

Contents

Preface vii

Acknowledgments xi

PART I: THE BASICS OF DATA USE AND TEACHING

1. **What Is Data?** 1

Levels of Data 1

What Educators Talk About When They Talk About Data 3

Validity and Efficiency in Data Use 4

Genuinely Big Data 6

Summing Up 8

New Direction: Building a *Data-Wise* Culture in Every School 9

2. **What Is Teaching?** 13

Deep Dynamics of Teaching 13

Surface Behaviors of Teaching 20

Summing Up 29

New Direction: Learning to Rehearse Teaching 31

3. **What Is Data Use in Teaching?** 33

The Innovation's Theory of Action 34

Origins of U.S. Data Use in Teaching 36

Data Use in Teaching in New York City 39

Teaching in a Material World 41

Data Use and Poverty in New York City 43

A Teacher's Voice 47

Summing Up 50

New Direction: Curating the Material World 53

PART II: DEEP DIVES INTO DATA USE IN ACTION

4. **Complicated Plumbing at the Heights** 57

Spreadsheet after Spreadsheet 57

Observation of a Lesson	60
Teaching in Centers	61
The Data Coach	63
Is There a “We” at Heights?	64
A Trust Leak	66
Summing Up	68
New Direction: Distributing Leadership	69
5. Action Space at Bayside	71
Deep Down	71
Higher Up in the Bayside System	75
The Bayside Action Space	83
Summing Up	87
New Direction: Encouraging Student Agency	89
6. A Vortex at South Falls	91
South Falls Elementary School	91
Teaching in a Vortex	96
Summing Up	103
New Direction: Building a Teacher Learning Community	104
7. Implementation at the Waterfront	107
Waterfront Middle School	107
Zooming Out	109
Zooming In	116
Summing Up	124
New Direction: Expanding Space and Time for Learning	125
Conclusion: Major Findings: The Reader’s Interpretive Turn	127
Appendix: Our Own Data (and Data Use)	131
References	135
Index	145
About the Authors	152

What Is Data?

The word *data* originated as the plural of the Latin word *datum*, but it is used now also as a singular noun, and we follow that convention in this book. At the most elemental or raw level, data consists of empirically discerned traces of some phenomenon of interest as captured by some device. In education, the device may be a voice recorder, clicker, notebook, camera, laptop, pop quiz, standardized test, teacher-created test, or notation by hand on a checklist or rubric. The use of the word *data* signifies that the traces of experience are to be taken as facts—not in some incontrovertible sense, but in the sense of legitimizing their usefulness for further thinking and action. As Bryk, Gomez, Grunow, and LeMahieu (2015) suggest, use of the word *data* also implicitly involves an embrace of variation—an acknowledgment that what teachers and schools pursue can only be achieved in terms of some distribution of outcomes. In our minds, the word is also associated with what is called *knowledge management*—learning to work smarter by sharing, discussing, and leveraging information (Brown & Duguid, 2000; Nonaka, 2007; Wenger, McDermott, & Snyder, 2002).

LEVELS OF DATA

Two levels of data are particularly relevant to schooling and teaching today. In this book, we use the terms *big-test data* and *intimate data* to refer to them. Note that in most parts of the book, we avoid the terms *formative* and *summative*, which are more associated with function than scale, and might prove confusing here. In this book, we continually zoom from big to intimate scale and vice versa. What we call *big-test data* is the kind that districts or charter management groups get back from states and standardized testing vendors, that they add to for purposes of reporting to state education departments and the federal government, and that they share with principals and increasing numbers of school-level data managers or “coaches.” This data derives largely from federally mandated annual testing in literacy and math, and from other policy-mandated testing—for example, of English language learners (ELLs).

Intimate data is student performance data that practitioners collect themselves. Some of this is associated with what is often called formal

classroom assessment—for example, teacher-created quizzes, tests, projects, or exit tickets (quick reports from students on what they learned from a particular episode of teaching). And some is associated with informal classroom assessment—by means, for example, of teacher observation and questioning, or post-teaching reflection. It is important to note that some intimate data is associated with *standardized* assessment. That is, it is collected and analyzed in formally specified ways, and often stored in online data management systems designed to analyze it. Still, teachers themselves collect it, and gain in the process an intimacy with what it reports. For example, a widely practiced form of standardized intimate data gathering in the schools we studied—both the elementary schools and the middle schools—involves *running records* or formal inventories by teachers of their students’ oral reading skills and reading problems.

