Big Data Comes to School: 
Reconceptualizing Evidence and Research in the Era of Technology-mediated Learning

Bill Cope, University of Illinois
Mary Kalantzis, University of Illinois

Introduction: The Challenge of Big Data

Over the past century, the learning sciences have come to be defined by an array of experimental, survey and anthropological methods. Recent literature on ‘big data’, however, suggests new possibilities educational evidence and research in the era of digitally-mediated learning and the large amounts of data collected incidental to that learning (DiCerbo and Behrens 2014; West 2012). These, in turn, are aspects of the broader social phenomenon of big data (Mayer-Schönberger and Cukier 2013; Podesta, Pritzker, Moniz, Holdern, and Zients 2014).

A measure of the newness of this phenomenon is AERA’s Handbook of Complementary Methods in Education Research (Green, Camilli, Elmore, and Elmore 2006), outlining the panoply of quantitative and qualitative methods used across the learning sciences. It is no criticism of the editors to note that this volume does not have a chapter on data science or learning analytics. The absence is simply a reflection on how much has changed in a few short years. In this this paper, we set out to explore the consequences of big data for educational research practice. We argue that a new generation of ‘educational data sciences’ might require a reconceptualization of the nature of evidence and the dimensions of our research practices.

The Rise of ‘Big Data’ and Data Science

Big data is big in the news. The Large Hadron Collider at the CERN labs in Switzerland can collect a petabyte of data per second, and in this river of data it has been possible to discern the Higgs boson, previously only predicted in theory by the Standard Model of physics. The Large Synoptic Survey Telescope now under construction will collect thirty terabytes of data per night, covering the whole sky in four days. Computers comparing patterns in these digital images will be able to identify transient phenomena in the sky that would not have been visible without this digitally-mediated looking. Sequencing the six billion base pairs of the human genome, which a decade ago took years and cost billions of dollars, can now be done in hours and for a mere thousand dollars. This achievement recedes into insignificance when one considers the task ahead to research the microbes living in the human gut. Together, these microbes have one hundred times more genetic indicators than their host; the microbic mix can be as much as fifty percent different from one individual to the next, and their composition is changing all the time. When these calculations can be made, there may be answers to questions about the effectiveness of antibiotics (Shaw 2014).

1 Article in review.
These are some of the ways in which we are coming to see previously invisible phenomena of our world through the medium of digitized data and information analytics. An emerging ‘big data’ literature goes so far as to claim that this represents a new phase in the development of human knowledge processes, changing our arguments about causality (Mayer-Schönberger and Cukier 2013), and even evolving into a new paradigm of science (Hey, Tansley, and Tolle 2009).

In this paper, we want to engage with this literature by exploring two questions. The first is: what are the continuities and differences between data sciences that address the natural and the social world? In answering this question, we will explore challenges and opportunities for the learning sciences. And our second is: does data science change the nature of evidence, and specifically evidence of learning? To which, our answer will be in equal measure ‘no’ and ‘yes’. The ‘no’ response is that the evidence we are seeking is the same—about the processes and outcomes of human learning. The ‘yes’ response is that in the era of digitally-mediated and incidentally recorded learning, data sciences may render anachronistic (expensive, inefficient, inaccurate, often irrelevant) many of our traditional research methods for gathering that evidence.

**Informationalizing the Social**

A large number of our social interactions today are ‘informationalized’—our emails, texts, tweets, Facebook posts, web navigation paths, and web purchases, all time-stamped and often also geo-located. By ‘informationalized’ we mean, our social interactions are created and transmitted through digital information platforms, which (and this is the decisive factor from the point of view of data science) incidentally record these interactions. Recording is easy and cheap. It happens in centralized or tightly distributed server farms, so the data can readily be stitched together for analysis.

Even though the storage of this data is principally of short term value to users, its value to hosts centers on its ‘informationalization’ over a longer timeframe. This is why commercial platform providers will often let us use their digital platforms for free. The informationalized data, recorded within frameworks that are designed for analysis, are valuable to them. We users care for the recording less than they do, at least in the longer term; in fact we mostly don’t need the recordings beyond the moment of communicative interchange. But they can use the data to serve advertising, do market research, or make recommendations that may draw us deeper into their platform or sell us a product.

The scale of social and behavioral data collection is enormous. Facebook’s data grows by 500 terabytes per day, including 2.7 billion ‘likes’. Wal-Mart handles 1 million customer transactions per day. Google processes 20 petabytes of data per day. 250 billion email messages are sent every day. From the point of the social sciences, the ‘big’ part of big data is less relevant than the differences between this data and traditional sources of evidence. The data is comprehensive (every Facebook user, every Wal-Mart customer). It is often complex and noisy, only small parts of which may render useful information (Lazer, Pentland, Adamic, Aral, Barabási, Brewer, Christakis, Contractor, Fowler, Gutmann, Jebra, King, Macy, Roy, and Alstyne 2009). And in the case of social and medical data, ethical issues of human agency and consent arise, issues which don’t present themselves when looking for elementary particles or galaxies—so social sciences such as education can only learn some big data lessons from their natural science peers.
As we hope to demonstrate, the learning sciences are a particularly rich area for the exploration of the nature of social evidence in the era of informationalization. Although the epistemological issues emerging apply to all social sciences, we believe education is especially significant for a number of reasons. The problems we face are as tricky as they could possibly be—so much so that if we can address education’s data science challenges, we will be addressing challenges also found in other social sites.

Education is significant too, because the stakes are so high. Pervasive access to technology-mediated learning environments, coupled with emerging data sciences, could transform the discursive and epistemic architecture of the modern classroom. The sources of this institutional form might be tracked back to the rise of mass-institutionalized education in the nineteenth century, with its textbooks, teacher-talk and tests. Or its origins might be traced back further to the Rule of St Benedict, who, in founding the institution of western monasticism, established a knowledge-transmission pattern of communicative and pedagogical relations between the teacher and the taught (Kalantzis and Cope 2012). Formal education is an old institution. Its roots are deep in our culture. But regime change is possible.

Education is also important because access and representation to meaning—the process of encountering things-to-be-known and representing what one has come-to-know—is increasingly mediated by networked, digital information and communications systems. Importantly today, and to emphasize the point again, these systems can and often do incidentally record everything that is happening.

And lastly, as we have argued elsewhere, education is significant because, it is uniquely a domain of meta-knowledge, a site where we deal with the fundamental dynamics of knowledge, its ontogenesis in persons and its phylogenesis in knowledge communities (Kalantzis and Cope 2014). How does informationalization affect these processes? There can be no harder and all-encompassing question than the question of human coming-to-know.

So, if education has been slow to address what is popularly called ‘big data’, if we are behind the natural sciences and the other social sciences in this regard, there are good reasons why we might now aspire to take a lead.

Big Data in Education
The recent realization of the significance of data science is not because technology-mediated learning has suddenly arrived on the educational scene. The project of using computers-in-education is now five decades old, beginning, historians of computing contend, in 1959 with the development of the PLATO learning system at the University of Illinois. Even before then, in a 1954 article published in the Harvard Educational Review, and reprinted in a book with the future-aspirational title, Teaching Machines and Programmed Learning, B.F. Skinner was foreshadowing the application of ‘special techniques ... designed to arrange what are called “contingencies of reinforcement” ... either mechanically or electrically. An inexpensive device,’ Skinner announced ‘... has already been constructed’ (Skinner 1954 (1960): 99,109-110). The book has a future-inspiring photo of such a machine.

