–231, 133. https://doi.org/10.1214/ss/1009213726

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and balue in the most common educational measure. *Review of Educational Research*, 86(4), 803–848. https://doi.org/10.3102/0034654316672069

Bruno, J. E., & Fox, J. N. (1973). Quantitative Analysis in Educational Administrator Preparation Programs. The ERIC Clearinghouse on Educational Management, University Council for Educational Administration. http://files.eric.ed.gov/fulltext/ED082270.pdf

Bulut, O., & Desjardins, C. (2019). Exploring, visualizing, and modeling big data with R. Author. https://okanbulut.github.io/bigdata/

Chang, J. (2015). Package ‘lda’: Collapsed Gibbs sampling methods for topic models. Version 1.4.2. (open access coding package). https://cran.r-project.org/web/packages/lda/lda.pdf

Chudgar, A., & Luschei, T. F. (2016). The untapped promise of secondary data sets in international and comparative education policy research. *Education Policy Analysis Archives*, 24(113), 1–16. http://epaa.asu.edu/ojs/article/view/2563

Clemson, B. (1978). Beyond management information systems. *Educational Administration Quarterly*, 14(3), 13–38. https://doi.org/doi:10.1177/0013161X7801400305

Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review*, 69(1), 21–26. https://doi.org/https://doi.org/10.1111/j.1751-5823.2001.tb00477.x

Cohen-Vogel, L., Tichnor-Wagner, A., Allen, D., Harrison, C., Kainz, K., Socol, A. R., & Wang, Q. (2015). Implementing educational innovations at scale: Transforming researchers into continuous improvement scientists. *Educational Policy*, 29(1), 257–277. https://doi.org/doi:10.1177/0895904814560886

Coleman, C. J. (2021). Exploring a generalizable machine learned solution for early prediction of student at-risk status. [Unpublished doctoral dissertation.] Columbia University. https://doi.org/10.7916/d8-5scb-n214

Coleman, C. J., Lutz-Coleman, L., Coleman, J., & Bowers, A. J. (2021). Understanding workshop participant movement through a temporal cluster analysis. In A. J. Bowers (Ed.), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists* (pp. 121–148). Teachers College, Columbia University. https://doi.org/10.7916/d8-jj2g-e225

Colson, E., Sibley, D., & Spiegel, D. (2021). The sobering truth about the impact of your business ideas. Radar/Business. https://www.oreilly.com/radar/the-sobering-truth-about-the-impact-of-your-business-ideas/

Conway, D. (2010). The data science Venn diagram. Author. http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Cope, B., & Kalantzis, M. (2016). Big data comes to school: Implications for learning, assessment, and research. *AERA*

Open, 2(2), 2332858416641907. https://doi.org/10.1177/2332858416641907

Corbett-Davies, S., & Goel, S. (2018). The measure and mismeasure of fairness: A critical review of fair machine learning. Stanford University. https://5harad.com/papers/fair-ml.pdf

Crisan, A., Fiore-Gartland, B., & Tory, M. (2021). Passing the data baton: A retrospective analysis on data science work and workers. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1860–1870. https://doi.org/10.1109/TVCG.2020.3030340

Crisan, A., & Munzner, T. (2019). Uncovering data landscapes through data reconnaissance and task wrangling. *Proceedings of the 2019 IEEE Visualization Conference* (VIS; pp. 46-50). https://ieeexplore.ieee.org/document/8933542

De Cnudde, S., Ramon, Y., Martens, D., & Provost, F. (2019). Deep learning on big, sparse, behavioral data. *Big Data*, 7(4), 286–307. https://doi.org/10.1089/big.2019.0095

Demchenko, Y., Grosso, P., de Laat, C., & Membrey, P. (2013). Addressing big data issues in Scientific Data Infrastructure. *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems* (CTS; pp. 48–55). https://ieeexplore.ieee.org/document/6567203

Denzin, N. K., & Lincoln, Y. S. (2000). Introduction: The discipline of practice and qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 1–28). Sage.