For the most part, we deal in this book with data at these two scales. Yet in using the term *big-test data*, we deliberately evoke the term *big data*—or what we call *genuinely big data*. This is the kind of data that comes in sets much too big for your laptop to process—in other words, a third scale of data. Examples of genuinely big data include data from the vast universe of digitized buying and selling, managed so as to predict, produce, and deliver shampoo (or something else) that you think you need; the billions of pixels from sophisticated cameras in NBA arenas, which are analyzed to discover best court positions for successful basketball shots; and, in education, the covertly captured user data from thousands of online learning programs that are analyzed in order to create new online programs or to improve existing ones. We have more to say about this third scale of data at the end of this chapter, though it is mostly speculative. In fact, the biggest data sets we saw in the schools we studied were merely compilations of big-test data and administrative data (for example, regarding attendance or ability/disability status). And as it turned out, school-level data managers actually did sometimes manage these data sets on their laptops.

Even if *big-test data* is not nearly as big as *big data*, the distinction between it and intimate data is important to make and to ponder. Big-test data plays a crucial role in U.S. schooling today, as does its intimate cousin. And the difference is palpable for people who work in schools. Calling attention to the difference is important, we think, for developing systems to support responsible data use at *both* scales. As it turns out, nearly everyone we interviewed or observed (both teachers and school administrators) regard both big-test data and intimate data as valuable. However, most also believe that school outsiders—whom our interviewees often referred to as “they” or “them”—do not appreciate the value to teaching of intimate data, particularly of the nonstandardized variety.

In coding our own data from this study, we distinguished between references (by any words) to big-test data, and references (by any words) to intimate data. Our initial analysis based on this coding suggested that our

research participants were 2.25 times more likely to talk about intimate data than about big-test data. As we looked more closely at the coded segments of transcript, however, we discovered that a reference to one of these (no matter how many times uttered) was never very far from a reference to the other. We think this suggests that school-based educators today are living with—and often thinking about—the associations of and tensions between big-test data and intimate data. We think this is healthy, and we wish more policymakers were doing the same.

WHAT EDUCATORS TALK ABOUT WHEN THEY TALK ABOUT DATA

In order to discern patterns of meaning in school-based educators' talk about data, we drew on the 75 analytic memos of our research transcripts that we coded as *what's data* (see the Appendix for further details of our research coding). We found that the number one data category educators talk about (at 29% of the memos) is data about what students know or don't know in terms of knowledge and skills that teachers specifically and deliberately aimed to teach them. This is hardly surprising, given that both big-test assessment and intimate classroom assessment focus resolutely on this target. Obviously, teachers need to know whether or not their students learned what they taught. There is a problem, however, if this focus crowds out opportunities for teachers to learn other crucial things too. Happily, one of these other crucial things comes next at a respectable 21%: data about students' thinking processes. Of course, this finding does not mean that educators in our study actually devoted this much of their engagement with students to figuring out what students were thinking. In fact, in our observations of teaching, we found that teachers often passed up opportunities to ask (in the face of "a wrong answer" or a confusing one) something like "Can you say more about what you're thinking?" Still, the code count here at least suggests strong interest in the matter.

What comes next in our memo code count, however, is disappointing. Only 6% of data-related talk was about students' misconceptions. Teachers' interest in misconceptions is crucial not only to help students unlearn "wrong answers," but at a deeper level, to help them displace naive mental models with more sophisticated ones (National Research Council, 2001). And even less of the data-related talk—at 4 percent—was about students' metacognitive skills and inclinations, or students' understanding of their own thinking and learning. Yet metacognition, according to learning science, is a major path to deep learning (National Research Council, 2001).

Beyond number counts, we were, of course, also interested in *how* the educators we studied talked about data, and with what degrees of subtlety and acknowledgment of complexity. One common pattern was what we came to think of as hyperrational, or cut and dried. Here, for example, is

a middle school principal responding to a question about how her school organizes for data use in teaching:

Okay, so once we get the state results—the exams—our data specialist makes charts, so that every teacher is given a data folder, and that data folder includes the results on the exam, the item analysis [which students answered which questions correctly and not], how many of the kids in that teacher’s classroom were ELLs, how many of the kids are SPED [eligible for special education services]. Then, in September, every teacher gives a diagnostic. In literacy, it’s a writing piece, and a running record in reading [individualized read-aloud with miscue analysis]; and in math it’s math skills; in social studies, it’s an essay about historical events. Then the teachers take that information, and they enter it into a tracking system that we’ve devised. It looks like this [shows a printout of a complex Google Docs spreadsheet]. And then we design the curriculum.

