2
The Anonymity Spectrum

Regardless of where you reside within the academy, if you are involved closely in the field of educational research you will have had to contend with sorting your work into methodological buckets. Are you, for instance, a quantitative researcher focused on large-scale samples of educational phenomena? Are you, instead, a qualitative researcher seeking to explore nuances of phenomena that eschew counting and numeracy? Or maybe you present yourself as the ever-elusive mixed-methods scholar, likely with heavier leanings toward one methodological paradigm than another, but aspirational in walking the straight and narrow of both approaches regardless. The factions of methodological work are important, but they often take up so much of our imagination as a field that it could be easy for early career members of the educational research community to see these factions as the primary means of understanding what you do and—worse—*why*.

But let's set aside any qualitative-vs.-quantitative feuding. It has gotten in the way of a better understanding of the data we harvest from the sowing of educational futures. As one possibility, we (Matthew and Antero) have been wondering what our research efforts might look like if we centered an understanding of how our data come to know the individuals and groups they target and how our

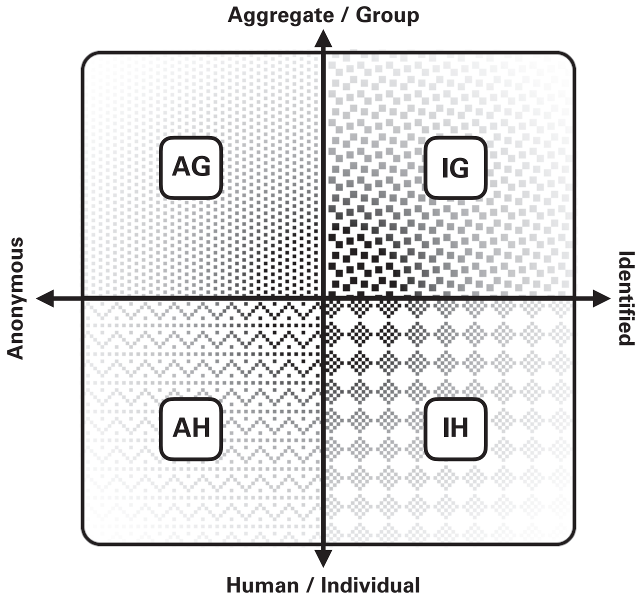

Figure 2.1
AnSpec (the Anonymity Spectrum).

data align with the outcomes of research processes. Thus, this chapter proposes a lens on data and education through which we can see a spectrum of design and interpretive possibilities in education, which we will call AnSpec (short for ANonymity SPECtrum). AnSpec is a two-dimensional space, as shown in figure 2.1. It allows us—as researchers, educators, and varied stakeholders invested in the lives of learners—to make intentional decisions about and with data that comprise various aspects of how we understand people: wholly, individually, specifically, and broadly. Organized across four quadrants, this spectrum helps elucidate the degree to which data are found within the range from anonymity to identifiability and from aggregated ("grouped") to individualized.

It is important for us to critique these axes. Why these specific axes? What do these axes erase? What values are implicit in these axes? What do they say about culture, history, or power? There are many ways of approaching the goals of an intervention, of a design, of a collection of numbers; AnSpec pulls into focus considerations of design and biases of data that are regularly overlooked or backgrounded in the design process. These axes represent an extant discourse; it is (necessarily) not the best possible discourse. Over the course of this book, we will both return to AnSpec as a lens and problematize it as a tool.

We can find numerous examples—which we refer to throughout the rest of this book—that exist along the outer bounds of the space. Large quantitative data sets and microanalytic qualitative studies parade the affordances of data along the periphery of this spectrum. Antero's ethnographic work often attempts to attain an $n = 1$ approach to data, getting as close as possible (which usually makes such data very identifiable as well). In the same way that we could map our own research into a few general clusters on this methodological map, we want to recognize that this is by no means a definitive reflection of current research practices. By no means are we arguing that AnSpec is the *best* representation of possible discourses for data, or for modeling data and privacy in education. Rather, this is one approach at *seeing* what is occurring from a vantage point not often perched upon within our fields.

So, what is the value?

Education and education research frequently background the decisions that AnSpec highlights: *Why* is a study collecting anonymous, grouped data? Why do studies about individualized student performance (like the Star test administered in Antero's classroom) build from and toward specific assumptions about student achievement? Alternatively, how might we understand the ongoing "reading wars" that have shaped US literacy education models for decades (e.g., Pearson, 2004) if the values imbued in these different stances were mapped across AnSpec?