If technology-mediated learning is by no means new, two developments of the past half-decade stand out: deep network integration of digital learning environments through ‘cloud computing’, and the generation of ‘big data’ that can be connected and analyzed
across different systems. The effects of each of these developments is destined to intensify over the next few years.

The significance of ‘cloud computing’ (Erl, Puttini, and Mahmood 2013) is social more than it is technological. We characterize this as a shift from personal computing to interpersonal computing. From the 1980s, personal computing provided mass, domestic and workplace access to small, relatively inexpensive computers. From the 1990s, the internet connected these for the purposes of communications and information access. Cloud computing moves storage and data processing off the personal computing device and into networked server farms. In the era of personal computing, data was effectively lost to anything other than individual access in a messy, ad hoc cacophony of files, folders, and downloaded emails. In the era of interpersonal computing, the social relations of information and communication can be systematically and consistently ordered.

This opens out the social phenomenon that is popularly characterized as ‘Web 2.0’ (O'Reilly 2005). It also opens out massively integrated social media. This turns data that was before this socially inscrutable, to socially scrubtable data. By interacting with friends using social media such as Facebook or Twitter, one is entering these providers’ data model, thereby making an unpaid contribution to that provider’s massive and highly valuable, social intelligence. By storing your data in webmail or web word processors, Google can know things about you that were impossible to know when you had your files on a personal computer and downloaded your emails, and this ‘social knowing’ has made it into a fabulously valuable advertising business.

So, our question is how could the social knowing that is possible in the era of interpersonal computing be applied to education? More and more learning happens in the cloud, not in separately installed programs or work files on personal computing devices. In education this includes: delivery of content through learning management systems; discussions in web forums and social media activity streams; web writing spaces and work portfolios; affect and behavior monitoring systems; games and simulations; formative and summative assessments; and student information systems that include a wide variety of data, from demographics to grades.

The idea of ‘big data’ captures the possibility of making better sense of the learning of individuals and groups because it is now better connected within the framework of interpersonal computing. It also represents a challenge. Though there is a mass of data, it is untidy, inconsistent and hard to read. We are accumulating a mountain of potential evidence of learning that may be of great value to teachers, learners and educational researchers.

**Disciplinary Realignments**

The emergence of big data in education has been accompanied by some significant disciplinary realignments. Leaders in the emerging field of learning analytics speak clearly to what they consider to be a paradigm change. Bienkowski and colleagues point out that “educational data mining and learning analytics have the potential to make visible data that have heretofore gone unseen, unnoticed, and therefore unactionable” (Bienkowski, Feng, and Means 2012). West directs our attention to “‘real-time’ assessment [with its] ... potential for improved research, evaluation, and accountability through data mining, data analytics, and web dashboards (West 2012). Behrens and
DiCerbo argue that “technology allows us to expand our thinking about evidence. Digital systems allow us to capture stream or trace data from students’ interactions. This data has the potential to provide insight into the processes that students use to arrive at the final product (traditionally the only graded portion). ... As the activities, and contexts of our activities, become increasingly digital, the need for separate assessment activities should be brought increasingly into question” (Behrens and DiCerbo 2013). Chung traces the consequences for education in these terms: “Technology-based tasks can be instrumented to record fine-grained observations about what students do in the task as well as capture the context surrounding the behavior. Advances in how such data are conceptualized, in storing and accessing large amounts of data (‘big data’), and in the availability of analysis techniques that provide the capability to discover patterns from big data are spurring innovative uses for assessment and instructional purposes. One significant implication of the higher resolving power of technology-based measurement is its use to improve learning via individualized instruction” (Chung 2013). DiCerbo and Behrens conclude: “We believe the ability to capture data from everyday formal and informal learning activity should fundamentally change how we think about education. Technology now allows us to capture fine-grained data about what individuals do as they interact with their environments, producing an ‘ocean’ of data that, if used correctly, can give us a new view of how learners progress in acquiring knowledge, skills, and attributes” (DiCerbo and Behrens 2014).

In the past few years, two new fields have emerged, each with their own conference and journal: ‘educational data mining’ and ‘learning analytics’. The principal focus of educational data mining is to determine patterns in large and noisy datasets, such as incidentally recorded data (e.g. log files, keystrokes), unstructured data (e.g., text files, discussion threads), and complex and varied, but complementary data sources (e.g., different environments, technologies and data models) (Baker and Siemens 2014; Castro, Vellido, Nebot, and Mugica 2007; Siemens and Baker 2013). Although there is considerable overlap between the fields, the emphasis of learning analytics is to interpret data in environments where analytics have been ‘designed-in’, such as intelligent tutors, adaptive quizzes/assessments, peer review and other data collection points that explicitly measure learning (Bienkowski, Feng, and Means 2012; Knight, Shum, and Littleton 2013; Mislevy, Behrens, Dicerbo, and Levy 2012; Siemens and Baker 2013; West 2012).

These new fields build upon older sub-disciplines of education. They also transform them by using quite different kinds of evidentiary argument from those of the past. Foundational areas for big data and data science in education include: formative and situated or classroom assessment (Black and Wiliam 1998; Pellegrino, Chudowsky, and Glaser 2001); technology-mediated psychometrics (Chang 2014; Rupp, Nugent, and Nelson 2012); self-regulated learning and metacognition (Bielaczy, Pirolli, and Brown 1995; Bransford, Brown, and Cocking 2000; Kay 2001; Kay, Kleitman, and Azviedo 2013; Lane 2012; Magnifico, Olmanson, and Cope 2013; Shepard 2008; Winne and Baker 2013); and analyses of complex performance and holistic disciplinary practice (Greeno 1998; Mislevy, Steinberg, Breyer, Almond, and Johnson 2002).

Old Research Questions and New

Big data can help us gather new evidence as we continue to do research into education’s ‘wicked problems’: performance gaps which produce cycles of underachievement;
cultural-racial differences in educational experience and outcome; or the role of education in reproducing or breaking cycles of poverty.

However, the very socio-technical conditions that have made big data possible, are sites for new educational practices that themselves urgently require research. Data science is uniquely positioned to examine these transformations. Its sources of evidence are intrinsic to these new spaces and its innovative methods of analysis essential.

