Data&Society RESEARCH INSTITUTE

The Leadership Conference

OPEN TECHNOLOGY INSTITUTE

Data & Civil Rights: Education Primer

by Andrea Alarcon, Elana Zeide, Alex Rosenblat, Kate Wikelius, danah boyd, Seeta Peña Gangadharan, and Corrine Yu
Produced for Data & Civil Rights Conference / October 30, 2014

Education is one of the primary mechanisms for creating equal opportunity and equity in America. In the six decades since the Supreme Court declared racial segregation unconstitutional in *Brown v. Board of Education*,[1] the United States has made much progress. However, pervasive racial injustice still exists in the public educational system. Continued racial isolation, the massive inequity in resources between primarily White and primarily minority schools, and the unequal treatment of racial minority students within schools undermine the economic, social, and political opportunities of minorities in the United States.

Many education reformers see the merging of student data, predictive analytics, processing tools, and technology-based instruction as the key to the future of education and a means to further opportunity and equity in education. However, despite widespread discussion of the potential benefits and costs of using data in educational reform, it is difficult to determine who benefits from reforms since there has been little assessment of these programs and few oversight mechanisms.

Background: Discrimination in Education

Major civil rights concerns and themes in U.S. education:

- *School Graduation Rates*: For the vast majority of ethnic and racial minorities, high school graduation rates have stagnated at about 60%, lagging behind the 83% rate for White students. High poverty schools have even lower graduation rates: 50% for Black students.[2]
- *Challenges to College Success*: Due to economic hardship and lack of academic preparation, 56% of Black students and 64% of Latino high school graduates enroll in postsecondary education, compared with 72% of White graduates. For full-time, first-time enrollees in a four-year institution, only 20% of Black students graduated in four years, less than half the rate (41%) of White students.[3]
- *Segregation*: Millions of American students continue to attend separate and unequal schools. As of 2010, 74.1% of Black students and 79.1 of Latino students attended majority-minority schools. Between 1980 and 2009, the number of Black students attending schools that are more than 90% segregated, rose from 33.2% to 38.1%. The

number of Latino students attending these schools increased from 28.8% to 43.1%.[4]

- *Lack of School Resources*: Minority students disproportionately attend underfunded and under-resourced schools.[5] For example, compared with more schools with affluent student populations, schools where more than three-quarters of the student population qualified as low income contained three times as many uncertified or out-of-field teachers in both English and science.[6]
- *Uneven School Discipline*: Schools disproportionately punish racial minority students through suspension and expulsion, harming their prospects for a good education.[7] Even in more affluent schools and schools with a majority of White students, students of color receive harsher punishments.[8]

Data-Driven Education

Education is increasingly data-driven.[9] The promise of data in education is that, combined with other technologies and data analytics tools, it can enable more personalized instruction, maximizing resources, and permitting greater access to affordable, or even free, education. Proponents advocate data-driven instruction and educational reform as a low-cost and more effective means to deliver better content to a wider variety of students, which, in turn, is seen as a means to encourage social and economic mobility.

Historically, it has been difficult to study specific aspects of the learning experience. While the concrete impacts of individual differences have been a gold standard, traditional learning methods have made measurement challenging. New learning analytics techniques, combined with both educational and extra-curricular data, introduce new possibilities for being able to better understand and measure individual differences.[10] Much existing student data exists in disconnected silos, but new technologies like cloud computing make aggregating and analyzing it more feasible. As information transfer, storage, and analysis has become less expensive, educators, schools and educational agencies have begun collecting, combining, and examining new types of data to evaluate and improve instructional, institutional, and public policies in education.[11] Educators and researchers can more easily aggregate previously separate streams of student data in order to better understand, measure, and intervene on the basis of individual differences among student learners.[12] Algorithmic analysis of these vast datasets can generate inferences about instructional content and methodology, student aptitude, progress, and engagement, helping to identify students' strengths, weaknesses and "learning styles," and guide students through the next-best piece of content so as to learn with maximum efficiency. [13]

Digital platforms can theoretically provide better instruction and guidance to a broader array of students and, in doing so, potentially enable social and economic mobility. Such tools also generate substantial information about students. Learning analytics, data mining, and other big data approaches to instruction, administration, and policy can create more individually-tailored instruction, maximize resources, and provide greater access to affordable or free education. The U.S. Department of Education promotes the collection and analysis of information generated by and about students as a means to improve evaluation of instructional, institutional, and public policies; provide low-cost education to a wide array of potential learners; and create more accurate tools for assessment than infrequent standardized tests.[14] They also predict that better data can help close achievement gaps, increase educational opportunities and college access, and reduce discrimination against underserved students.

