BIG DATA IN EDUCATION

Context:

Within the last 15 years, advances in computing, data capturing, and data storage technology have redefined how we observe, interpret and shape the most critical pillars of our lives. The emergence and rapid adoption of this technology has been coined as the era of big data (Capatosto, 2017; Wang, 2016; Data-Pop Alliance, 2015). Specifically, big data in education is a tool used to predict student outcomes, shape policy decisions, and improve learning.

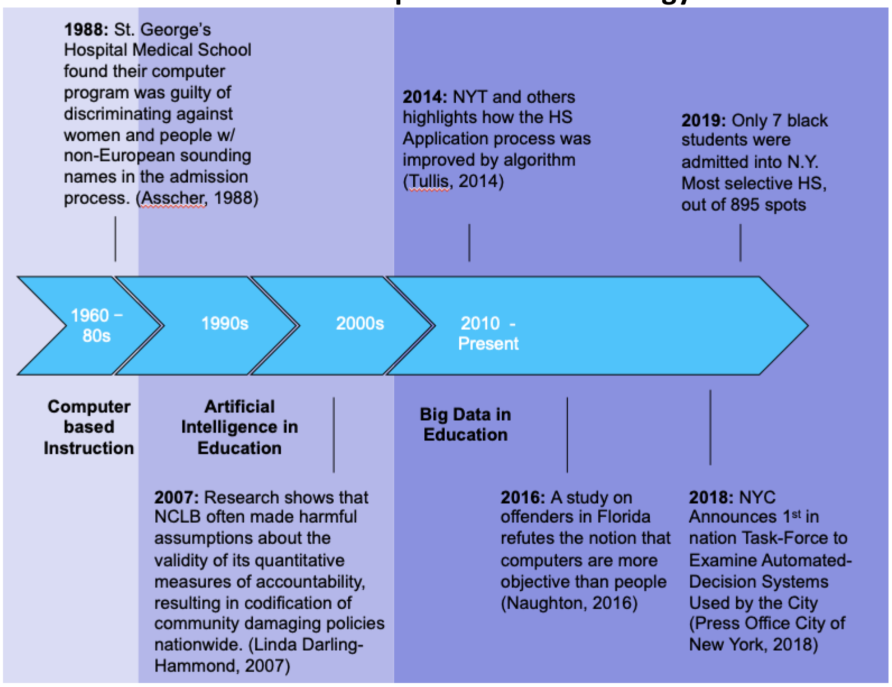
In 2004, New York City became the first city to adopt algorithmic-based high school choice selection, as a way to streamline the complex high school admission process. Since adopting the system the number of students who did not receive an offer from one of their chosen schools fell drastically—from 30,000 students in 2003 to 3,000 students in 2004. (Atila Abdulkadiroglu, Parag A. Pathak, and Alvin E. Roth, 2005) Within the same year additional cities including Boston, Denver, and New Orleans have adopted similar systems with the goal of improving enrollment efficiency and fairness.

The emergence of big data is seen as a critical tool in education, with the potential to create new learning experiences for children that meet their individual needs. State and district policy-makers around the country are partnering with big data practitioners, implementing data-driven technologies with ambitious goals.

Unfortunately many state and district led big data initiatives fail to identify and acknowledge the harm these systems can create by codifying implicit beliefs into seemingly objective, complicated, math-based systems.

In theory, big data allows data-driven school systems to be more accountable to evidence and capable of providing predictive technology that better supports all students. In practice, big data analytics is a **decision-making support** and **communication tool**, **and it comes with risks**. Around the nation, policy-makers, educators and communities are experiencing big data from vastly different altitude of impact; for policy-makers, the success of big data in public education is measured by narrowly focused key performance indicators (KPIs) related to test score growth and resource efficiency, often ignoring equity concerns. By implementing systems that exclude the voices of the communities (particularly communities of color) most impacted by big data and predictive analytics, public officials create echo chambers and produce interventions that rarely result in scalable equitable outcomes.

A Brief Timeline of Computational Technology and Bias



Evidence:

In 2013 district leaders in Dallas Independent School District (serving nearly 157,000 students in 230 schools) invested in a data-driven initiative, implementing personalized learning models in 5 schools. The Dallas ISD felt the initial results of the initiative were promising, as students across the 5 target schools saw growth relative to their peers; promising metrics and family support have led to rapid expansion. (Roger, 2018) As of May 2018 the program now serves over 20,000 students in over 100 Dallas ISD schools.