One such transformation is the rise of blended and ubiquitous learning (Cope and Kalantzis 2009), and the blurring of pedagogies for in-person and remote learning interactions, such as the ‘flipped classroom’ (Bishop and Verleger 2013). However, after half a century of application in traditional educational sites, the overall beneficial effects of computer-mediated learning remain essentially unproven. In his examination of 76 meta-analyses of the effects of computer-assisted instruction, encompassing 4,498 studies and involving 4 million students, John Hattie concludes that “there is no necessary relation between having computers, using computers and learning outcomes”. Nor are there changes over time in overall effect sizes, notwithstanding the increasing sophistication of computer technologies (Hattie 2009: 220-1). Warschauer and Matuchniak similarly conclude that technology use in school has not been proven to improve student outcomes, though different kinds of pedagogical applications of technology do (Warschauer and Matuchniak 2010). More recently, in a review of technology integration of schools, Davies and West conclude that although “students ... use technology to gather, organize, analyze, and report information, ... this has not dramatically improved student performance on standardized tests” (Davies and West 2014). If traditional research methods and sources of evidence have only offered disquieting ‘not proven’ verdicts about technology use in general, the big data analyses that technology-mediated learning environments make possible may allow us to dig deep into the specifics of what within educational technologies works, and what does not. Such research may furnish important insights where gross effect studies have not.

Specific developments in new fields of educational innovation also offer both challenges and opportunities for educational data science. We have seen over the past decade the rapid growth of purely online or virtual schools, from the K-12 level to higher education (Molnar, Rice, Huerta, Shafer, Barbour, Miron, Gulosino, and Horvitz 2014). We have also witnessed the emergence of massive, open, free educational offerings, such as MOOCs (DeBoer, Ho, Stump, and Breslow 2014; Peters and Britez 2008). These developments are widely regarded to be transformative, or potentially transformative. Innovative modes of research are required before we can conclude about their effects.

In both online and face-to-face contexts, we have also seen the introduction of interactive digital resources including games and simulations; classroom interactions via discussion feeds and forums that elicit more consistent and visible participation; recursive feedback systems which extend and in some cases transform traditional modes of formative and summative assessment (Cope, Kalantzis, McCarrthy, Vojak, and Kline 2011; DiCerbo and Behrens 2014; Mislevy, Almond, and Lukas 2004; Quellmalz and Pellegrino 2009); and adaptive, personalized or differentiated instruction which calibrates learning to individual needs (Conati and Kardan 2013; Shute and Zapata-Rivera 2012; Walkington 2013; Wolf 2010). Such models and processes of instructional delivery are variously labeled ‘constructivist’, ‘connectivist’ or ‘reflexive’ (Kalantzis and Cope 2012; Siemens 2005). Such innovative pedagogical frameworks for learning are posited to be
peculiar to, or at least facilitated by, technology-mediated learning. Big data analyses will help us determine which pedagogies have what effects and how they have these effects.

As these platforms and environments all generate large amounts of data, what follows are expectations that teachers become data literate (Twidale, Blake, and Gant 2013), in support of processes of evidence-based decision making (Mandinach and Gummer 2013). We need to research how teachers become most effective in these big data environments.

In the next sections of this paper, will attempt to answer two large questions—what kinds of data are these? and how do the methods used to analyze these data challenge our traditional research practices?

### What are Big Data?

**What’s Different about ‘Big Data’**

So big data are here, and more are coming to education. But what is different about these kinds of data? To start to answer this question, a definition: in education, ‘big data’ are:

a) the purposeful or incidental recording of interactions in digitally-mediated, cloud-interconnected learning environments; b) the large, varied, immediately available and persistent datasets generated; c) the analysis and presentation of the data generated for the purposes of learner and teacher feedback, institutional accountability, educational software design, learning resource development, and educational research.

It’s not the bigness that makes big data interestingly different from data generally. Libraries have always managed a lot of data. Personal computers and standalone mainframes have long managed a lot of data. Beyond ‘big’ per se, these are some differences:

- **The datapoints are smaller.** This is in many ways more significant than the bigness. In fact, it is the only (both incidental and consequential) reason why the data have become bigger. Small might mean an answer to a question, a move in a simulation, or a comment in a thread in a discussion board. Smaller still might be a keystroke, a timestamp, a click in a navigation path, or a change captured in the edit history of a wiki or blog. Learning has not become bigger. It’s just that the things we can record incidental to the learning process have become smaller, and these add up to a lot more data than we have ever had before—more data than a human can deal with, without computer-synthesized analytics.

- **Much of the data generated and recorded are unstructured.** From a computing point of view, log file records (for instance keystrokes, timestamps, clicks, URLs) and records of natural language are unstructured data. These data are not self-describing. If computers are to assist in reading these masses of unstructured data, their meanings need to be inferred from pattern analysis. In unstructured data, there will invariably be a lot of ‘noise’ surrounding every potentially meaningful signal.

- **When the data are structured, different data sources are structured in different ways.** Item-based, standardized tests, for example, generate structured data. These data are self-describing: the answer to this question is right or wrong; the percentage of right versus wrong answers for each examinee speaks for itself; an examinee is located in a cohort distribution of relative test success and failure. But
how do we find patterns relating to questions about the same or similar things in
different tests? How do we compare students or cohorts when they have done
different tests? The data might be structured, but when each dataset is an island,
and when the data models are different, we can’t answer some critically important
educational questions. But now that all these data are kept in potentially
accessible cloud architectures, we should be able to align data even when the data
models diverge—but new data science methods are needed to do this.

- **Evidence of learning is embedded in learning.** In the traditional model of
evidence-gathering and interpretation in education, researchers are independent
observers, who pre-emptively create instruments of measurement, and insert these
into the educational process in specialized times and places (a pre-test or post-test,
a survey, an interview, a focus group). The ‘big data’ approach is to collect data
through practice-integrated research. If a record is kept of everything that happens,
then it is possible analyze what happened, _ex post facto_. Data collection is
embedded. It is on-the-fly and ever-present. With the relevant analysis and
presentation software, the data is readable in the form of data reports, analytics
dashboards and visualizations.

- **The evidence is more importantly social than it is technical.** Big data is where
artificial intelligence meets ‘crowdsourced’ (Surowiecki 2004) human
intelligence. Millions of tiny human events are recorded at datapoints that can be
aggregated. Take natural language: one student recommends a wording change in
the context of a whole text and codes the reason; another accepts the change. In
this scenario, a ‘machine learning’ algorithm has collected one tiny but highly
meaningful piece of evidence. It happens again, with one and then many other
students. The more it happens, the more powerful the evidence becomes.
Aggregated to the thousands or millions, these data provide crucial evidence for
teachers, educational program designers or researchers. We can also use these
data to make proactive writing recommendations to future learners, or formative
assessment (Cope et al. 2011).

- **Data and intervention are not separate.** The agenda of big data is not simply to
observe the world. It is to change the world. With the right analytics and
presentation tools, its very presence changes the world. The interventions might
be small—a tiny piece of feedback for a learner, or creating a decision point about
the next step in learning for an individual student in a personalized learning
environment. Or it may be for the purposes of learner profiling when teacher and
student are working on the co-design of individual learning programs. Or it may
be to generate data for the purposes of administrative accountability. Interventions
for and based on data are by their nature designed to change things. For better or
worse, big data and social engineering go hand in hand. In education, as a
consequence, we need protocols to assure users that data pervasively collected
incidental to their everyday learning activities, will be used for their benefit, and
not their detriment—for instance, to discriminate through profiling. This is a
major concern of a White House investigation of big data (Podesta et al. 2014).