Data can also facilitate "personalized" instruction and guidance, identify at-risk students for early intervention, and assess instructional techniques and curricula.[15] Data mining can support a variety of education-related functions, including building student models to individualize instruction, map learning domains, evaluate pedagogical support, and contribute to learning science.[16] Analytics techniques can be used to create models to predict registration, student performance, and retention.[17] The wealth of new information about students is used to detect cheating or plagiarism, create college or course recommendation engines, and identify abnormal results.[18] It can also be used for administrative, recruiting, and fundraising purposes.[19]

Education researchers and companies promise that personalization will bring about a more adaptive, responsive, and efficient school system and ameliorate inequalities in the educational system by providing underserved students with school environments better tailored to their unique needs. "Personalized learning" adjusts instructional content, pace, and complexity to meet an individual learner's needs and objectives. It typically involves delivering students differentiated instruction or guidance from their peers based on information regarding their overall proficiency, interests, preferred learning style, motivation and demographic. "Adaptive learning" platforms use algorithms to automate personalization based on a student's prior interaction with a software program, a student can receive different lessons, sequences, pacing, and challenges, among other things.[20] These approaches can be applied not only to specific classes, but also to entire curricula and career paths.

Tutoring programs like Khan Academy provide supplemental instruction through digital platforms. Massive Open Online Courses (MOOCs) are typically large-scale distribution systems for course content, frequently affiliated with a traditional university. Much of the content produced through MOOCs is free and open-access, theoretically enabling anyone with internet access to get high quality education and learn at their own pace. MOOCs also rely on data to optimize instruction. The combination of digitally delivered instruction and personalized coursework may be especially beneficial for those who otherwise do not have access to high quality education.

Educational institutions now use big data techniques to recruit and enroll students. Use of data analytics has also become intertwined with questions of access to higher education technologies, including the identification and recruitment of underserved minorities and better assessment of the financial need of students. Proponents of data-driven services also encourage use of technology to help students – especially underserved and first generation students – better navigate the college admissions and financial aid processes, by creating more transparency regarding available options, the percentage of students graduating with debt, and streamlining the application process.

Already, some schools have begun to implement analytics tools to intervene in students' trajectories and prevent underachievement. At the high school level, the Early Warning Indicator and Intervention System analyzes attendance records, behavior problems, and course performance to measure dropout risk. It has succeeded in a number of cases in helping educators get students identified as "at risk" back on track.[21] At the college level, Georgia State utilizes predictive analytics methods to increase semester-to-semester retention rates by 5% and reduced time-to graduation by half a semester.[22]

Concerns and Challenges

Critics worry that data-driven education will entrench existing inequalities and contribute to a problem of cumulative disadvantage.[23] Though data-driven systems may appear "neutral" and irrefutably scientific,[24] they—and the ways in which educators and reformers adopt and implement them—reflect particular norms and values about what educational opportunity and equity means.[25] The complexity of algorithmic analysis makes detection of bias and discrimination difficult, as the technical processes of data mining can obscure intentionally differential treatment of members of protected classes or result in unintentional disparate impact based on the types of data used and ways in which algorithms are trained. Statistics may be manipulated to support or obscure invidious discrimination, while empirical and normative assumptions and values are embedded in big data techniques.

Though data-driven education has the potential to improve access to and the quality of teaching in underserved communities, it may also perpetuate persistent labeling, deepen rather than lessen concerns about resources, violate peoples' expectations of privacy, and enable inappropriate or harmful repurposing of educational data in non-educational contexts. For example, students or their guardians may find it impossible to eschew or reverse flawed algorithmic assessments. The identification of students as "at risk" might not allow them to remove any harmful record of their failures if they improve later on.[26] Students may see labels as self-fulfilling prophecies and predictive analytics may prime educators to make prior judgments about students' capabilities and character.