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), 745–766. https://doi.org/10.1080/10618600.2017.1384734

Douglass Horsford, S. (2017, 2017/04/03). A race to the top from the bottom of the well? The paradox of race and U.S. education reform. *The Educational Forum*, 81(2), 136–147. https://doi.org/10.1080/00131725.2017.1280754

Du, X., Yang, J., Shelton, B. E., Hung, J.-L., & Zhang, M. (2021). A systematic meta-review and analysis of learning analytics research. *Behaviour & Information Technology*, 40(1), 49–62. https://doi.org/10.1080/0144929X.2019.1669712

Engelhardt, C., Biernacka, K., Coffey, A., Cornet, R., Danciu, A., Demchenko, Y., Downes, S., Erdmann, C., Garbuglia, F., Germer, K., Helbig, K., Hellström, M., Hettne, K., Hibbert, D., Jetten, M., Karimova, Y., Hansen, K., Kuusniemi, K., Elisa, M., . . . & Zhou, B. (2022). D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions (V1.1 DRAFT). EU Open Research Repository. https://doi.org/10.5281/zenodo.5837500

Faria, A.-M., Sorensen, N., Heppen, J., Bowdon, J., Taylor, S., Eisner, R., & Foster, S. (2017). Getting students on track for graduation: Impacts of the Early Warning Intervention and Monitoring System after one year. IES Regional Educational Laboratory Program. https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=388

Farley-Ripple, E. N., Jennings, A., & Jennings, A. B. (2021). Tools of the trade: a look at educators' use of assessment systems. *School Effectiveness and School Improvement*, 32(1), 96–117. https://doi.org/10.1080/09243453.2020.1777171

Farmer, J. (1970). Why planning, programming, budgeting systems for higher education? Western Interstate Commission for Higher Education. http://eric.ed.gov/?id=ED063860

Feng, S., & Law, N. (2021). Mapping artificial intelligence in education research: A network-based keyword analysis. *International Journal of Artificial Intelligence in Education*, 31, 277–303. https://doi.org/10.1007/s40593-021-00244-4

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *WIREs Data Mining and Knowledge Discovery*, 9(6), e1332. https://doi.org/https://doi.org/10.1002/widm.1332

Figlio, D., & Loeb, S. (2011). School accountability. In E. A. Hanushek, S. J. Machin, & L. Woessmann (Eds.) *Handbook of the Economics of Education*, (Vol. 3, pp.383–421). North Holland.

Firestone, W. A., & González, R. A. (2007). Culture and processes affecting data use in school districts. *Teachers College Record*, 109(13), 132–154. https://doi.org/10.1177/016146810710901308

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, 44(1), 130–160. https://doi.org/10.3102/0091732x20903304

Furr, R. M. (2021). *Psychometrics: An introduction*. SAGE publications.

Fyfe, E. R., de Leeuw, J. R., Carvalho, P. F., Goldstone, R. L., Sherman, J., Admiraal, D., Alford, L. K., Bonner, A., Brassil, C. E., Brooks, C. A., Carbonetto, T., Chang, S. H., Cruz, L., Czymoniewicz-Klippel, M., Daniel, F., Driessen, M., Habashy, N., Hanson-Bradley, C. L., Hirt, E. R., . . . & Motz, B. A. (2021). ManyClasses 1: Assessing the generalizable effect of immediate feedback versus delayed feedback across many college classes. *Advances in Methods and Practices in Psychological Science*, 4(3), 25152459211027575. https://doi.org/10.1177/25152459211027575

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. https://doi.org/http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007

Gans, J. S., Stern, S., & Wu, J. (2019). Foundations of entrepreneurial strategy. *Strategic Management Journal*, 40(5), 736–756. https://doi.org/https://doi.org/10.1002/smj.3010

Gardner, J., Brooks, C., Andres, J. M., & Baker, R. S. (2018, 10–13 Dec. 2018). MORF: A framework for predictive modeling and replication at scale with privacy-restricted MOOC data. *Proceedings of 2018 IEEE International*

Conference on Big Data (Big Data; pp. 3235–3244). https://ieeexplore.ieee.org/document/8621874

Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. (pp. 225–234). https://doi.org/10.1145/3303772.3303791

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gegenheimer, K. (2021). Challenges and successes in education leadership data analytics collaboration: A text analysis of participant perspectives. In A. J. Bowers (Ed.), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists* (pp. 110–120). https://doi.org/10.7916/d8-jj2g-e225

Ghosh, S., Thomke, S., & Pourkhalkhali, H. (2020). The effects of hierarchy on learning and performance in business experimentation. *Academy of Management Proceedings*, 2020(1), 20500. https://doi.org/10.5465/ambpp.2020.159

Gleason, P., Crissey, S., Chojnacki, G., Zukiewicz, M., Silva, T., Costelloe, S., & O'Reilly, F. (2019). Evaluation of support for using student data to inform teachers' instruction. NCEE 2019–4008. https://eric.ed.gov/?id=ED598641

Grabarek, J., & Kallemeyn, L. M. (2020). Does teacher data use lead to improved student achievement? A review of the empirical evidence. *Teachers College Record*, 122(12). https://doi.org/10.1177/016146812012201201

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234. https://doi.org/10.3102/0013189x11413260

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences.* Princeton University Press.

Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1–30. http://www.jstatsoft.org/v40/i13/

Haertel, E. (2013). How is testing supposed to improve schooling? *Measurement: Interdisciplinary Research and Perspectives*, 11(1–2), 1–18. https://doi.org/10.1080/15366367.2013.783752

Hakimi, L., Eynon, R., & Murphy, V. A. (2021). The ethics of using digital trace data in education: A thematic review of the research landscape. *Review of Educational Research*, 91(5), 671–717. https://doi.org/10.3102/00346543211020116

Hall, A., & Huennekens, B. (2016). The use of common education data standards to support data integration. Center for the Integration of IDEA Data.

Halverson, R. (2010). School formative feedback systems. *Peabody Journal of Education*, 85(2), 130–146. https://doi.org/10.1080/0161956100368527

Halverson, R. (2014). Data-driven leadership for learning in the age of accountability. In A. J. Bowers, A. R. Shoho, & B. G. Barnett (Eds.), *Using Data in Schools to Inform Leadership and Decision Making* (pp. 255–267). Information Age Publishing.

Hao, J., & Mislevy, R. J. (2021). A data science perspective on computational psychometrics. In A. A. von Davier, R. J. Mislevy, & J. Hao (Eds.), *Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment: With Examples in R and Python* (pp. 133–158). Springer International Publishing. https://doi.org/10.1007/978-3-030-74394-9_8

Hawn Nelson, A., Jenkins, D., Zanti, S., Katz, M., Berkowitz, E., Burnett, T., & Culhane, D. (2020). A toolkit for centering racial equity throughout data integration. The Anne E. Casey Foundation. https://www.aecf.org/resources/a-toolkit-for-centering-racial-equity-within-data-integration/

Hernández-Torrano, D., & Courtney, M. G. R. (2021). Modern international large-scale assessment in education: an integrative review and mapping of the literature. *Large-scale Assessments in Education*, 9(1), 17. https://doi.org/10.1186/s40536-021-00109-1

Herodotou, C., Naydenova, G., Boroowa, A., Gilmour, A., & Rienties, B. (2020). How can predictive learning analytics and motivational interventions increase student retention and enhance administrative support in distance dducation*? The Journal of Learning Analytics*, 7(2), 72–83. https://doi.org/10.18608/jla.2020.72.4

Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., Wild, J., Krauss, S., Goretzko, D., & Stachl, C. (2021). Machine learning for the educational sciences. *Review of Education*, 9(3), e3310. https://doi.org/https://doi.org/10.1002/rev3.3310

Hu, Q., & Rangwala, H. (2020). Towards fair educational data mining: A case study on detecting at-risk students. *Proceedings of The 13th International Conference on Educational Data Mining* (EDM 2020; pp. 431–437).

Ifenthaler, D. (2021). Learning analytics for school and system management. In S. Vincent-Lancrin & R. Baker (Eds.), *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots (pp. 161–172). OECD Publishing*. https://doi.org/10.1787/589b283f-en

Ifenthaler, D., & Tracey, M. W. (2016). Exploring the relationship of ethics and privacy in learning analytics and design: implications for the field of educational technology. *Educational Technology Research and Development*, 64(5), 877–880. https://doi.org/10.1007/s11423-016-9480-3

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: a systematic review. *Educational Technology Research and Development*, 68(4), 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015, 2015/06/01/). Significance and challenges of big data research. *Big Data Research*, 2(2), 59–64. https://doi.org/https://doi.org/10.1016/j.bdr.2015.01.006

Jockers, M. L., & Underwood, T. (2015). Text-mining the humanities. In S. Schreibma, R. Siemens, & J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 291–306). https://doi.org/https://doi.org/10.1002/9781118680605.ch20

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Jorion, N., Roberts, J., Bowers, A. J., Tissenbaum, M., Lyons, L., Kuma, V., & Berland, M. (2020). Uncovering patterns in constructionist collaborative learning activities via cluster analysis of museum exhibit log files. *Frontline Learning Research*, 8(6), 77–87. https://doi.org/10.14786/flr.v8i6.597