- **Everyone is a data analyst now.** The same embedded data that a researcher can
use, a teacher can use too. In fact, they should use it, to know their learners and
recalibrate their teaching. In this evidentiary environment, the teacher can be,
should be, positioned as researcher. And this same data can be presented to students, both in recursive feedback or formative assessment systems, and progress overviews. Then the student too, will be positioned as a researcher of sorts—of their own learning. With big data, traditional researcher/practitioner and observer/subject positions are blurred. This is not a feature of the data per se, but highlights a dimension of accessibility that to some degree also determines the shape, form and purpose of the data.

Big data has come to school. Or at least if the hardware has not yet been delivered, it is on its way. However, we have a long way to go to create software and research processes adequate to its promise. On which subject, we need to make a declaration of interest. Since 2009, with the support of a series of research and development grants from the Institute of Education Sciences\(^2\) and the Bill and Melinda Gates Foundation, we have built a big data environment called Scholar, capable of collecting evidence and serving analytics data from perhaps a million semantically legible datapoints for a single student in their middle or high school experience; or in a class in a term; or a school in a week. It’s been a big learning journey, and we’ve barely begun, but that’s another story (Cope and Kalantzis 2013).

**Data Sources and Data Types**

We’ve briefly mentioned sources of big data already, but now we want to be more analytical, to build a typology of educational datapoints in technology-mediated learning environments. We’ve addressed scale and persistence as key features of big data already. Now we want to focus on its range and diversity.

Just because the data is born digital does not mean it can be readily compiled into meaningful, integrated views. Far from it, the data comes in surprisingly different shapes and forms, as different in fact as the educational software environments in which it is generated. Dealing with the bigness and the storage of this recorded data is easy compared to the issue of its data diversity. This is perhaps the most fundamental challenge for the emerging educational data sciences, and one to which we will return in the last section of this paper.

The following typology is designed to capture the principal educational data types. However, in the nature of typologies, the realities are more complex. Different data types overlap—the one data collection environment often contains a variety of data types. Certainly the experience of a whole day in a school with one-to-one computer access will draw learners past quite a few of these data types. We have ordered the following list of ten data types based on functional juxtapositions, rough similarity and in some cases, a degree of overlap.

---

\(^2\) US Department of Education Institute of Education Sciences: 'The Assess-as-You-Go Writing Assistant: a student work environment that brings together formative and summative assessment' (R305A090394); 'Assessing Complex Performance: A Postdoctoral Training Program Researching Students’ Writing and Assessment in Digital Workspaces' (R305B110008); ‘u-Learn.net: An Anywhere/Anytime Formative Assessment and Learning Feedback Environment' (ED-IES-10-C-0018); 'The Learning Element: A Lesson Planning and Curriculum Documentation Tool for Teachers' (ED-IES-10-C-0021); and 'InfoWriter: A Student Feedback and Formative Assessment Environment for Writing Information and Explanatory Texts' (ED-IES-13-C-0039). Scholar is located at http://CGScholar.com
1. Analytics Tools in Learning Management Systems

Here, we base our analysis on an investigation of data analytics tools, either available or in development in some of the most widely used learning management systems: Blackboard, Desire2Learn, Canvas and Coursera. These environments often fold in other data types such as selected response tests that we deal with as our next data type. We will focus here on data that is characteristic of these environments as learning management systems. In the collection of structured data, these systems can and often do include student demographic data, courses taken, media accessed, discussion areas joined, assignments given, files uploaded, and grades assigned. In the collection of unstructured data, these systems can collect login timestamps, keystroke data, and clickstream data. These unstructured data are of no particular significance until they are synthesized and presented as indicators of levels of student engagement, records of progress and predictors of performance. Learning analytics dashboards present their readings of this structured and unstructured data on a per student, per course and whole-institution implementation levels.

![Fig.1: Student Risk Quadrant in D2L ‘Insights’ Product](image)
2. Adaptive Survey-Psychometric Measurement Systems

Frequently located within or alongside learning management systems, selected response assessments rely upon long-established traditions of survey psychometrics. The general ‘assessment argument’ (Pellegrino, Chudowsky, and Glaser 2001) underlying survey psychometrics requires: an observational opportunity by requiring examinees to respond to a series of test items that validly samples a curriculum; an interpretation process which makes sense of individual and cohort scores; and inferences about cognition based on these interpretations. This is the process used by latent-variable psychometric models, and various accompanying statistical techniques. Computer technologies have revolutionized pencil-and-paper ‘bubble tests’. Computer Adaptive Tests (CAT) tailor the test to the trait level of the person taking the test. Computer Diagnostic Tests differentiate subset areas of knowledge within a test (Chang 2014; Chang 2012). It is possible to embed such tests within instruction, offering immediate answers to students, and so to move away from the “Teach/Stop/Test” routine of conventional, summative assessment (Woolf 2010). Such is the project of companies such as Knewton, who are embedding adaptive tests into textbooks (Waters 2014). Or, the other way around, questions and responses to questions can be used to direct you to places in a textbook (Chaudhri, Cheng, Overholtzer, Roschelle, Spaulding, Clark, Greaves, and Gunning 2013). Advanced applications of these technologies include machine learning environments in which difficulty ranking of selected response items is crowdsourced based on patterns of response to particular questions in relation to student profiles (Segal, Katzir, Gal, Shani, and Shapira 2014). In the classroom, the teacher can solicit and record on-the-fly selected response feedback through ‘clickers’ (Blasco-Arcas, Buil, Hernández-Ortega, and Sese 2013).

3. Learning Games

Testing has a game-like quality—playing the game of school, in a competitive environment, and succeeding (and also ‘losing’) in the competitive environment where grades are spread across a normalized distribution curve. However, when educators advocate ‘gamification’ (Magnifico, Olmanson, and Cope 2013), they mean ‘game’ in a narrower sense, principally variants of the genre ‘video game’ (Gee 2005). This genre was first developed in the PLATO computer learning system at the University of Illinois in the 1960s—a salutary story, that computer games were first created in an e-learning system, now that we are trying to bring them back into educational settings where they began. Video games were massively popularized with the rise of personal computing in the 1980s, and now reaching an audience larger than Hollywood. The data types generated are also fundamentally different from tests. James Gee identifies 36 learning principles that are intrinsic to most video games, even first person shooter games, many of which are absent or at best weakly present in what he characterizes as ‘traditional schooling’ (Gee 2004). Adding our own gloss, each one of these principles reflects a moment of machine-mediated reflexive data collection; each is a recordable teaching moment. We will highlight just a few—Principle 11: intrinsic rewards are reflected in staged achievement levels; Principle 16: there are multiple ways to make move forward; Principle 21: thinking and problem solving are stored in object-representations; Principle 24: learning situations are ordered such that earlier generalizations are incrementally fruitful for later, more complex cases; Principle 28: learners can experiment and discover...
Every move leaves a recorded trace of thinking, a trace of deep significance. These data are massive and complex, in the case of one user, let alone many. Our challenge is to use these data which are intrinsic to the recursivity (and fun!) of games, not just to drive the logic of the game, but also as sources of evidence of learning. Such data might also be used to support the creation of better educational games (Mislevy, Oranje, Bauer, Davier, Hao, Corrigan, Hoffman, DiCerbo, and John 2014).