Corpora (that is, large data sets) carry unspoken rationales. In this sense, rather than further endorse dominant modes of data collection, AnSpec functions to critique and analyze existing data sets and to facilitate more purposeful data decisions in our scholarship moving forward. As we explore the possibilities of data for transforming educational systems and contexts, we need to admit that—even when we are wary of blind faith in algorithms and their authors—we are drawn to the potential of using logged data to support people in learning and schooling. At the same time, educators, researchers, and the public today tend to use the term *data* with enough vague hand-waving as to render the term meaningless. As Hume (2007) suggested, what we take as evidence is not defined by abstract reason but instead by whatever habits we use to support our arguments. The evidence—or "data"—can be evaluated, analyzed, and (mis-)used from any perspective or framework; any understanding of data is an appeal to habits and mores. In the rest of this book, we write with a more familiar sense of *data* as "numbers generated about a person." This is made more complicated by the fact that anything encoded digitally—which is to say, almost all the modern world—is encoded numerically. That observation is not a "gotcha" or an idle notion. As computing power has grown, phenomena that seemed relatively resistant to "data analysis" (e.g., essays, photographs, scents) have become so obviously numeric so quickly that it is sometimes a surprise to find them in a purely analog form within an organization such as a school or business. The quantization of data is inherently reductive, but so are words, images, and thoughts. Any perspective is a limited perspective.

A myriad of research considerations, protocols, and workshops are offered and sold to interested and frustrated scholars, educators, and policy analysts alike. Does the world of justice-driven education research *really* need another framework for interpretation? Not necessarily! However, we suggest that what separates AnSpec from other approaches is how its design centers the possibilities of *what if* while acknowledging the horizons and limitations of what we do (or don't) see as we conduct our research. We must acknowledge the humanity at the center of research. AnSpec seeks to reorient some of our scholarly practices around the *whom* and the *what for* of our work. Clearly, we will be exploring research applications across the four quadrants of the spectrum in substantial detail in the pages that follow. However, we first label and describe our axes here.

Complexifying Anonymity: The Axis of Identification

When we think of our personal information as data, it is worth asking what about us is being identified. Who are we? In daily life, we may range from being uncomfortable giving our last name (to, say, an unknown caller) to filling out a form with security questions that require information about a pet or favorite aunt. Often, these two responses may well be given to the same company in different forms. In fact, Matthew had a bad experience canceling a gym membership in the past, so when he signed up for a new gym, he went so far as to pay for it with a burner card. (This card maneuver did not help when he finally canceled this membership.)

Identification is a mode of specificity about personal information, even when flexible and contextual. Goffman (1974) shows that “who we are” is contingent on the “frame” in which we find ourselves. Sometimes we are parents, sometimes we are professors, sometimes we are really bad at sports. These identities are not all simultaneously enacted when one of us plays a video game online, though they undoubtedly affect our behaviors. Indeed, the inflexibility of identificatory data is a growing issue. If, for example, marketing algorithms infer that you are a mom, you will likely stop receiving “dad” or “nonparent” shopping suggestions.[1]

Identification is a range, and there are tiers to it: some identity data are effectively irrevocable. You must really trust someone to be willing to give them your bank account information or Social Security Number. If you choose to do so, your life may be further shaped by that decision. Perhaps you are bankrupt now? (We hope not.) Perhaps you've revealed sensitive information about yourself in one situation (if you are pregnant, disabled, or economically disadvantaged, for example) that found its way to a university institutional review board or potential employer. Or perhaps inferences based upon your browsing and shopping histories delivered incorrect information to third parties. Often the identifications made about you are wrong. They are automatically inferred based on statistics. Anecdotally, Matthew knows a woman in her seventies whom an ad network has labeled a man in his thirties. (This is likely due to her love of first-person-shooter video games.) These identificatory labels are typically not factually incorrect on average—indeed, a lot of financial investment goes into making them "correct enough" for sales and marketing purposes—but, unsurprisingly for a regression to the mean, they consistently tend toward the (pre-)dominant population. The central limit theorem abides. If 97 percent of people who play game X are verified as men, then, without further information, almost any algorithm that uses those data will decide that, if you choose to play this game, you are probably a man. Gender scholars—and millions of others—have broken down these dichotomies and demonstrated the complexities of gender as a construct (e.g., Butler, 2002), but most marketing algorithms do not care who you are so much as how you will generate money. Indeed, if you are willing to buy what the other 97 percent of people labeled as "men" are also buying, you, to this algorithm, are an "effectively-a-man": gender, in this case, is a marketing construct.