Kang, S., & Bowers, A. J. (2021). NSF education data analytics collaborative workshop: How educators and data scientists meet and create data visualizations. In A. J. Bowers (Ed.), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists* (pp. 68–84). Teachers College, Columbia University. https://doi.org/10.7916/d8-jj2g-e225

Khan, M. O. R. (2021). Using data to pair students and teachers for enhanced collaborative growth. In A. J. Bowers (Ed.), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists* (pp. 195–206). Teachers College, Columbia University. https://doi.org/10.7916/d8-jj2g-e225

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Gasevic, D., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Tsai, Y.-S. (2022) Explainable Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, 3, 100074. https://doi.org/10.1016/j.caeai.2022.100074

Kizilcec, R. F., & Davis, D. (2022). Learning analytics education: A case study, review of current programs, and recommendations for instructors. In O. Viberg & Å. Grönlund (Eds.), *Practicable Learning Analytics*. Springer Nature. https://rene.kizilcec.com/wp-content/uploads/2022/09/kizilcec2022LAEducation.pdf

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., Turkay, S., Williams, J. J., & Tingley, D. (2020). Scaling up behavioral science interventions in online education. Proceedings of the National Academy of Sciences, 117(26), 14900–14905. https://doi.org/10.1073/pnas.1921417117

Kjell, O., Giorgi, S., & Schwartz, A. (2020). Package 'text': Analyses of text using natural language processing and machine learning (open access coding package). https://ftp.heanet.ie/mirrors/cran.r-project.org/web/packages/text/text.pdf

Knight, S., & Shum, S. B. (2018). Theory and learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 17–22). Society for Learning Analytics Research. https://doi.org/10.18608/hla17

Knowles, J. E. (2015). Of needles and haystacks: Building an accurate statewide dropout early warning system in Wisconsin. *Journal of Educational Data Mining*, 7(3), 18–67. http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/JEDM082

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333–353. https://doi.org/10.1002/wcs.1350

Kohavi, R., & Longbotham, R. (2016). Online controlled experiments and A/B testing. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning and data mining* (pp. 1–8). Springer. https://doi.org/10.1007/978-1-4899-7502-7_891-1

Koning, R., Hasan, S., & Chatterji, A. (2022). Experimentation and start-up performance: Evidence from A/B testing. *Management Science*, 68(9), 6355–7064. https://doi.org/10.1287/mnsc.2021.4209

Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., & Dawson, S. (2018). Understand students' self-reflections through learning analytics. LAK '18: *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 389–398). https://doi.org/10.1145/3170358.3170374

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: a cognitive presence case. *LAK '16: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 15–24). https://doi.org/10.1145/2883851.2883950

Krumm, A. E., & Bowers, A. J. (2022). Data-intensive improvement: The intersection of data science and improvement science. In D. J. Peurach, J. L. Russell, L. Cohen-Vogel, & W. R. Penuel (Eds.), *Handbook on Improvement Focused Educational Research* (pp. 465–484). Rowman & Littlefield.

Krumm, A. E., Means, B., & Bienkowski, M. (2018). *Learning analytics goes to school: A collaborative approach to improving education*. Routledge.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Alstyne, M. V. (2009). Computational social science. *Science*, 323(5915), 721–723. https://doi.org/doi:10.1126/science.1167742

Lee, L.-K., Cheung, S. K. S., & Kwok, L.-F. (2020). Learning analytics: Current trends and innovative practices. *Journal of Computers in Education*, 7(1), 1–6. https://doi.org/10.1007/s40692-020-00155-8

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684. https://doi.org/10.1177/2053951718756684

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–26. https://dl.acm.org/doi/10.1145/3359284

Liberman, M. (2010). Obituary: Fred Jelinek. *Computational Linguistics*, 36(4), 595–599.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 415–463). Springer. https://doi.org/10.1007/978-1-4614-3223-4_13

Liu, R., Koedinger, K., Stamper, J., & Pavlik, P. (2017). Sharing and reusing data and analytic methods with LearnSphere. In X. Hu, T. Barnes, A. Hershkovitz, & L. Paquette (Eds.) *Proceedings of the 10th International Conference on Educational Data Mining* (pp. 475–476).

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). Descriptive analysis in education: A guide for researchers. NCEE 2017–4023 https://ies.ed.gov/ncee/pubs/20174023/pdf/20174023.pdf

Lomotey, K., & Lowery, K. (2013). Black students, urban schools, and Black principals: Leadership practices that reduce disenfranchisement. In H. R. Milner (Ed.*) Handbook of Urban Education* (pp. 363–388). Routledge.