By “hyperrational,” we mean a system description that spells out data flow (as in charts, folders, and tracking systems), but glosses over interpretation (as in “then we design the curriculum”), or put another way, a focus on data collection but not data use. Such talk about data seems to suggest that data is an agent rather than information for an agent. It puts data, rather than teaching and learning, at the heart of *data use in teaching*.

VALIDITY AND EFFICIENCY IN DATA USE

The usefulness of data systems in all fields of practice depends on how well these systems ensure the validity of the data they process, and also on how efficiently they support the practices they are intended to serve—in this case, teaching and learning.

Validity

As a construct in assessment, validity has to do with the relationship of data to legitimate inference, and with the reliability of data systems over time in maintaining this relationship. We’re talking about a kind of truthfulness here, where the word *truth* covers such questions as: what writing really is, what students’ responses to a particular writing assessment really signify about them as writers, and whether the assessment captures this reliably across time and cases.

Teachers and school leaders need to consider validity continually—in-deed, whenever they consider questions of what students now understand, and what they should learn next. But this requires three things that tend to

be poorly distributed in U.S. schools. First, access to deep content knowledge—that is, content knowledge with strong disciplinary or interdisciplinary footing. This is needed to discern the value of particular big-test items to a larger domain, to create valid intimate assessment items, and even to devise and carry out valid teaching probes.

Second, many schools have poor collective understanding of standardized testing. In our research, for example, we found instances of schools basing decisions about what to teach in a given year on individual items from the previous year's big test. But the validity of a standardized test overall in accurately sampling some domain does *not* extend to individual test items. For example, knowledge of a particular word—say, the verb *articulate*—shouldn't be added to the 4th-grade curriculum this year just because it appeared on last year's 4th-grade English Language Arts (ELA) test. Some other word meant to discriminate among levels of 4th-graders' vocabulary will replace it next year.

Finally, many schools need to expand their collective understanding of how validity figures in nearly every teacherly move, insofar as the move is intended to probe for evidence of understanding. This involves coming to terms with the ways in which teaching and assessment intersect in practice—sometimes beneficially, as in skillful formative assessment on the spot, and sometimes detrimentally, as in too much practice testing.

These knowledge distribution problems may seem formidable—though only, we think, from the perspective of how U.S. schools have traditionally been designed—namely as cellular institutions. In fact, these problems are all learning problems—ones that schools can organize themselves to address—for example, by means of teacher learning groups, better materials curating, and collegial coaching. We have more to say about these strategies in our New Directions.

Efficiency

As we use the term in the context of data use in teaching, efficiency has to do with suiting the data system to the circumstances of practice. If the teacher is a teacher of English language learners and wants to track students' growth in reading, then—as one ELL teacher we interviewed told us—she cannot depend exclusively on the students' annual ELA state testing data (NY-SESLAT in New York State). She needs other data, too, that tracks development more finely and offers her and her students more timely signals of the students' growth. So, for example, this teacher maintains folders on all her students. They include running record data, evaluated writing assignments, evidence of growth in managing text complexity, and anecdotal memos that she writes following every individual session with a student. This teacher can produce and manage such a data set efficiently because she works with her students one-on-one and in small groups. However, the 7th-grade ELA

teacher whose class this ELL teacher “pushes into” (to support some number of her ELL students there) cannot do the same. She works with so many more students, and more often in large groups rather than one-on-one. Yet she can still make a point every week of tracking at least one literacy indicator for every ELL student in her room—via a Guided Reading session report, an exit ticket, an observation over the student’s shoulder, or a deliberately elicited oral comment—and she can record these data and share them with the ELL teacher.

We found that most of the practitioners we studied, even if they lacked precision in describing validity and efficiency, seemed aware of the need for both in data use in teaching. For example, one principal told us that “the state testing gives you something more standardized,” by which we think he meant *valid*, “but I think,” he added quickly, that “what is necessary is to look at the specific tools, and see how reliable and viable they are”—by which we think he meant *efficient*.

GENUINELY BIG DATA

We described, above, the difference between big-test data and intimate data, but we also acknowledged that genuinely big data may also come to play an increasingly important role in schooling (National Academy of Education, 2017). In fact, it is already playing a circumspect role. Even though big data transactions in any field are still technically challenging, there is a lot of talent available for dealing with the challenges—and no longer just in advertising, sports, manufacturing, and science.

Education Week’s Benjamin Herold reports, for example, on an ambitious start-up called AltSchool, funded by \$133 million of venture capital—mostly from Facebook founder Mark Zuckerberg. AltSchool employs data scientists and engineers to develop tools and other potential products for future schools based on data generated by students learning now at the company’s network of lab schools in the San Francisco Bay Area and New York City. The data is collected by multiple devices embedded in the schools’ environments. These measure students’ engagement levels, keystrokes, vocabulary use, eye movements, skill development, knowledge acquisition, and more (Herold, 2016a; Herold & Doran, 2016). In this case, big data is harvested and aggregated intimate data, often exceedingly intimate data, typically captured in online learning environments, and typically focused on learning processes.

