

Workshop Discussion Notes: Education

Data & Civil Rights
October 30, 2014 – Washington, D.C.
<http://www.datacivilrights.org/>

This document was produced based on notes taken during the Education workshop of the Data & Civil Rights conference. This document represents a general summary of the discussion that took place. Not all attendees were involved in every part of the conversation, nor does this document necessarily reflect the views and beliefs of individual attendees. All workshop participants received workshop materials prior to the event to spark discussion. The primer can be found at: <http://www.datacivilrights.org/pubs/2014-1030/Education.pdf>

Overview

Trust, control, and fairness were the three main themes of this workshop discussion. The participants discussed education data in the context of a two-tiered educational system, and they observed that the solutions being offered for the problematic parts of student data collection and analysis are not equally effective for higher and lower-income schools. In particular, giving students, parents, and teachers more access, control, and ownership of their educational data is often proffered as a palliative and a desirable outcome of the debates over student data that center on transparency and accountability, but this improved access to data will primarily empower those who know how to understand, analyze, and challenge that data, while further marginalizing those who do not have the ability, capacity, or resources to grapple with it. In other words, student data can be an effective tool for furthering the educational goals of involved, empowered parties. Moreover, this tool is particularly legible in the area of personalized learning. For less empowered parties, the benefits of more access to and ownership of data are unclear and diminished.

The group also discussed this divide in the access/control continuum and how it plays into the efficacy of using “informed consent” to bridge the gap between the fears and the hopes of data-driven education technologies. For those who can understand the way these new systems work, and how to negotiate them, informed consent may be valuable, but for those who do not a strong understanding of how data-driven systems work, the concept is less applicable for fostering a positive reception of or engagement with EduTech. Additionally, more information does not necessarily create trust in these data-driven developments.

Marginalized student populations may be distrustful of the organizations and platforms that mediate their education because of the ways they have been historically excluded from high-quality educational opportunities, and EduTech systems may not be designed in ways that are cognizant of student needs and situations across the board. For instance, being tardy to class may be different for a student who has major caretaking responsibilities at home than for a more privileged student, but the system might log it in the same way for both parties, and impose consequences accordingly. The anticipated solution is that even more data needs to be collected on students who do not fit the expectations of these systems, but then already marginalized and distrustful student populations are being asked for even more data and deploying further surveillance that can seem invasive from a privacy perspective. There is a lot of concern that the

disparities in education that exist in current systems will be re-mapped onto new forms of EduTech, and that the root causes of these will be harder to address because people will be posing challenges to opaque algorithms, rather than to specific stakeholders. The challenge of identifying or resolving what needs to be fixed, and how, may be beyond the technical understandings of the people on the ground (students, parents, and teachers).

Themes and Discussion Topics

The conversation centered on a series of questions and responses that are addressed below:

1. Is personalization helping or undermining civil rights goals?
 - a. Personalization can help make visible what opportunities there are for improvement.
 - i. However, there is no good data for early or higher education.
 - ii. There is a need to train educators on data literacy and on how to use information.
 - iii. However, having an outcome-only focus in the use of data can be problematic.
 1. For example, universities might not let people major in a certain subject if the data does not predict success.
 - b. Personalization could harm individuals by making assumptions for them about what materials they should see
 - i. The question is how to use data as decision support tools to improve decision making, but not remove agency from the picture.
 - c. Lack of transparency
 - i. Makes it difficult to understand why certain materials are provided to students.
 - ii. Students, teachers, and parents should not be entirely ignorant of student data, and an opaque algorithm should not be blindly believed.
2. What data do we have, or not? Where are there big data gaps, or data locked behind different structures?
 - a. Data is rarely used to make the case for equity in education.
 - i. There is a long history of using data to marginalize and discriminate.
 - b. In some ways, we have too little data on everyone, and use big, blunt data points to make decisions that those points cannot bear, for example, SATs.
 - i. Educational achievement should also take into account other data points.
3. How do we consider notions of fairness in data driven markets when we move out of the private sector and into the public?
 - a. Data is going to be used, and that cannot be stopped. We need to see where interventions can be made before data is misused.
 - i. When we see data being used, how do we try to guarantee structures are in place to promote the values of civil rights and equity?
 - b. Transparency is needed, including oversight and audit mechanisms.

- i. Further, we need to help schools navigate vendor agreements and learn what questions to ask.
 - c. Educators and administrators may feel that, based on their professional opinion, they can do a better job than the data, introducing a new source of bias into the system as they override the data-based decisions.
 - i. This may change as new educators enter the system and are taught about how they can use data.
 - d. Focus should be on the individual first, not how data can be used for general good.
 - i. Collective and individual benefits are often conflated, and are not always evenly distributed.
 - e. Need substantial controls to foster trust of data.
 - i. Problem is that there is no federal monitoring system, and education is instead governed from state to state and district to district.
 - ii. The Department of Education does not have very strong mechanisms for enforcing civil rights.
 - iii. Groups with the most at stake have the deepest distress because data has been used against them in the past.
 - iv. Need to teach with technological reasons that data is safe.
 - 1. There is no argument you can make about de-identification that is going to make people comfortable with information being captured if they do not understand the benefit.
 - a. Unlike with data collection in the private sector, there is no immediate quid pro quo, and the benefit people get from handing over their data is unclear.
 - b. On the other hand, people often do not understand how private companies use their data, so their consent is not necessarily meaningful.
 - v. Need to avoid the urge to rush to benefit and push data to realize a benefit immediately. We need to think critically about use and not be distracted by possible benefits.
 - f. People also distrust data because it is mediated through subjective decision makers that people, particularly minorities, do not trust.
 - i. Many people do not think that data is fairly collected.
 - ii. Data is also often mistreated. It operates in the same social construct as everything else, and results that tell a story contrary to that construct are sometimes ignored.
- 4. What would you like to see of some of the approaches discussed if we realize technology is not going away and data will continue to be part of the conversation?
 - a. There can be value for subjective decision making in education, or teachers would not have a role. That variable should be built in.
 - b. Need to control against biases:
 - i. Need to discuss conscious and subconscious biases.

- ii. Need to consider which mechanisms and rules are effective, and create a system with a focus on accountability.
 - 1. Such a system should look for equity based on multiple well measured outcomes.
- c. Transparency itself is not a panacea, but accountability is important.
 - i. We should think about the kind of mechanisms that can be articulated in plain language and made available for policy makers, educators, and parents and students.
 - 1. For example, parents and students should be able to access their own learning information and challenge and correct it using a modern interface.
 - ii. However, even with full access to data, electronically sophisticated people cannot always work out what needs to be done.
 - 1. School districts can drown parents and students with data so they cannot sort what matters from what does not.

Areas for Further Exploration

The topics of transparency and accountability came up recurrently as areas for further explorations. Neither is a panacea for the challenges of data-driven systems, but there should be more conversations around the kinds of transparency and accountability mechanisms that can be articulated in plain language and the mechanisms of oversight that would be effective. This could help build trust, and ways for more marginalized student populations to participate in technical reforms to education at the level of a user interface.

The other topics that participants expressed a strong interest in discussing further is the way that data systems can reveal how children rank in comparison to each other in a much more granular way, and how they might be labeled in ways that are harmful or which follow them throughout their education. How could a database full of student data be used to create predictive algorithms that generate inclusive or exclusive opportunities? In what ways could controls be established on the mobility of student data, or to otherwise limit its usage to prescribed intentions?