Loukina, A., Madnani, N., & Zechner, K. (2019). The many dimensions of algorithmic fairness in educational applications. In H. Yannakoudakis, E. Kochmar, C. Leacock, N. Madnani, I. Pilán, & T. Zesch (Eds.) *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications.* (pp. 1–10).

Mac Iver, M. A., Stein, M. L., Davis, M. H., Balfanz, R. W., & Fox, J. H. (2019). An efficacy study of a ninth-grade early warning indicator intervention. *Journal of Research on Educational Effectiveness*, 12(3), 363–390. https://doi.org/10.1080/19345747.2019.1615156

Madaio, M., Blodgett, S. L., Mayfield, E., & Dixon-Román, E. (2021). Confronting structural inequities in AI for education. Cornell University. arXiv preprint arXiv:2105.08847. https://arxiv.org/abs/2105.08847

Marshall, J. C. (1997). Data-based decision making. In S. D. Caldwell (Ed.), Professional development in learning-centered schools (pp. 150–167). National Staff Development Council.

Meyers, C. V., Moon, T. R., Patrick, J., Brighton, C. M., & Hayes, L. (2021). Data use processes in rural schools: management structures undermining leadership opportunities and instructional change. *School Effectiveness and School Improvement*, 33(1) 1–20. https://doi.org/10.1080/09243453.2021.1923533

Michael, G. (2017). A framework for considering comprehensibility in modeling. *Big Data*, 4(2), 75–88. https://doi.org/10.1089/big.2016.0007

Molloy, J. C. (2011). The open knowledge foundation: Open data means better science. *PLos Biology*, 9(12). https://doi.org/10.1371/journal.pbio.1001195

Mukherjee, R., & Kar, P. (2017). A comparative review of data warehousing ETL tools with new trends and industry insight. *Proceedings of the 2017 IEEE 7th International Advance Computing Conference* (IACC; pp. 946–948). https://ieeexplore.ieee.org/document/7976740

Munzner, T. (2014). *Visualization analysis and design*. CRC Press. https://doi.org/10.1201/b17511

Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1), 237. https://doi.org/10.3390/app11010237

Neff, G., Tanweer, A., Fiore-Gartland, B., & Osburn, L. (2017). Critique and contribute: A practice-based framework for improving critical data studies and data science. *Big Data*, 5(2), 85–97. http://online.liebertpub.com/doi/abs/10.1089/big.2016.0050

Noguera, P. A., & Alicea, J. A. (2021). The role of education in reducing racial inequality: Possibilities for change. In *Handbook of urban education* (pp. 26–49). Routledge.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books.

Palardy, G. J., Rumberger, R. W., & Butler, T. (2015). The effect of high school socioeconomic, racial, and linguistic segregation on academic performance and school behaviors. *Teachers College Record*, 117(12), 1–52. http://www.tcrecord.org/Content.asp?ContentID=18151

Paquette, L., Ocumpaugh, J., Li, Z., Andres, A., & Baker, R. (2020). Who's learning? Using demographics in EDM research. *Journal of Educational Data Mining*, 12(3), 1–30. https://doi.org/10.5281/zenodo.4143612

Parker, P. D., Jerrim, J., Chmielewski, A. K., & Marsh, H. W. (2017). Predicting university entry using machine-based models and solutions. In I. Schoon & R. K. Silbereisen (Eds.), Pathways to adulthood: Educational opportunities, motivation and attainment in times of social change (pp. 92–110). UCL Institute of Education Press. https://researchbank.acu.edu.au/cgi/viewcontent.cgi?article=2568&context=ftp_pub

Pea, R., & Jacks, D. (2014). The learning analytics workgroup: A report on building the field of learning analytics for personalized learning at scale. Stanford University. https://ed.stanford.edu/sites/default/files/law_report_6pager_09-02-2014.pdf

Piety, P. J. (2013). Assessing the educational data movement. Teachers College Press.

Piety, P. J. (2019). Components, infrastructures, and capacity: The quest for the impact of actionable data use on P–20 educator practice. *Review of Research in Education*, 43(1), 394–421. https://doi.org/10.3102/0091732x18821116

Piety, P. J., Hickey, D. T., & Bishop, M. (2014). Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems. *LAK '14: Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 193–202). https://doi.org/10.1145/2567574.25675

Piety, P. J., & Pea, R. D. (2018). Understanding learning analytics across practices. In D. Niemi, R. D. Pea, & B.