4. Simulations
Simulations share a close family resemblance with games, but for our data typology, they are also importantly different. Whereas games are fictional spaces with predetermined rules, a player or players in highly structured roles, strict and competitive scoring structures and a predetermined range of outcomes, simulations model the empirical world, with few rules beyond the coherences, little or no competitive scoring and open-ended outcomes (Sauvé, Renaud, Kaufman, and Marquis 2007). Learners might work their way through simulations, in which models are presented in partial scaffolds, but there is latitude for each participant to explore alternatives and do their own modeling (Blumschein, Hung, Jonassen, and Strobel 2009). The key feature of simulations is their direct reference to the empirical world, either presenting empirical data or eliciting new data. The distinctive reference points for learning in this data type are empirical evidence, navigation paths taken, and the models created in the play between simulation and user. Simulations afford possibilities for assessment that transcend traditional tests (Clarke-Midura and Dede 2010).

5. Intelligent Tutors
Intelligent tutoring systems guide a learner through a body of knowledge, serving content, requesting responses, making hints, offering feedback on these responses, and designing stepwise progression through a domain depending on the nature of these responses (Aleven, Beal, and Graesser 2013; Chaudhri, Gunning, Lane, and Roschelle 2013). A wrong move in solving a problem might produce an opportunity for further revision; a correct solution might mean that a learner can progress onto a more complex problem or a new topic. In this way, a recursive data collection process is built into the tutor. This is the basis for the learner-adaptive flexibility and personalized learning progressions offered by such systems (Koedinger, Brunskill, Baker, and McLaughlin 2013; Woolf 2010); but the data generated is rarely put to use beyond these immediate processes for individual students. Intelligent tutors work best in problem domains where highly structured progressions are possible, such as mathematics. They are less applicable in areas where progression cannot readily be assembled into a linear sequence of knowledge components, such as writing.

6. Natural Language Processors
Technologies of automated writing assessment have been shown to be able to grade essays to a degree of reliability that is equivalent to trained human raters (Burstein and Chodorow 2003; Chung and Baker 2003; Cotos and Pendar 2007). Thus far, these technologies have been less successful in providing meaningful feedback to writers beyond the mechanics of spelling in grammar (McNamara, Graesser, McCarthy, and Cai 2014; Vojak, Kline, Cope, McCarthey, and Kalantzis 2011; Warschauer and Grimes...
So-called natural language processing technologies offer two quite different mechanisms of analysis. One is rule-based analytics of which grammar and spell checkers are canonical examples. The second is the application of and statistical methods of corpus comparison for the purposes of grading (for instance, a new essay with this grade is statistically similar to another essay that a human grader has rated at a certain level) and latent semantic analysis (for instance, where the meanings of homonyms disambiguated or synonyms are aligned) (McNamara, Graesser, McCarthy, and Cai 2014; Vojak et al. 2011). Promising areas of development include analyses of conceptual structures or topic models in written texts (Paul and Girju 2010), argument (Ascaniis 2012), and sentiment analysis, in discussion forums for instance (Wen, Yang, and Rose 2014).

7. Semantic Mapping
A concept or information map is a spatial array that represents the component parts of knowledge (facts, concepts) as nodes, connecting these via directional links that specify the relation between nodes (Novak and Cañas 2008; Tergan 2005). Mind or concept mapping or advanced organizers were introduced from educational psychology in the 1960s, with the aim of aiding the cognitive process of “subsumption,” in which new ideas reorganize existing schema (Ausubel 1963; Ausubel 1978; Ausubel, Novak, and Hanesian 1978). Numerous educational technology tools have been developed that employ concept mapping, from hypertext stacks, when concept mapping was first introduced to computer-mediated learning, to more recent e-learning software systems that support ‘mind mapping’ (Bredeweg, Liem, Beek, Linnebank, Gracia, Lozano, Wißner, Bühlting, Salles, Noble, Zitek, Borissova, and Mioduser 2013; Cañas, Hill, Carff, Suri, Lott, Gómez, Eskridge, Arroyo, and Carvajal 2004; Chang, Sung, and Lee 2003; Kao, Chen, and Sun 2010; Liu 2002; Su and Wang 2010; Tzeng 2005). The authors of this paper have developed one such tool, InfoWriter, in which learners during their writing highlight and diagram the ideas that the writing represents (Olmanson, Kennett, McCarthey, Searsmith, Cope, and Kalantzis 2014 (in review)). It is also possible to machine-generate semantic maps from text using natural language processing methods (Girju, Badulescu, and Moldovan 2006).

8. Social Interaction Analyses
Online learning spaces frequently support various forms of peer interaction. One such form of interaction is online discussion forums. These present as unstructured data. Patterns of peer interaction can be mapped—who is participating, with whom, to what extent (Speck, Gualtieri, Naik, Nguyen, Cheung, Alexander, and Fenske 2014; Wise, Zhao, and Hausknecht 2013). Natural language processing methods can be used to parse the content of interactions (Xu, Murray, Woolf, and Smith 2013). Online learning environments can also support computer-supported collaborative learning that aligns with what is frequently labeled as twenty-first century knowledge work, which characteristically is distributed, multidisciplinary and team-based (Liddo, Shum, Quinto, Bachler, and Cannavacciulo 2011; Strijbos 2011). Learners may be provided with different information and conceptual tools, or bring different perspectives to solve a problem collectively that none could solve alone. Such environments generate a huge
amount of data, the product of which is a collectively created artifact or solution that cannot be ascribed to individual cognition (Bull and Vatrapu 2011; Perera, Kay, Koprinska, Yacef, and Zaiane 2009). The processes of collecting and analyzing such data have been termed ‘social learning analytics’ (Ferguson and Shum 2012).

9. Affect Meters
Motivation and affect are key factors in learning (Magnifico, Olmanson, and Cope 2013). Computer-mediated learning environments can monitor student sentiments with affect meters of one kind or another: emote-aloud meters and self-reports on affective states that address a range of feelings, including, for instance, boredom, confusion, interest, delight, fear, anxiety, satisfaction, frustration (Baker, D’Mello, Rodrigo, and Graesser 2010; Chung 2013; D’Mello 2013; Wixon, Arroyo, Muldner, Burleson, Rai, and Woolf 2014). Log files can also provide indirect evidence of patterns of engagement, or more specific information such as the extent to which a learner relies on help offered within the environment (Rebolledo-Mendez, Boulay, Luckin, and Benitez-Guerrero 2013). In the context of social learning, web reputation technologies (Farmer and Glass 2010) can be applied to a spectrum of behaviors, ranging from helpfulness meters, offering feedback on feedback, to flagging inappropriate comments and potential cyberbullying (Espelage, Holt, and Henkel 2003).