Critics are concerned that the hype surrounding these techniques will be used to reinforce present structural inequities. New technology also amplifies concerns about the allocation of resources (e.g., implementing and operating personalized learning systems over staffing classrooms), the role of predatory lenders in shaping marginalized students' access to higher education, and the mechanisms by which data will be used to include and exclude students seeking access to higher education. For example, The National Research Center for College and University Admissions sells a model to predict which high school students are most likely to enroll at a particular college.[27] Data points in the system include geography, academic interest, whether students would prefer a private college, and family income. Although the process has helped target students who might not have thought about going to college at all, it also "raises the threat that some colleges with worrisome bottom lines will only go after certain kinds of students — like students who can pay their way without scholarships — at the expense of others." Meanwhile, threats to student privacy could also make it difficult for students to advance, including less privileged youth. Data collected in the course of education may also be repurposed to drive decontextualized decision-making based on potentially outdated or non-representative data. For example, there is some worry that information like attendance records will affect financial decisions in other domains. Already, the availability of educational data is shaping who is recruited, selected, and financially supported to access to higher education. It is often unclear how accountability in these multi stakeholder systems should work.

More broadly, while some stakeholders believe that individualization of education is productive, others believe that it fragments society and narrows learning opportunities. True personalization would not just pace learning more effectively or target the more adequate educational methods, but would provide students with different content and educational opportunities depending on an assessment of their capabilities. Personalized educational tools may create filter bubbles that increase, rather than decrease, stratification.[28] Data-driven education privileges certain kinds of learning over others—namely those that are easy to measure and statistically analyze, and may forestall students' educational development, growth, and opportunity.

Companies are pushing for the aggregation of student data into analytics tools to improve their algorithms and assess the quality of their predictions. At the same time, concerns persist over who should have access to what data, for what purposes, and with what oversight. Part of what is at stake is that the use of data to drive instruction and educational policy-making relies on various assumptions. They include the idea that learning progress may be measured quantifiably, that data-analysis is accurate, that predictive models are useful, and that individualization of instruction will improve the educational success for at-risk and under-served students. Not all stakeholders view education in these terms.

The controversy over the inBloom, Inc. data repository is indicative of deep anxiety about the use of student information. A non-profit entity, inBloom contracted with states and districts to provide interoperable and secure data storage that would facilitate data use and analysis. Parents and privacy experts raised concerns about the scope of inBloom's potential data collection, focusing on the fact that the database structure had the capacity to collect up to 400 fields, ranging from test scores and special-education enrollment to whether children got free lunches.[29] They also objected to the role of third party actors in the educational system, the potential that private companies might profit from student information, and the lack of opt-out mechanisms. Opposition to inBloom eventually prompted all its clients to withdraw from the program, and the company shut down in April 2014. Due in part to this incident, U.S. states considered 110 bills explicitly addressing student data in 2014.[30]

Debate over the use of data in educational contexts has long been a polarizing topic, even absent civil rights concerns. Untangling the ways in which big data tools and techniques advance civil rights issues versus stratifies society becomes imperative to improving education, equity, and opportunity for all individuals.

Questions for Data, Civil Rights, and Education

1. What are the primary benefits and challenges of data-driven educational tools? How do we assess these changes?
2. How does data-driven educational policy and instruction increase or reduce discrimination or inequality of opportunity in the educational system?
3. If we want to optimize data use to improve equality of opportunity, equity, and reduce discrimination what are the variables we would want to consider?
4. Given the uncertainty of potential outcomes and the lack of empirical evidence to support data-driven educational reform, what is the best way to proceed with policy-making?
5. How can we ensure transparency, accountability, and due process in automated algorithmic systems when they influence a student’s path through the education system?
6. What policies or tools can we have in place to remedy errors, or to hold data-driven decision-making processes accountable? How do we identify which part of the calculation lead to a discriminatory result?
7. How can we foster transparency and accountability of third parties? How useful is transparency and access to one’s own information?
8. Who should have access to data about students, in what contexts, and for what purposes?
    a) How do we make certain to protect marginalized youth, including those from abusive households, those whose guardians are state actors, and those who are trying to alter their path?
    b) What rights should students have to distribute, sell, or otherwise make available their own data when asked (e.g., by college recruiters, coaches, etc.)?
9. How can data analytics be used to address historically marginalized communities and promote civil rights?