The policies and practices of the Dallas ISD are unique, rooted in their commitment to using big data technology as a tool to address education equity in collaboration with community partners and educators.

A 2019 British Journal of Educational Technology study reviewing the current state of big data in education and its potential for positive change. According to this critical review, Big Data offers a number of opportunities to education, however Big Data in education and educational research are two separate areas of inquiry, requiring different sets of skills and knowledge (Daniel, 2019). While educational research is broadly concerned with the investigation of various aspects of education, focusing on the context of data, requiring expertise in education, Big Data in education deals with the analysis of large and complex data using Data Science techniques (2019).

Since 2015 the research and implementation of big data in education has focused on themes such as (Daniel, 2019; Dede, 2016):

- Developing digital learning technology that allow for the capturing of new types of data.
- Improving evidence-based decision making
- Reconceptualizing existing data to create automated predictive interventions
- Data privacy and ownership

Absent from most research is the role bias plays in big data. A key principle of big data research and technology is the devaluing of data context in analysis (Daniel 2019). Most practitioners argue that 'numbers speak for themselves' (Anderson 2008) and human reasoning (and experience) simply get in the way of objective insight. Others are cautiously optimistic regarding the potential of big data in education, but warn that by simply focusing on technical fixes and not the ethics behind the technology we risk causing harm (Eynon, 2013).

This trend of asymmetrical positive thinking around the objectivity of numbers can lead to unfounded assertions and real harm to families. In 2014, government officials in St. Pauls established a multi-agency data-sharing, Joint Powers Agreement (JPA) aimed at using big data and predictive analytics to predict and flag youth "at risk" of future delinquency. This plan was originally established as part of a Community Innovation Project, an effort with the goal of 'better supporting youth and families'. (Melisizwe, 2019) Unfortunately this agreement excluded the participation and oversight of community members, families and advocates. This exclusion incited fear among communities who understood that the technology would essentially function as racial profiling of children predicted to engage in crime. (In Equality / Stop the Cradle to Prison Algorithm Coalition, 2018)

Those in charge of the JDA were focused on using big data to inform decision-making and reduce racial bias by move away from individual discretion, and increasingly rely on tools like risk assessments and predictive algorithms. (2018) Unfortunately this did not address community concerns around algorithm bias being hard-wired into the technology due to the bias in the historical data. A 2016 analysis showed that black youth in Ramsey County (part of the JDA) are 4.44 times more likely to be arrested than white youth, and 3.54 more likely to be admitted to detention. (Racial Disparities of Black Youth in Ramsey County, RRI, 2016) A 2016 RAND analysis of big data technology similar to that of the JDA found that those who were "flagged" in the system were more heavily surveilled and arrested. (Saunders, J., Hunt, P., & Hollywood, J. S., 2016) While the St. Paul School District believed it was building at tool for positive change, having invested millions in the technology, they had failed to effectively grapple with the ethical concerns of a broad sweeping data-sharing agreement and did not address the concerns of the community regarding systemic-oppression being codified into the data.

Ultimately community organizers like the Stop the Cradle to Prison Algorithm (CPA) Coalition were able to dissolve the agreement citing evidence of racial bias, child privacy violations, and a lack of transparency and collaboration.

STEINHARDT
Metropolitan Center for Research
on Equity and the Transformation of Schools

But what happens when those most impacted by big data use big data for education equity?

In 2014, Dr. Meredith Broussard illuminated inequalities in the Philadelphia district school system by analyzing how resources are allocated. Broussard, with support from community volunteers, was able to use the available data systems (along with technology they designed) to highlight the fact that the district was missing 75% of the recommended curriculum and that $0 were being allocated to textbooks. (Broussard, 2014)In 2015, following the work of Broussard and others, the district increased its discretionary fund dedicated to textbooks from $0 to $165. Meredith and her community of activists were able to use big data to communicate a massive inequality to public officials that directly impacts both academic growth and resource efficiency.

Bottom Line:
Big data in education is a tool designed to derive insight using subjective statistics and trends. Whether the intervention is academic or social-economic, transparency and trust will always lead to a more positive impact than pure statistics based governance. Data-driven education policy, like all public policy cannot be equitable if there aren't formal partnerships between impacted communities and government, using explicit language to outline goals and potential drawbacks.