10. Body Sensors
Body sensors are also used to measure affect, as well as patterns of engagement in e-learning environments. Connected to screen work, these may include eye tracking, body posture, facial features, and mutual gaze (D’Mello, Lehman, Sullins, Daigle, Combs, Vogt, Perkins, and Graesser 2010; Grafsgaard, Wiggins, Boyer, Wiebe, and Lester 2014; Schneider and Pea 2014; Vatrapu, Reimann, Bull, and Johnson 2013). Student movement not connected with screen presentations include wearable technologies such as bracelets (Woolf 2010: 19), RFID chips in student ID cards (Kravets 2012), group interactions in multi-tabletop environments (Martinez-Maldonado, Yacef, and Kay 2013), the ‘internet of things’ and the quantified self-carried in phones and watches (Swan 2012), and detectors that capture patterns of bodily movement, gesture and person-to-person interaction (Lindgren and Johnson-Glenberg 2013).

Implications for Educational Research

To say these education data are big, is an understatement. However, the object of our research interest—learning—is no more complex a human practice than it ever was. It’s just that the grain size of recordable and analyzable data has become smaller. Whereas it was never practicable to record every pen stroke made by every learner, every keystroke of every learner is incidentally recorded in log files and thus are potentially open to analysis. Small bodily movements can be recorded. Every word of student online interaction and work can be recorded and analyzed. The consequent data can now be analyzed on-the-fly, as well as over long periods of time when the data is persistent, recorded by default even when not by design. Already, some studies have dealt with
datasets consisting of as many as one hundred million datapoints (Monroy, Rangel, and Whitaker 2013).

Even more challenging than the bigness, we have just seen how varied are the sources of evidence. Not only does each represent a different technology cluster; it also reflects a different perspective on the social relations of knowledge and learning. How do you bring these data together to form an overall view of an individual learner, or a cohort, or a demographically defined group, or a teacher, or a school, or a kind of intervention, or a type of educational software?

These developments challenge educational researchers to extend and supplement their methods for eliciting evidence of learning. Following are ten methodological extensions of, or supplements to, traditional educational research methods when big data has come to school.

1. Multi-scalar Data Collection
Evidence of learning has always been gleaned in a series of feedback loops at different scales, from the microdynamics of instruction, to assessment, to research. In the microdynamics of instruction, learners get feedback, for instance in the patterns of classroom discourse where a teacher initiates a question, student responds with an answer, teacher evaluates the response (Cazden 2001). Assessment is feedback on another scale (Mislevy 2013)—the test at the end of the week or the term that determines how much of a topic has been learned. Research provides feedback on the effectiveness of the learning process overall, such as the effectiveness of an intervention in improving outcomes for a cohort of learners. Reading the evidence at each of these scales has traditionally involved different data collection instruments and routines, implemented in specialized times and places, and often by different people in different roles. Educational researchers, for instance, have traditionally collected their evidence of learning using independent, stand-alone and external observational protocols or measurement artifacts. Embedded data collection, however allows simultaneous collection of data that can be used for different purposes at different scales. Data collection timeframes, sites and roles are conflated. Traditional educational distinctions of evidentiary scale are blurred. We highlight now three significant consequences of this blurring.

First, and finest level of granularity, the traditional instruction/assessment distinction is blurred. In digitally-mediated learning environments, every moment of learning can be a moment of feedback; and every moment of feedback can be recorded as a datapoint. The grain size of these datapoints may be very small, so small and thus soon become so many that they are of necessity the kind of evidence that would have almost entirely been lost to the traditional teacher, assessor or researcher. For instruction and assessment to become one, however, these need to be ‘semantically legible datapoints’. Our definition of a semantically legible datapoint is ‘learner-actionable feedback’. Every such datapoint can offer an opportunity that presents to the learner as a ‘teachable moment’. These datapoints can take forms as varied as the ten data types we described earlier, involving either or both machine response to learner action or machine-mediated human response, thereby harnessing both collective human intelligence and artificial intelligence. Such learning environments, where the distinctions between instruction and assessment are so blurred (Armour-Thomas and Gordon 2013), might even require that we move away
from the old terminology. Perhaps we might need to replace the instruction/assessment dualism with a notion of ‘reflexive pedagogy’.

Second, the distinction between formative and summative assessment is blurred. Semantically legible datapoints that are ‘designed in’ can serve traditional formative purposes (Black and Wiliam 1998; Wiliam 2011). They can also provide evidence aggregated over time that has traditionally been supplied by summative assessments. This is because, when structured or self-describing data is collected at these datapoints, each point is simultaneously a teachable moment for the learner, and a waypoint in a student’s progress map that can be analyzed in retrospective progress analysis. Why, then, would we need summative assessments if we can analyze everything a student has done to learn, the evidence of learning they have left at every datapoint? Perhaps, also, we need new language for this distinction? Instead of formative and summative assessment as different collection modes, designed differently for different purposes, we need a language of ‘prospective learning analytics’, and ‘retrospective learning analytics’, which are not different kinds of data but different perspectives and different uses for a new species of data framed to support both prospective and retrospective views. One example: the criteria and level descriptions in a rubric are spelt out differently when they have this both-ways orientation, prospective/constructive as well as retrospective/judgmental (Cope and Kalantzis 2013).

Third, the distinction between assessment and research data collection processes is also blurred. The only difference between assessment and research might be the scale of analysis. However, the evidence used by researchers working in the emerging field of educational data science is grounded in the same data—the data of reflexive pedagogy and retrospective learning analytics. The only difference would be that, at times, researchers may be watching the same data for larger patterns on a different scale—across cohorts, between different learning environments and over time. This domain, we might call ‘educational data science’.

2. A Shift of Focus in Data Collection
Semantically legible data is self-describing, structured data. The meanings should be immediately evident to all parties—learners, their teachers and other parties interested in student learning. However, inferring meanings beyond identically replicated sites of implementation is complex—not only as complex as the variety of immediately incommensurable data types that we outlined in the previous section, but also the widely varied data models used by software environments within a consistent data type—how do you map field to field across databases, tag to tag across differently structured text? How do you align or corroborate data that is distributed across different databases? Things get even more complicated when we add the ‘data exhaust’ emanating from computer-mediated learning environments as unstructured data, where signals have to be differentiated from the surrounding noise.

So, we have a dramatically expanded range of collectable data. But, educational data science is left with a lot of work to do to make sense of these data beyond localized return of self-describing data, then to present these data in meaningful ways to learners, teachers and in research reports.

The size of the challenge is expanded, not simply in proportion with the range of the collectable data. Our ambitions also expand. Learning analytics is expected to do a better
job of determining evidence of deep learning than standardized assessments—where the extent of knowing has principally been measured in terms of long term memory, or the capacity to determine correct answers (Knight, Shum, and Littleton 2013). Can big data help us to shift the focus of what is assessed? As Behrens and DiCerbo characterize the shift to big data, we move from an item paradigm for data collection with questions that have answers that can be current and elicit information, to an activity paradigm with learning actions that have features, offer evidence of behavioral attributes, and provide multidimensional information (Behrens and DiCerbo 2013; DiCerbo and Behrens 2014). How, raising our evidentiary expectations, can educational data sciences come to conclusions about dimensions of learning as complex as mastery of disciplinary practices, complex epistemic performances, collaborative knowledge work and multimodal knowledge representations? The answer may lie in the shift to a richer data environment and more sophisticated analytical tools, many of which can be pre-emptively designed into the learning environment itself, or ‘evidence-centered design’ (Mislevy, Behrens, DiCerbo, and Levy 2012; Rupp, Nugent, and Nelson 2012).