If the labeling largely occurs on the tail end of an algorithm's determinants in various social contexts today, the axis of identification is moved toward the agentic process of supporting people in controlling how they might be identified or valued in research contexts. We then have four nonoverlapping categories: 1) how *you* care about yourself as a person; 2) how others care about you as a person; 3) how researchers or educators understand you as a person; and 4) how the data have constructed you. *As such, we consider identity data to be the set of characteristics, labels, information, or behaviors drawn from you and/or applied to you.*

Importantly, identity data changes quickly. What is not identifiable today may be identifiable tomorrow. For instance, people often use "voice" as an identity characteristic when we are physically proximal to each other. If a loved one calls our name, we often know who it is. Computers do not typically, at least at this time, register this type of identity characteristic. Though it seems as if voices are not so obviously unique at a global scale, there is no apparent reason to believe that they will not be identifiable someday soon. What counts as identity data today? What counted yesterday? What will count tomorrow? We will come back to these questions as we progress through the book.

Complexifying Individuals: The Axis of Aggregation

If data on one person are individualized, then data about multiple people are called *aggregated data*. That said, the concept of aggregation (or "grouping") contains obvious gradations. A group of two people ("Emma" and "Alexander") exists more or less only at the personal identity level: we can ask questions about them such as "What are your names? Where are y'all from?" and get answers that anyone who speaks the same language can understand. Scaling that up to a global level is absurd. What does it mean to name seven billion people? You could not feasibly hear every name even if you spent your entire life listening. That said, the group of Emma and Alexander and the group of "Humans on Earth" are both aggregations of a sort.

Aggregating at the level of groups—*stamp collectors*, *women*, *humans*, *those who have just broken the vase*—carries serious sociopolitical implications, but every model of cognition and every algorithm requires some level of abstraction. That is, aggregation is abstraction; Quine (2013) suggested that it is, in part, the naming itself that brings things into being as an object of consciousness. Similarly, we unconsciously group people by any number of their characteristics such as their appearance, behavior, and our prior experiences with them. Cognitive load theory roughly suggests that we can only keep a few things "in mind" at a given time: if we see twenty people, we have grouped them into fewer than ten categories, whether we like it or not.

This kind of cognitive grouping (or aggregation) happens, quite literally, always. While the bar for "perceptible features our mind might care about" would be much lower in a lightless, silent, odorless room than, say, Times Square, no matter where you are or what you are doing, there are more than ten features perceptible to your mind. Algorithms are built by conscious humans trying to solve problems, and their grouping schemes will often look much more like dominant, labeled grouping schemes (e.g., race, language, and gender) than we might expect. As suggested so eloquently by Noble (2018), algorithms cause action in the world based on these grouping schemes. Everything from social media to school to marketing to predictive policing rely on human-cognition-scale groupings, and it is humans who must parse the results. Matthew's videogame-loving acquaintance in her seventies will be offered advertisements for consumable goods that would be unsuited to her, such as student loan debt consolidation (not to suggest that she couldn't go back to school for another doctorate!). That said, there are statistical regularities to these explicit, demographic groups: Black people have been systematically oppressed contingent to their identification as Black, so Black may be a relevant and just category to consider, and/or it may just encode and reinforce racial bias. Due to that aggregation-oppression, Black Americans have been systematically barred from many opportunities. A relatively simplistic (and broadly unhelpful) response from many companies has been to aggregate in ways that do not lend themselves easily to human-comprehensible labels—or at least not to *report* those labels, so they do not report how they are replicating oppression. However, ontogeny recapitulates phylogeny; that is to say, the ways that things come into being are manifested in the ways that they act in the world. It is a trick, therefore, to pretend that when we aggregate without using political terms, we can avoid racism. (This is sometimes called "race evasiveness," per Annamma et al., 2017.) For instance, one may avoid using explicit racial/ethnic categories (such as Black) but substitute, for instance, users of specific personal care products designed for Black communities. It becomes clear that if we are not conscious about how we are replicating oppression, and actively keep working to counter that, we are almost certainly going to replicate the oppression. Hiding it in black boxes does not help; it hurts. So, the thorny question remains! We will aggregate—our brains do not allow otherwise. But how do we aggregate without recapitulating systemic oppression?