Research Links:
Dougherty, Shaun M., Joshua Samuel Goodman, Darryl V. Hill, Erica G. Litke, and Lindsay Coleman Page. 2014. Middle School Math Acceleration and Equitable Access to 8th Grade Algebra: Evidence from the Wake County Public School System. HKS Faculty Research Working Paper, Harvard University.

Office of Transformation and Innovation Author: Courtney Rogers Contributors: Angie Gaylord & Cecilia Oakeley 2018

Best Policy/Practice Links:
Dignity in Schools

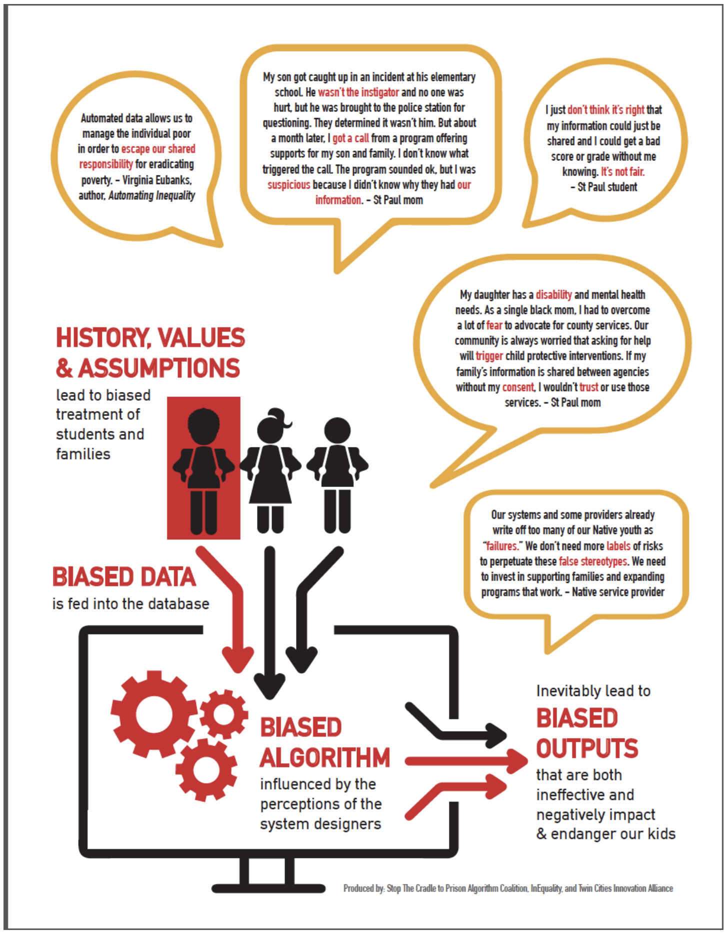
Data & Society

Dallas Independent School District

Youth Justice Project

Multimedia:
1. https://nyupress.org/9781479837243/algorithms-of-oppression/
2. See attached JPA graphic

NYU **STEINHARDT**
Metropolitan Center for Research
on Equity and the Transformation of Schools



Author:
Robert Azeem Jackson III

Citation:

Anderson, C. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." Wired. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory/.

Atila Abdulkadiroglu, Parag A. Pathak, and Alvin E. Roth, "The New York City High School Match," *American Economic Review* 95 (2) (2005): 364-–367, available at https://seii.mit.edu/wp-content/uploads/2011/12/Paper-New-York-City-High-School-Math.pdf.

"Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data." Data-Pop Alliance White Paper Series. Data-Pop Alliance (Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute) and Internews. September 2015

Broussard, M. (2014, July 15). Why Poor Schools Can't Win at Standardized Testing. *The Atlantic*. Retrieved June 12, 2019, from https://www.theatlantic.com/education/archive/2014/07/why-poor-schools-cant-win-at-standardized-testing/374287/

Broussard, M. (2019). *Artificial unintelligence: How computers misunderstand the world*. Cambridge, MA: The MIT Press.

Capatosto, Kelly (2017). "Foretelling the Future A Critical Perspective on the Use of Predictive Analytics in Child Welfare." Kirwan Institute Research Report. Retrieved from http://kirwaninstitute.osu.edu/wp-content/uploads/2017/05/ki-predictive-analytics.pdf

Daniel, B. K. (2015). Big Data and analytics in higher education: opportunities and challenges. *British Journal of Educational Technology*, **46**, 904–920. doi:10.1111/bjet.12230

Daniel, B. (2017) Big Data and data science: A critical review of issues for educational research. Br. J. Educ. Technol. 2017.