However, as Andrew Ho (2017) points out, big data can also take the form of massive administrative data (for example, data related to eligibility for free or reduced-price lunch, gender, race, disability status, ELL status, attendance, and, of course, personal identifiers like names and addresses).

And he notes that variables in one kind of big data can be linked to variables in another by means of unique identifiers. This is, of course, one basis of the promise that researchers and policymakers (and—yes—marketers too) see in big data. And, as Ho points out, it is also the basis of concern about big data as a potential violator of family and student privacy.

Indeed, it may be that innovations like AltSchool will at some point encounter political opposition based on privacy concerns. This is what happened to the educational big-data management company InBloom. That effort was also well-funded—with \$100 million from the Bill and Melinda Gates Foundation and the Carnegie Corporation of New York. Its aspiration was to serve as a central data repository and analyst of student performance and other student data collected not only by districts and states but by vendors the districts and states work with. It aimed to encourage open-source tool development, to facilitate communication between vendors and districts, and to relieve districts and states of the enormous costs of managing and securing their own student data (Herold, 2014). However, within just over a year of its launch, InBloom disappeared under a storm of privacy concerns (Ho, 2017; National Academy of Education, 2017).

Meanwhile, it is important to note that privacy concerns are not the only threat to (and from) big-data projects, in education as elsewhere. The projects can also fail on efficiency grounds. Despite the validity they amass through sheer scale, big-data projects may nonetheless fail an efficiency test in dealing with what Atul Gawande (2009) calls the extreme complexity of practice at the ground level. He refers to medical practice, but we believe—and explain in the next chapter—that a similar level of complexity attends teaching practice too.

Science journalist Amy Standen (2014) tells the story of a big-data strategy involving the assemblage of a massive “electronic cohort” of pediatric lupus patients. Although the effort likely saved the life of a young lupus patient at Stanford’s Lucille Packard Children’s Hospital, as reported in the *New England Journal of Medicine* (Frankovich, Longhurst, & Sutherland, 2011), it is *not* routinely used today even at Stanford. Today, Standen reports, the hospital trusts in what it regards as the more efficient method of a small team of doctors conferring on the basis of their own relevant cases, their familiarity with ordinary experimental medical research, and their collaborative examination of the patient. There will likely come a day when they deal also with predictions and probabilities drawn from big data, though they will likely still include the other kinds of data in their deliberations—even perhaps giving precedence to them. In the end, efficiency depends on circumstances—what the situation demands at this moment, and what time and constraints on attention permit.

So it goes in educational contexts too, as educational researcher James Paul Gee argues in an interview with journalist Benjamin Herold. It is foolish, Gee says, to think that keystroke and chat window dialogue data

captured in online courses, or biometric data on posture or skin temperature captured in gaming environments, are inherently more valid in describing or boosting learning than is the ordinary data that comes from simply observing and interacting with students as they learn (Herold, 2016b). Essentially, he asks, who would give precedence to the former over the latter? We would say that much depends on the *use* of the data. Do you want to predict massive patterns of behavior? If so, consult genuinely big data. But if you are a teacher, you are more likely interested in knowing how to help Jorge learn how to read. And intimate data is what you need for that. Indeed, as Ho (2017) suggests, genuinely big data may put you in exactly the wrong mindset—for example, a tacit presumption that Jorge is demographically a long-shot in terms of learning to read well.

SUMMING UP

More than 15 years ago, when Milbrey McLaughlin and Joan Talbert were conducting the voluminous research that informs their groundbreaking book, *Building School-Based Teacher Learning Communities* (2006), one teacher asked them, what *is* data? A teacher would not likely ask this question now. That's not just because the word *data* is ubiquitous generally now (for example, in the teacher's monthly personal media bills), but also because it has become ubiquitous in teaching. On the other hand, this chapter was designed with the idea in mind that the question is still very useful, and a reluctance to ask it is dangerous. Teachers cannot engage in thoughtful data use in teaching without knowing some of the distinctions that this chapter makes—as related, for example, to big-test data, intimate data, and genuinely big data; or to validity and efficiency in data use. The chapter also deals with the important question of what teachers talk about when they talk about data, and implicitly about what they *might* talk about. What follows is the first of the book's New Directions, which portrays a context for such talk.