One answer is to make a conscious point to notice instances of systemic racism and actively confront that racism to whatever degree possible (per Kendi, 2019, for example). Rather than letting statistical aggregation find the "best fit" labels, we evaluate situations for the potential effects of bias, apply explicit labels, and try to mitigate or even reverse historical damage. Data are still required for this evaluation, but the act of aggregation or disaggregation itself becomes a political act.

We consider aggregated data *to be data that groups people*.

Identification and aggregation are rarely independent of one another. Indeed, their interaction produces a variety of unlikely results, some of which we will discuss later in this book. The rest of this chapter offers a quadrant-by-quadrant tour of AnSpec, while much of the rest of this book explores specific, applied examples of how AnSpec shapes contexts of learning, research design, and analysis. In the rest of this chapter, we focus on behaviors and characteristics of clustered data.

The Four Quadrants of AnSpec

Referring to figure 2.1, the four quadrants of AnSpec are named after the general areas that are represented. However, we want to re-emphasize that AnSpec is not a pair of binaries. While data can be imagined as either anonymous or identified and individual or aggregate, there are degrees to all of these considerations, and it is this indeterminacy that helps elucidate reimagined contexts and opportunities for design.

Anonymous Grouped Data

Anonymous Grouped (AG) data are those data that are purely anonymous and always exist in aggregate. An example of AG data is the data collected from a traffic counter: one of those strips that lies across a road and counts the number of wheels that roll over it. It is typically difficult (or even impossible) to identify any specifics about the individual humans who generated these data. One cannot guess their names, addresses, or, really, anything except for the most general possible information about whether they drove over a traffic counter in the middle of a city. It is possible that these individuals do not even live or work in the city, depending on the location of the traffic counter. Perhaps you know some external information about the socioeconomic status or cultural background of people who live in the neighborhood, but nothing about the data itself would be enough to identify anyone.

Often, AG tasks require local information that requires little contextualized information; if you are a potato seller and people in your town start eating 500 potatoes a day, you do not need to know much about those people, but you know that you should buy more potatoes. In the case of the traffic data, it is rare that one would use an individual line of these data for much, as it would mostly just read "a wheel went over me at time X" and, given that tractor trailers exist (which have many more wheels per vehicle than your typical family car), one might not even be able to estimate the number of vehicles, depending on the specific path. Despite the anonymity, these kinds of data can make lasting changes on the landscape of education, as we'll continue to explore.

AG data currently affect how we use data for learning and will continue to do so in the future. One nice example of the use of AG data for learning environments is the example of a room at the British Museum (in Savoy et al., 2014). To better understand what to highlight, the museum wanted data about who was perusing which exhibits. Unfortunately, the data from the exhibit itself was minimal, and they collected location data at the broadest level: how much do people walk through or stand in each spot? Taken at an individual level, such data are nonsensical; there is not an obvious reason to make it identifiable. However, at the aggregate level, the data helped the museum identify a fundamental misstep in spatial design. The data about visitor walking paths revealed that visitors were frequently missing a whole chunk of space in the exhibit. What's more, while walking through the space, visitors often even failed to see that they were missing a set of exhibits, depending on the angle to which they were walking into the space. By using the pathing data, the museum found that, with some relatively minimal rearrangement, they could enable people to see all the exhibits. The museum's analysis of location-tracking and path-tracking data highlight some potentially beneficial applications for AG data in educational contexts.[2]

Corporations also use AG data extensively. One example is a classic "A/B test," in which one group ("A") gets one form of the software relatively invisibly, and another group ("B") gets a close contrast case. This is one of the most common forms of testing for software companies, and these tests tend to be anonymous and grouped. Most large online retailers reportedly require A/B tests for every public-facing minor change to their sites—even just a simple wording change in the middle of a license agreement. Google has probably tested you today: you may have searched, and a link was almost imperceptibly shorter or longer. If tested on enough people (~100 million, perhaps), they can decide if this nearly imperceptible difference produced any meaningful change in user behavior, such as clicking on an ad or spending more time with a site. Because these corporations are testing at a global scale, anonymity is effectively assured; they could not care less if you, in particular, clicked on an ad in this A/B test, as the difference may be 0.001 percent or less between the two groups, or the difference may be a couple of milliseconds of load time on average; however, these small differences may still translate into millions of dollars in profit over the life of that button. At a much smaller scale with more contrasts, similar A/B tests are at work across education, especially when it comes to educational technologies. Large edtech companies routinely analyze Web traffic, currently an evergreen resource for AG data. AG A/B tests would seem to be broadly neutral (or even good); as we will discuss later regarding anonymous individualized data, however, A/B tests at scale are not inherently innocent.