David Gillborn, Paul Warmington & Sean Demack (2018) QuantCrit: education, policy, 'Big Data' and principles for a critical race theory of statistics, Race Ethnicity and Education, 21:2, 158-179, DOI: 10.1080/13613324.2017.1377417

Dede, C., Ho, A., & Mitros, P. (2016). Big Data analysis in higher education: promises and pitfalls. EDUCAUSE review August 2016 (pp. 8–9). Retrieved September 1, 2016, from http://er.educause.edu/articles/2016/8/big-data-analysis-in-higher-education-promises-and-pitfalls

Dougherty, Shaun M., Joshua Samuel Goodman, Darryl V. Hill, Erica G. Litke, and Lindsay Coleman Page. 2014. Middle School Math Acceleration and Equitable Access to 8th Grade

Algebra: Evidence from the Wake County Public School System. HKS Faculty Research Working Paper, Harvard University.

Herold, B. (2014, May 2). InBloom's Collapse Shines Spotlight on Data-Sharing Challenges. Retrieved June 9, 2019, from https://www.studentprivacymatters.org/newsclips/inbloom-specific-newsclips/

Kalil, T. (2012, March 29). Big Data is a Big Deal. Retrieved June 11, 2019, from https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal

IMPROVING OUTCOMES FOR KIDS & FAMILIES Beyond Predictive Analytics & Data Sharing Policy Brief by IN EQUALITY / Stop the Cradle to Prison Algorithm Coalition KEY, 2019.

KSTP. (2019, January 28). St. Paul, Ramsey County, school officials dissolve joint powers agreement. *ABC Eye Witness News 5*. Retrieved June 10, 2019, from https://kstp.com/news/st-paul-ramsey-county-school-officials-dissolve-joint-powers-agreement/5225446/

Linda Darling-Hammond (2007) Race, inequality and educational accountability: the irony of 'No Child Left Behind',Race Ethnicity and Education, 10:3, 245-260, DOI: 10.1080/13613320701503207

Mayer-Schonberger, V., and K. Cukier. 2013. Big Data: A Revolution That Will Transform How We Live, Work and Think. London: Hodder & Stoughton. Kindle Edition

McNeel, B. (2018, December 11). Dallas Hits on Successful School Turnaround Model With ACE, but It Comes at a Steep Price. Could a Wider Expansion Across Texas Now Be Its Best Bet to Survive? *The 74*. Retrieved June 8, 2019, from https://www.the74million.org/article/dallas-hits-on-successful-school-turnaround-model-with-ace-but-it-comes-at-a-steep-price-could-a-wider-expansion-across-texas-now-be-its-best-bet-to-survive/

Melisizwe, T. (2019, January 29). Coalition to Stop the Cradle to Prison Algorithm Celebrates Hard-Won Victory with the Dissolution of Problematic Data-Sharing Agreement. Retrieved June 9, 2019, from https://dignityinschools.org/coalition-to-stop-the-cradle-to-prison-algorithm-celebrates-hard-won-victory-with-the-dissolution-of-problematic-data-sharing-agreement/

Naughton, J. (2016, June 26). Even algorithms are biased against black men. *The Guardian*. Retrieved June 11, 2019, from https://www.theguardian.com/commentisfree/2016/jun/26/algorithms-racial-bias-offenders-florida

ONeil, C. (2018). *Weapons of math destruction: How big data increases inequality and threatens democracy*. London: Penguin Books. doi:https://doi.org/10.1111/newe.12047

Office of Transformation and Innovation: Courtney Rogers Contributors: Angie Gaylord & Cecilia Oakeley 2018

Press Office City of New York (2018, May). Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City. Retrieved July 28, 2019, from https://www1.nyc.gov/office-of-the-mayor/news/251-18/mayor-de-blasio-first-in-nation-task-force-examine-automated-decision-systems-used-by

Saunders, J., Hunt, P., & Hollywood, J. S. (2016). Predictions put into practice: A quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology, 12*(3), 347-371. doi:10.1007/s11292-016-9272-0

Tullis, T. (2014, December). How Game Theory Helped Improve New York City's High School Application Process. *The New York Times*. Retrieved August 1, 2019, from https://www.nytimes.com/2014/12/07/nyregion/how-game-theory-helped-improve-new-york-city-high-school-application-process.html

Wang, Y. (2016). Big Opportunities and Big Concerns of Big Data in Education. *TechTrends, 60*(4), 381-384. doi:10.1007/s11528-016-0072-1