Saxberg (Eds.), *Learning Analytics in Education* (pp. 215–232). Information Age Publishing.

Popham, W. J. (2016). *The ABCs of educational testing: Demystifying the tools that shape our schools*. Corwin Press.

Pratt, M. (2021). Look who's talking—Facilitating data conversations that match data visualizations with educators' needs. In A. J. Bowers (Ed.), *Data Visualization, Dashboards, and Evidence Use in Schools: Data Collaborative Workshop Perspectives of Educators, Researchers, and Data Scientists* (pp. 161–176). Teachers College, Columbia University. https://doi.org/10.7916/d8-jj2g-e225

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59. https://doi.org/10.1089/big.2013.1508

Raghu, M., & Schmidt, E. (2020). A Survey of Deep Learning for Scientific Discovery. Cornell University arXiv. https://doi.org/10.48550/arXiv.2003.11755

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.

Reardon, S. F. (2019). Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age. RSF: The Russell Sage Foundation Journal of the Social Sciences, 5(2), 40–68. https://doi.org/10.7758/rsf.2019.5.2.03

Robinson, C. D., Gallus, J., Lee, M. G., & Rogers, T. (2021). The demotivating effect (and unintended message) of awards. Organizational Behavior and Human Decision Processes, 163, 51–64. https://doi.org/https://doi.org/10.1016/j.obhdp.2019.03.006

Rodolfa, K. T., Unanue, A. D., Gee, M., & Ghani, R. (2019). An experience-centered approach to training effective data scientists. *Big Data*, 7(4), 249–261. https://doi.org/10.1089/big.2019.0100

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open*, 2(4). https://doi.org/10.1177/2332858416673968

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management*, 54(5), 758–790. https://doi.org/https://doi.org/10.1016/j.ipm.2018.01.010

Saide, S., & Sheng, M. L. (2020). Toward business process innovation in the big data era: A mediating roles of big data knowledge management. *Big Data*, 8(6), 464–477. https://doi.org/10.1089/big.2020.0140

Savi, A. O., Williams, J. J., Maris, G. K., & van der Maas, H. (2017). The role of A/B tests in the study of large-scale online learning. OSF Preprints. https://doi.org/10.31219/osf.io/83jsg

Schildkamp, K., Karbautzki, L., & Vanhoof, J. (2014). Exploring data use practices around Europe: Identifying enablers and barriers. *Studies in Educational Evaluation*, 42, 15–24. https://doi.org/10.1016/j.stueduc.2013.10.007

Schutt, R., & O'Neil, C. (2013). Doing data science: Straight talk from the frontline. O'Reilly.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. https://doi.org/10.1214/10-STS330

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly Media. https://www.tidytextmining.com/index.html

Simon, H. A. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), *Computers, Communications, and the Public Interest* (pp. 37–52). The Johns Hopkins Press. http://knowen-production.s3.amazonaws.com/uploads/attachment/file/2005/DESIGNING%2BORGANIZATIONS%2Bfor%2BInformation-Rich%2Bworld%2B--%2BSImon.pdf

Singer, J. D. (2019). Reshaping the arc of quantitative educational research: It's time to broaden our paradigm. *Journal of Research on Educational Effectiveness*, 12(4), 570–593. https://doi.org/10.1080/19345747.2019.1658835

Singh, L., Deshpande, A., Wenchao Zhou, Banerjee, A., Bowers, A. J., Friedler, S., Jagadish, H. V., Karypis, G., Obradovic, Z., Vullikanti, A., & Zuo, W. (2017). NSF BIGDATA PI Meeting - Domain-Specific Research Directions and Data Sets. Association for Computing Machinery (ACM) SIGMOD Record, 47(3), 32-35. http://doi.org/10.1145/3316416.3316425

Singh, L., Deshpande, A., Zhou, W., Banerjee, A., Bowers, A., Friedler, S., Jagadish, H. V., Karypis, G., Obradovic, Z., Vullikanti, A., & Zuo, W. (2019). NSF BIGDATA PI meeting—Domain-specific research directions and data sets. SIGMOD Record, 47(3), 32–35. https://doi.org/10.1145/3316416.3316425

Singh, S., & Mahmood, A. (2021). The NLP cookbook: Modern recipes for transformer based deep learning architectures. *IEEE Access*, vol. 9, 68675-68702. https://doi.org/10.1109/ACCESS.2021.3077350