Popular sentiment today suggests that people are often resistant to having their location tracked. A general allergy to surveillance—anonymous or individualized—is perfectly healthy; the authors share this affliction. However, in situations with AG location data, such methods may be harmless. Making students and other users aware that their data is being collected or tracked is an ethical consideration that differs greatly depending on whether a notification is in the pages-long gibberish of a digital user agreement, part of a complicated protocol for an institutional review board, or ignored entirely by a municipal worker placing a traffic counter on a road in a nearby development area.

This is not to say that all data uses require complex technology such as location tracking: for example, schools often decide to replace a set of textbooks based on which ones have been thrown away or have gone missing. Assuming that such accounting of books lives in some administrator's Excel sheet, over time such information becomes grouped and anonymous logged data. The changing nature of information over its lifespan means that data may move from one quadrant to another.

Identifiable Grouped Data

Identifiable Grouped (IG) data, such as surveys and polls, are a common form of mass data collection. Most of these data are only valuable at scale. The census is, perhaps, the most canonical example: census data are detailed and specific enough that any single line of the data would likely be sufficient to identify an individual. Information such as address, zip code, telephone number, name, age, and income level are grouped for analysis, but exist at an identifiable level.

IG data are, somehow, both dangerous and anodyne. Improbably quoting Stalin while ruminating on both COVID-19 and *The Fast & the Furious* film franchise, a ten-year-old recently said to Matthew, "A single death is a tragedy, a million deaths is a number." The fact that this quote survived (filtered) through a grade-school classroom shows the power of IG data. Indeed, identifiable grouping is wielded by entities like the US's Immigration and Customs Enforcement (ICE) in justifying the rounding-up and imprisonment of people. Sources of such data do not allow for understanding individuals, only for categorization and, often, weaponization.

Let's be clear: demographers are not evil by trade, and, without clear data on who is where, doing what, we have radically impoverished (and presumably dominance-biased) views of the world. Governments that recognize the power of their populace and the diversity of identity can address and support that identity as much as target it. The potential for IG data to uplift as well as to oppress depends on the nature of the questions being asked, the purposes of data collection, and the autonomy and agency of the organization at the center of such work. A government census, for instance, can wholly exclude or bring in nuanced perspectives of a country and how it should be organized.

However, there is something quite powerful about undermining the IG nature of data. There is (or was) a Twitter account (Zhang, 2015) that attempts to humanize the census, and it is quite powerful and revealing. Making clear the actual individuals who are at the heart of each entry in the US census, the account reframes viewers' interpretations of what politicians talk about when lumping individuals into broad categories. A sample tweet: "I worked less than half of last year. I studied civil engineering. I moved last year. My commute is 12 mins long. I drive by myself."[3] There is an unmistakably sad tone to these tweets, though they simply report the public data from the census. By humanizing a single line of data, the data that a census collects makes each individual into a story—a movement toward disaggregation.

IG data powers most advertisements we see, both offline and online. A company wanting to advertise on Facebook, say, picks sets of demographic categories ("commute less than 20 minutes" or "African-American") and ads are sent to the feeds that match those demographic categories. Users of Facebook must trust Facebook (or be unbothered by *not trusting* Facebook) with their specific demographic and individual information, presumably with the hope that their data are somewhat aggregated so that advertisers do not just show up at their house and attempt to sell them T-shirts, fertility monitors, or machine-washable shoes. In short, IG data are everywhere, and they contain within themselves their contradiction: the gloss of aggregation emotionally enables us to give identifiable information under the (potentially mistaken) belief that we are not then personally identifiable.

Identifiable Individualized Data

Identifiable individualized ("IH" for Individual Human) data are data about individuals: your social security number, your demographics, your name and address, your grades, your family members. These data can do immense good or immense evil, and they are a powerful tool. The tradeoffs for IH data might seem reasonably obvious to many readers; they exist as a function of trust. There are things that identify you that you should (and probably do) only tell your partner, children, or parents. Very few people give their social security number out freely. In fact, it makes national headlines when people do this: the *New York Times* reported in 2008 that when the owner of an identity fraud service named "LifeLock" did this, his account was hacked within days.