---

[1] Brown v. Board of Education, 347 U.S. 483 (1954).
[2] Presidential Task Force on Education, *Ethnic and Racial Disparities in Education: Psychology's Contributions to Understanding and Reducing Disparities*, (2012). http://www.apa.org/ed/resources/racial-disparities.aspx.
[3] National Assessment of Education Progress, U.S. Department of Education, *Falling Further Behind: Combating Racial Discrimination in America*, (2012), "Discrimination in Education." http://www.civilrights.org/publications/reports/cerd-report-falling-further-behind/discrimination-in-education.html.
[4] The Civil Rights Project, "Civil Rights Project, E Pluribus...Separation: Deepening Double Segregation for More Students," by Gary Orfield, John Kucsera, and Genevieve Siegel-Hawley (Los Angeles, CA: 2012). http://civilrightsproject.ucla. edu/research/k-12-education/integration-and-diversity/mlk-national/e-pluribus...separation-deepening-double-segregation-for-more-students/.
[5] The Civil Rights Project, "Why segregation matters: Poverty and educational inequality," by Gary Orfield and Chungmei Lee (Los Angeles, CA: 2005). http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/why-segregation-matters-poverty-and-educational-inequality/orfield-why-segregation-matters-2005.pdf; Elizabeth Lamura, "Our Children, Ourselves: Ensuring the Education of America's At-Risk Youth," *Buffalo Public Interest Law Journal* (2012-2013), 117, 127.
[6] US Department of Education, *The Condition of Education 2004*, By John Wirt, et al., NCES 2004-077 (Washington, D.C., 2004). http://nces.ed.gov/pubs2004/2004077.pdf; The Education Trust, "Teaching Inequality: How Poor and Minority Students Are Shortchanged on Teacher Quality: A Report and Recommendations by the Education Trust," by Heather G. Peske and Kati Haycock (Washington, D.C., 2006). http://files.eric.ed.gov/fulltext/ED494820.pdf.
[7] "While boys receive more than two out of three suspensions, black girls are suspended at higher rates (12%) than girls of any other race or ethnicity and most boys." Office for Civil Rights, U.S. Department of Education, "Civil Rights Data Collection, Data Snapshot: School Discipline 1," (Washington, D.C., 2014). http://ocrdata.ed.gov/Downloads/CRDC-School-Discipline-Snapshot.pdf.
[8] The Civil Rights Project, "Sent home and put off-track: The antecedents, disproportionalities, and consequences of being suspended in the ninth grade," by Robert Balfanz, Vaughan Byrnes, and Joanna Fox (Los Angeles, CA: 2013). http://civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/state-reports/sent-home-and-put-off-track-the-antecedents-disproportionalities-and-consequences-of-being-suspended-in-the-ninth-grade
[9] Natasha Singer, "Colleges Awakening to the Opportunities of Data Mining," *The New York Times*, July 22, 2012. http://www.nytimes.com/2012/07/22/education/edlife/colleges-awakening-to-the-opportunities-of-data-mining.html.
[10] Ryan S. Baker, "Data Mining for Education," in *International encyclopedia of education (3rd edition)*, ed. B McGaw, P. Peterson, and E. Baker. (Oxford, UK: Elsevier, 2010). http://www.columbia.edu/~rsb2162/Encyclopedia%20Chapter%20Draft%20v10%20-fw.pdf
[11] Jules Polonetsky and Omer Tene, "The Ethics of Student Privacy: Building Trust for Ed Tech," *International Review of Information Ethics* 21 (2014): 25, accessed August 5, 2014. http://www.i-r-i-e.net/inhalt/021/IRIE-021-Polonetsky-Tene.pdf.
[12] Baker, "Data Mining."
[13] Jeffrey A. Johnson, "The Ethics of Big Data in Higher Education." *International Review of Information Ethics* 21 (2014): 25, accessed August 5, 2014. http://www.i-r-i-e.net/inhalt/021/IRIE-021-Johnson.pdf.
[14] Office of Educational Technology, U.S. Department of Education, "Enhancing teaching and learning through educational data mining and learning analytics: An issue brief," by Marie Bienkowski, Mingyu Feng, and Barbara Means (2012), 1-57. http://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf.
[15] Office of Educational Technology, "Enhancing Teaching."; "Competency-Based Learning or Personalized Learning," U.S. Department of Education, accessed May 1, 2014, http://www.ed.gov/oii-news/competency-based-learning-or-personalized-learning.
[16] Johnson, "Ethics of Big Data," 7; Baker, "Data Mining."
[17] Varun Kumar and Anupama Chadha, "An Empirical Study of the Applications of Data Mining Techniques in Higher Education," *International Journal of Advanced Computer Science and Applications* 2, 3 (2011): 80–84.
[18] Brijesh Kumar Baradwaj and Saurabh Pal, "Mining Educational Data to Analyze Students' Performance." *International Journal of Advanced Computer Science and Applications* 2, 6 (2011): 63–69.
[19] Meredith Deliso, "How Big Data Is Changing the College Experience," *Online Degrees*, August 23, 2012. http://www.onlinedegrees.org/how-big-data-is-changing-the-college-experience/; Marc Parry, "College Degrees,