A key opportunity arises if we focus our evidentiary work on the knowledge artifacts that learners create in digital media—a report on an science experiment, an information report on a phenomenon in the human or social world, a history essay, an artwork with exegesis, a video story, a business case study, a worked mathematical or statistical example, or executable computer code with user stories. These are some of the characteristic knowledge artifacts of our times. In the era of new media, learners assemble their knowledge representations in the form of rich, multimodal sources—text, image, diagram, table, audio, video, hyperlink, infographic, and manipulable data with visualizations. They are the product of distributed cognition, where traces of the knowledge production process are as important as the products themselves—the sources used, peer feedback during the making, and collaboratively created works. These offer evidence of the quality of disciplinary practice, the fruits of collaboration, capacities to discover secondary knowledge sources, and create primary knowledge from observations and through manipulations. The artifact is identifiable, assessable, measurable. Its provenance is verifiable. Every step in the process of its construction can be traced. The tools of measurement are expanded—natural language processing, time-on-task, peer- and self-review, peer annotations, edit histories, navigation paths through sources. In these ways, the range of collectable data surrounding the knowledge work is hugely expanded.

Our evidentiary focus may now also change. We can manage our ambitions by focusing them on less elusive forms of evidence than traditional constructs such as the ‘theta’ of latent cognitive traits in item response theory, or the ‘g’ of intelligence in IQ tests. In the era of digital we don’t need to be so conjectural in our evidentiary argument. We don’t need to look for anything latent when we have captured so much evidence in readily analyzable form about the concrete product of complex knowledge work, as well as a record of all the steps undertaken in the creation of that product.

We also need to know more than individualized, ‘mentalist’ (Gergen 2013) constructs can ever tell us. We need to know about the social sources of knowledge, manifest in quotations, paraphrases, remixes, links, citations, and other such references. These things don’t need to be remembered now that we live in a world of always-accessible information; they only need to be aptly used. We also need to know about collaborative
intelligence where the knowledge of a working group is greater than the sum of its individual members. We now have analyzable records of social knowledge work, recognizing and crediting for instance the peer feedback that made a knowledge construct so much stronger, or tracking via edit histories the differential contributions of participants in a jointly created work.

In these ways, artifacts and the processes of their making may offer sufficient evidence of knowledge actions, the doing that reflects the thinking, and practical results of that thinking in the form of knowledge representations. As we have so many tools to measure these artifacts and their processes of construction in the era of big data, we can safely leave the measurement at that. Learning analytics may shift the focus of our evidentiary work in education, to some degree at least, from cognitive constructs to what we might call the ‘artifactual’. Where the cognitive can be no more than putative knowledge, the artifactual is a concretely represented knowledge and its antecedent knowledge processes.

3. A Wider Range of Sample Sizes
In our traditional research horizons, ideal sample size = just enough. In quantitative analyses, we undertook power analyses to reach an ‘n’ point where the numbers gave us enough information to be correct within a small margin of error, and making ‘n’ any bigger would be unlikely to change the conclusion. Traditional experimental methods aimed to optimize research resources and maximize the validity of outcomes by minimizing the sample size ‘n’ and justifying that size with power analyses. In qualitative analyses, we did just enough triangulation to be sure.

In the era of big data, there is no need to figure or justify an optimal sample size, because there is no marginal cost of making n = all. There are no possibilities of sample error or sample bias. This sample might be all the users of a piece of software, or all the students in a school or district. If we want to do tightly controlled experimental work, for instance to test whether the beta version of a new feature should be implemented, within n = all we can create A/B differences in the software. Then we need to be sure that each of A and B are big enough to support the conclusions we hope to reach (Tomkin and Charlevoix 2014).

At the same time, n = 1 becomes a viable sample size, because massive amounts of data can be gleaned to create a composite picture of an individual student. From a data point of view, the learner now belongs to a ‘school of one’ (Bienkowski, Feng, and Means 2012). Single cases can be grounded in quantitative as well as qualitative data. And, of course, as soon as n = 1 and n = all become equally viable, so does every data size between. There is no optimal n.

4. Widening the Range of Timeframes in Intervention-Impact Measurement Cycles
Typically, education experiments have required one semester or even a year or more to demonstrate an overall effect—of a new curriculum, or the application of new educational technology for instance. The whole experiment is carefully mapped out before the start. You can’t change your research questions half way through. You can’t address new questions as they arise. Research is a linear process: research design => implementation => analysis => report results.
However, when structured data instrumentation is embedded and unstructured data collected and analyzed, non-linear, recursive micro intervention-result-redesign cycles are possible. This can make for more finely grained and responsive research, closely integrated into the design and phased implementation of interventions. Such an approach builds upon traditions of design experiments (Laurillard 2012; Schoenfeld 2006) and micro-genetic classroom research (Chinn 2006).

Methodologies requiring longer research time cycles are out of step with contemporary software design methodologies. A first generation of software design followed a linear engineering logic whose origins are in construction and manufacturing industries: requirements specification => technical specification => alpha complete code => beta implementation testing => release => maintenance. This approach is often termed ‘waterfall’—and is reflected in the long versioning and release cycles of desktop software programs.

A new generation of software development tools has moved to an iterative, recursive methodology termed ‘agile development’ which emphasizes rapid, frequent and incremental, design, testing and release cycles (Martin 2009; Stober and Hansmann 2009). This why changes in social media platforms are continuous and often barely visible to users. In this context, research is ideally embedded within development, and is itself as agile as the development processes it aims to assist. Partial implementation in real use cases generate user data, and this in turn produces new research questions and new design priorities—a process that repeats in rapid cycles. In our Scholar project, for instance, the research and development cycles last two weeks. ‘User stories’ are generated based on issues arising in the previous two weeks of implementation, coding occurs, then after two weeks they are either released to n = all, or to a subgroup if we want to compare A/B effects to decide whether the change works for users. Research and design are fluid, re-planned every two weeks. Design is in fact a process of co-design, based on an experimental co-researcher relationship with users. Research happens in ‘real time’, rather than fixed, pre-ordained researcher time.

So, in today’s digital learning environments, the research timeframes frequently need to be shorter. Their logic has to be recursive rather than linear. However, data persistence also offers the possibility of undertaking analyses that are longitudinal. This new kind of longitudinal research need not be pre-determined by research design. Rather, it may consist of retrospective views of data progressively collected and never deleted. In these ways, with its range of shorter to longer timeframes, big data offers a multi-scalar temporal lens, where it is possible to zoom in and out in one’s view of the data, reading and analyzing the data across different timescales.

5. More Widely Distributed Data Collection Roles
In a conventional division of research labor, the researcher is independent, counseled to observe and analyze with a disinterested objectivity. To recruit research subjects as data collectors, interpreters and analysts would be to contaminate the evidence. Researcher and research subject roles are supposed to be kept separate.