Singleton, A., & Arribas-Bel, D. (2021). Geographic data science. *Geographical Analysis*, 53(1), 61–75. https://doi.org/https://doi.org/10.1111/gean.12194

Soland, J., Domingue, B., & Lang, D. (2020). Using machine learning to advance early warning systems: Promise and pitfalls. *Teachers College Record*, 122(14), 1–30. https://doi.org/10.1177/016146812012201403

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S.-F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, 65, 3–14. https://doi.org/10.1016/j.imavis.2017.08.003

Srivastava, A. N., & Sahami, M. (2009). *Text mining: Classification, clustering, and applications*. CRC Press.

Steyvers, M., & Griffiths, T. (2007). *Probabilistic topic models. Handbook of latent semantic analysis*, 427(7), 424–440. https://doi.org/10.4324/9780203936399

Stodden, V., Guo, P., & Ma, Z. (2013). Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. *PLoS ONE*, 8(6), e67111. https://doi.org/10.1371/journal.pone.0067111

Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., Heroux, M. A., Ioannidis, J. P. A., & Taufer, M. (2016). Enhancing reproducibility for computational methods. *Science*, 354(6317), 1240–1241. https://doi.org/10.1126/science.aah6168

Streifer, P. A. (2004). *Tools and techniques for effective data-driven decision making*. Scarecrow Education.

Streifer, P. A., & Schumann, J. A. (2005). Using data mining to identify actionable information: Breaking new ground in data-driven decision making. *Journal of Education for Students Placed at Risk*, 10(3), 281–293. https://doi.org/10.1207/s15327671espr1003_4

Stuit, D., O'Cummings, M., Norbury, H., Heppen, J., Dhillon, S., Lindsay, J. J., & Zhu, B. (2016). Identifying early warning indicators in three Ohio school districts. REL 2016–118. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. http://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/REL_2016118.pdf

Sun, M., Liu, J., Zhu, J., & LeClair, Z. (2019). Using a text-as-data approach to understand reform processes: A deep exploration of school improvement strategies. *Educational Evaluation and Policy Analysis*, 41(4), 510–536. https://doi.org/10.3102/0162373719869318

Tarka, P., & Jędrych, E. (2020). On the unstructured big data analytical methods in firms: Conceptual model, measurement, and perception. *Big Data*, 8(6), 478–500. https://doi.org/10.1089/big.2020.0123

Taylor, D. (2016). Battle of the data science Venn diagrams. KDNuggets News. https://deeplearning.lipingyang.org/wp-content/uploads/2017/10/Battle-of-the-Data-Science-Venn-Diagrams.pdf

Thomke, S. H. (2020). Experimentation works: The surprising power of business experiments. Harvard Business Press.

Tissenbaum, M. (2020). I see what you did there! Divergent collaboration and learner transitions from unproductive to productive states in open-ended inquiry. *Computers & Education*, 145, 103739. https://doi.org/https://doi.org/10.1016/j.compedu.2019.103739

Tissenbaum, M., Berland, M., & Lyons, L. (2017). DCLM framework: understanding collaboration in open-ended tabletop learning environments. *International Journal of Computer-Supported Collaborative Learning*, 12(1), 35–64. https://doi.org/10.1007/s11412-017-9249-7

L. Towne, L. L. Wise, & T. M. Winters (Eds.) (2004). Advancing scientific research in education. National Academies Press. https://nap.nationalacademies.org/catalog/11112/advancing-scientific-research-in-education

Tufféry, S. (2011). *Data mining and statistics for decision making* (2nd ed.). Wiley.

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1), 1–67. http://www.jstor.org/stable/2237638

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Pearson.

Ullman, J. D. (2020). The battle for data science. *IEEE Data Eng. Bull.*, 43(2), 8–14. http://sites.computer.org/debull/A20june/p8.pdf

Urick, A. (2018). Large-scale data with large-scale implications. In C. R. Lochmiller (Ed.), *Complementary Research Methods for Educational Leadership and Policy Studies* (pp. 143–172). Palgrave-MacMillan.

van der Zee, T., & Reich, J. (2018). Open education science. *AERA Open*, 4(3), 2332858418787466. https://doi.org/10.1177/2332858418787466

Vanhoutte, E. (2016). The gates of hell: History and definition of digital| humanities| computing. In M. Terras, J. Nyhan, & E. Vanhoutte (Eds.) *Defining Digital Humanities: A Reader* (pp. 135–172). Routledge.