Designed by the Numbers," *The Chronicle of Higher Education*, July 18, 2012. https://chronicle.com/article/College-Degrees-Designed-by/132945/.

[20] Audrey Watters, "Student Data is the New Oil: MOOCs, Metaphor, and Money" (presented as part of *Conversations about Online Learning* at Columbia University, New York, NY, October 16, 2013). http://hackeducation.com/2013/10/17/student-data-is-the-new-oil.

[21] Mark MacCarthy, "Student Privacy: Harm and Context." *International Review of Information Ethics* 21 (2014): 7; Sammy Mack, "Putting Student Data To The Test To Identify Struggling Kids," *National Public Radio*, April 8, 2014. http://www.npr.org/2014/04/08/300587823/putting-student-data-to-the-test-to-identify-struggling-kids.

[22] L.S. Hall, "The Other Half of the Battle: funders Bet on Push to Keep Low-Income Kids in College," *Inside Philanthropy,* September 25, 2014. http://www.insidephilanthropy.com/home/2014/9/25/the-other-half-of-the-battle-funders-bet-on-push-to-keep-low.html.

[23] Oscar H. Gandy, Jr., *The Panoptic Sort: A Political Economy of Personal Information. Critical Studies in Communication and in the Cultural Industries* (Boulder, CO: Westview Press, Inc, 1993).

[24] Elana Zeide, "The Proverbial Permanent Record," *SSRN*, October 9, 2014. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2507326; Kate Crawford and Jason Schultz, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms," *Boston College Law Review* 55.1 (October, 2014).

[25] Anya Kamenetz, "The Test- Why Our Schools Are Obsessed with Standardized Testing—But You Don't Have to Be," *PublicAffairs--New York* (forthcoming, January, 2015).

[26] MacCarthy, "Student Privacy."

[27] Ry Rivard, "Predicting Where Students Go," *Inside Higher Education*, October 19, 2014. https://www.insidehighered.com/news/2014/09/19/colleges-now-often-rely-data-rather-gut-hunt-students.

[28] Jules Polonetsky and Omer Tene, "The Ethics of Student Privacy: Building Trust for Ed Tech," *International Review of Information Ethics* 21 (2014): 25, accessed August 5, 2014. http://www.i-r-i-e.net/inhalt/021/IRIE-021-Polonetsky-Tene.pdf; Viktor Mayer Schönberger and Kenneth Cukier, "Learning with Big Data" (New York: Eamon Dolan/Houghton Mifflin Harcourt, 2014).

[29] Olga Kariff, "Privacy Fears Over Student Data Tracking Lead to InBloom's Shutdown," *Business Week,* May 1, 2014. http://www.businessweek.com/articles/2014-05-01/inbloom-shuts-down-amid-privacy-fears-over-student-data-tracking.

[30] Natasha Singer, "With Tech Taking Over in Schools, Worries Rise," *The New York Times*, September 14, 2014. http://www.nytimes.com/2014/09/15/technology/with-tech-taking-over-in-schools-worries-rise.html; Data Quality Campaign, "State Student Data Privacy Legislation: What Happened in 2014 and What is Next?" http://dataqualitycampaign.org/files/State%20Student%20Data%20Privacy%20Legislation%20Resource.pdf.