A major question of IH data revolves around what "counts" as purely IH data. Do you give your name and address freely? Matthew gave his name, address, and phone number to a cable company representative recently, knowing that his privilege as an adult cis-het man endows him with a smaller set of identities to be exploited—he is treated as a "default." We have had family members targeted for harassment simply for being educated women online, colleagues targeted for being Black online, and friends doxxed for expressing what-should-be-uncontentious views (such as "kids dying in school shootings is a bad thing that happens"). Sharing IH data is fraught, and it means something different to many people, just as "identity" does. Obviously, "being Black" and "being a birdwatcher" exist at different levels and modes of identity, but both may serve as identifying or sensitive if they can be tied to behaviors. Do you want people to find you and exploit the triangulation of your data?

IH data serves as a nexus of how the power relations that govern society and culture at large come to impinge on the agency of the individuals within it. It is too often used to consolidate power in existing aggregations. That is, power and capital tend toward individualizing an already powerful white man while marginalizing most others.

Anonymous Individualized Data

Anonymous individualized ("AH" for Anonymous Human) data are those data through which there is effectively no possibility of finding any identifiable information about a person, but the situation or data could lead to changes for that person. For instance, when you select “no dairy” for the menu at a wonderful restaurant, Clover, in Cambridge, MA, the information is logged and the menu is adapted to fit these specific needs. Many videogames use such data extensively. It is hard to determine someone’s name or address or many meaningful features about them based on how they play a game like Tetris, but the game can adapt to their specific play style; for instance, the game might drop the T block more often or less often based on how often it is part of a successful Tetris.

AH data is useful in our quest to try to implement just data analyses, but it has a core weakness: it is inherently unstable. The more AH data one collects, the more it tends to transform into IH data: by triangulating seemingly innocuous or opaque data, you can often learn a lot about something or someone. Classen et al. (2015) were able to identify individuals by vibrations on windows; Zeng et al. (2010) were able to identify individuals simply by evaluating how long it takes for their web searches to finish loading. Data are rarely perfectly “clean”; rather, they bring along the detritus of ancillary information. Over time and with different lenses, data that once lived in one AnSpec quadrant metamorphosizes; the AH caterpillar may balloon into an IH butterfly, given the unsuspecting tools or contexts. Often, you can jettison or avoid some ancillary information, but it’s rare that companies, for instance, would want to jettison those data, or that one could prevent the slide toward identifiability. This slide is one we will continue to explore in the next chapters.

Limitations and Missing Axes

The quadrants in AnSpec bleed into one another. Because our contexts shape how our information becomes legible to different algorithms, these axes become unfixed in time and space. Such are some of the key limitations inherent in each of these quadrants.

The processes of quantization are essentializing processes. Knowing that data vary in quality, quantity, type, and structure, the AnSpec framework is a starting place for outlining how individuals are valued (literally and symbolically) in our work as educational researchers. Through essentializing data, what features do we keep? What is foregrounded? What is backgrounded? Indeed, does the quantization of human experience itself inherently reduce the human experience? To say “yes” would be a simple denial of the last few centuries, but some days that may not be such a terrible thing. In the next two parts of this book, we will explore relationships among data, security, and justice; play through examples, applications, and code; and practice using AnSpec as a tool. We will attempt to frame justice as a process and an orientation rather than an example of our nifty tool, but it is incumbent on all of us to keep that in mind.

The Left Hand of Data

Designing Education Data for Justice

By: Matthew Berland, Antero Garcia

Citation:

The Left Hand of Data: Designing Education Data for Justice
By: Matthew Berland, Antero Garcia
DOI: 10.7551/mitpress/14381.001.0001
ISBN (electronic): 9780262377645
Publisher: The MIT Press
Published: 2024

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open

The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Berland, Matthew, author. | Garcia, Antero, author.
Title: The left hand of data : designing education data for justice / Matthew Berland and Antero Garcia.
Description: Cambridge, Massachusetts : The MIT Press, [2024] | Includes bibliographical references and index.
Identifiers: LCCN 2023030088 (print) | LCCN 2023030089 (ebook) | ISBN 9780262547529 (paperback) | ISBN 9780262377652 (epub) | ISBN 9780262377645 (pdf)
Subjects: LCSH: Education—Data processing. | Education—Research. | Educational evaluation.
Classification: LCC LB1028.43 .B45 2024 (print) | LCC LB1028.43 (ebook) | DDC 370.285—dc23/eng/20230718
LC record available at https://lccn.loc.gov/2023030088
LC ebook record available at https://lccn.loc.gov/2023030089