Vassiliadis, P. (2009). A survey of extract—transform—load technology. *International Journal of Data Warehousing and Mining*, 5(3), 1–27. https://doi.org/10.4018/jdwm.2009070101

Vaughan, J. W., & Wallach, H. (2020). A human-centered agenda for intelligible machine learning. In M. Pelillo & T. Scantamburlo (Eds.) *Machines We Trust: Perspectives on Dependable AI*. MIT Press. (E-book, Ch.8). http://www.jennwv.com/papers/intel-chapter.pdf

Vincent-Lancrin, S. (2021). Frontiers of smart dducation technology: Opportunities and challenges. In *OECD Digital Education Outlook 2021: Pushing the Frontiers with AI, Blockchain, and Robots* (pp. 19–42). OECD Publishing. https://doi.org/10.1787/589b283f-en

Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, 2019(2) 494–619. https://ssrn.com/abstract=3248829

Wagenmakers, E.-J., Sarafoglou, A., Aarts, S., Albers, C., Algermissen, J., Bahník, Š., van Dongen, N., Hoekstra, R., Moreau, D., van Ravenzwaaij, D., Sluga, A., Stanke, F., Tendeiro, J., & Aczel, B. (2021). Seven steps toward more transparency in statistical practice. *Nature Human Behaviour*, 5(11), 1473–1480. https://doi.org/10.1038/s41562-021-01211-8

Wan, T. (2021, January 13). A record year amid a pandemic: US Edtech Raises $2.2 billion in 2020. EdSurge. https://www.edsurge.com/news/2021-01-13-a-record-year-amid-a-pandemic-us-edtech-raises-2-2-billion-in-2020

Wan, Z., Xia, X., Lo, D., & Murphy, G. C. (2021). How does machine learning change software development practices? *IEEE Transactions on Software Engineering*, 47(9), 1857–1871. https://doi.org/10.1109/TSE.2019.2937083

Wang, Y., & Bowers, A. J. (2016). Mapping the field of educational administration research: A journal citation network analysis of the discipline. *Journal of Educational Administration*, 54(3), 242–269. https://doi.org/10.1108/JEA-02-2015-0013

Wang, Y., Bowers, A. J., & Fikis, D. J. (2017). Automated text data mining analysis of five decades of educational leadership research literature: Probabilistic topic modeling of EAQ articles from 1965 to 2014. *Educational Administration Quarterly*, 53(2), 289–323. https://doi.org/10.1177/0013161x16660585

Watts, D. J. (2017\). Should social science be more solution-oriented? *Nature Human Behaviour*, 1(1), 0015. https://doi.org/10.1038/s41562-016-0015

Wayman, J. C., Shaw, S., & Cho, V. (2017). Longitudinal effects of teacher use of a computer data system on student achievement. *AERA Open*, 3(1). https://doi.org/doi:10.1177/2332858416685534

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, 7, 1832. https://doi.org/10.3389/fpsyg.2016.01832

Wickham, H. (2010). A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, 19(1), 3–28. https://doi.org/10.1198/jcgs.2009.07098

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23. https://doi.org/10.18637/jss.v059.i10

Wickham, H., & Grolemund, G. (2017). *R for data science: import, tidy, transform, visualize, and model data*. O'Reilly Media. https://r4ds.had.co.nz/index.html

Wilkinson, L. (2012). The grammar of graphics. In J. E. Gentle, W. K. Härdle, & Y. Mori (Eds.), *Handbook of Computational Statistics: Concepts and Methods* (pp. 375–414). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21551-3_13

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18

Williamson, B. (2016). Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy*, 31(2), 123–141. https://doi.org/10.1080/02680939.2015.1035758

Williamson, B. (2017). Who owns educational theory? Big data, algorithms and the expert power of education data science. *E-Learning and Digital Media*, 14(3), 105–122. https://doi.org/10.1177/2042753017731238

Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995

Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective. *Educational Technology Research and Development*, 64(5), 881–901. https://doi.org/10.1007/s11423-016-9463-4

Wu, C. F. J. (1997). Statistics = data science? [Lecture for H. C. Carver Professorship, University of Michigan].

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, 45, 1–37. https://doi.org/10.1017/s0140525x20001685

Yau, N. (2013, December 18). Data scientist surpasses statistician on Google Trends. FlowingData. https://flowingdata.com/2013/12/18/data-scientist-surpasses-statistician-on-google-trends