With the rise of big data, we begin to recruit our research subjects—students as teachers—as data collectors. This is supported by a logic of the ‘wisdom of crowds’ in online and big data contexts which, in any event, dethrones the expert (Surowiecki 2004; Vempaty, Varshney, and Varshney 2013). To the extent that it is still required, expert human judgment can be meaningfully supplemented by non-expert judgments such as
those of students themselves (Strijbos and Sluijsmans 2010). Web 2.0 technologies have demonstrated the effectiveness of non-expert reputational and recommendation systems (Farmer and Glass 2010; O'Reilly 2005). In our own research on science writing in the middle school, we have shown that when rating level descriptors are clear, mean scores of several non-expert raters are close to those of expert raters (Cope, Kalantzis, Abd-El-Khalick, and Bagley 2013). These findings are corroborated in many studies of peer assessment (Labutov, Luu, Joachims, and Lipson 2014; Piech, Huang, Chen, Do, Ng, and Koller 2013).

In fact, we can even argue that new power arises when the range of data collection and analysis roles is extended. All now are interested parties, even the researcher, who may be able to declare an interest to explore ways to improve an online learning environment. Including a range perspectives (technically, permissions and roles in a software systems), may lead to the design of moderation processes that produce truer results.

6. Cyberdata
In coining the term 'cybernetics', Norbert Weiner attempted to capture the logic of self-adjusting systems, both mechanical and biological (Weiner 1948 (1965)). The Greek kybernetis, or oarsman, adjusts his rudder one way then another, in order to maintain a straight path. One aspect of big data is that it is self-adjusting, and where the machine is capable of learning, either supervised or unsupervised by humans. This domain is also at times called 'artificial intelligence'. For instance, Google Translate undertakes pattern analyses across a massive database of translated documents, and uses the results of these machine analyses to translate new documents—an example of unsupervised machine learning. Automated writing assessment technologies compare human-graded assessments to as-yet ungraded texts, in order to assign a grade based on statistical measures of similarity—an example of supervised machine learning (Vojak et al. 2011). These kinds of data we term cyberdata, because the instruments of data collection and analysis are at least to some degree themselves machine-generated. We researchers now work with machines as collaborators in data collection and analysis. The machine does not just collect data; it becomes smarter in its collection and analysis the more data it collects (Chaudhri, Gunning, Lane, and Roschelle 2013; Woolf 2010; Woolf, Lane, Chaudhri, and Kolodner 2013).

7. Blurring the Edges Dividing Qualitative from Quantitative Research
In the traditional division of research types, by dint of pragmatic necessity, quantitative research has larger scale/less depth, while qualitative research has smaller scale/more depth. Mixed methods are an often uneasy attempt to do a bit of both, supplementing the one method with the other, corroborating conclusions perhaps but mostly leaving the data and analyses divided into separate islands.

With a new generation of big data, we can perform analyses on large unstructured or hard-to-analyze datasets that were the traditional focus of qualitative research, such as large bodies of natural language. We can get good-enough transcriptions of speech. We can analyze other non-quantitative data such as image and sound. With these tools for reading, there are no logistics of scale for qualitative research. If Structured Query Language worked for a previous generation of structured datasets, NoSQL processes
unstructured, lightly structured and heterogeneous data. Now, the edges dividing qualitative and quantitative research blur, and this is particularly important when our computer-mediated learning environments frequently contain data relating to an experience of learning that is amenable to both modes of analysis.

8. Rebalancing the Empirical and the Theoretical
Here’s an irony of big data. It’s not just quantitative. It demands more conceptual, theoretical, interpretative, hermeneutical—indeed qualitative—intellectual work than ever.

Our counterpoint for this case is a celebrated 2008 article by the editor of Wired magazine, Chris Anderson, ‘The End of Theory’ (Anderson 2008). In it, he claimed that the data was now so big—so complete and so exhaustive—that it could be allowed to speak for itself. In reality the opposite is true. Empirical evidence of the Higgs Boson would not have been found in the avalanche of data coming from the Large Hadron Collider if there had not already been a theory that it should exist. The theory that conceived its possibility made it visible (Hey, Tansley, and Tolle 2009). The theory in this case was a form of reasoning, played through in mathematical models for sure, but not simply generated from empirical data.

And even with the most comprehensive data and sophisticated methods of calculation, nothing but qualitative thinking will get us past spurious correlations, or help us to explain the meanings behind the numbers.

9. A Wider Range of Causal Arguments
This leads us to questions of causation, or the meaning we might validly ascribe to the data. Statisticians have generally been reticent to make generalizations about the ‘how’ that is causation. Instead, they take the more cautious route of calculating the ‘what’ of correlation. At one point only do they throw caution to the wind—they are willing to say that randomized controlled experiments are strong enough to demonstrate cause (Pearl 2009: 410). Under these conditions alone can ‘an intervention, such as a curricular innovation … be viewed as the cause of an effect, such as improved student learning.’ A causal effect can be inferred when the ‘difference between what would have happened to the participant in the treatment condition and what would have happened to the same participant if he or she had instead been exposed to the control condition’ (Schneider, Carnoy, Kilpatrick, Schmidt, and Shavelson 2007: 9, 13).

In the era of big data, however, randomized controlled experiments are at times nowhere near cautious enough; at other times they can be helpful, if for instance parts of their logic are translated into A/B studies (where n=all is divided into ‘A’ and ‘B’ groups, and functional ‘beta’ variations introduced into the ‘B’ group) and at other times again they are paralyzingly over-cautious. Or, another way of saying this, in the era of big data, valid causal arguments can be developed across a wider range of frames of reference.

Here is how randomized controlled experiments are not cautious enough: typically, they base their causal inferences on quantitative generalizations about whole interventions for averaged populations over identical timeframes. Gross cause may thus be claimed to have been demonstrated. However, in this methodology the intervention is causal black box. A single, gross cause is associated with a single gross effect, without
providing causal explanation of what happened inside the box (Stern, Stame, Mayne, Forss, Davies, and Befani 2012: 7).

In a situation where data collection has been embedded within the intervention, it is possible to track back over every contributory learning-action, to trace the microdynamics of the learning process, and analyze the shape and provenance of learning artifacts. This is the sense in which randomized controlled experiments are not cautious enough in their causal generalizations.

Moreover, the dynamics of cause within the black box will never be uniform, for individuals or subcategories of individual. Big data is inevitably heterogeneous, and contemporary data science addresses heterogeneity. Finely grained causal analysis can now be done, including discovery of contributory casual patterns that could not have been anticipated in research hypotheses, but which may stand out in ex post facto micro-genetic causal analysis. Now we can and should be much more measured in our causal inferences.

In these ways, it is possible to drill down into big educational data, reaching tiny moments of learning at points where the microdynamics of learning become more visible. Building back up from these it is possible to trace the macro-dynamics that constitute overall effects. Educational data science needs to apply and extend methods of network mapping, systems analysis, model development, diagramming and visualization in order to support such fine-grained causal explanations (Maroulis, Guimerà, Petry, Stringer, Gomez, Amaral, and Wilensky 2010).

We still may want to include randomized experimental intervention in our expanded frame of reference for causal analysis. There may be interesting high level generalizations that can be deduced in randomized A/B studies for instance, undertaken with rigor equal to any experimental study. However, when n = all, we don’t need to support our methodology with power analyses. At this level of focus, a modified experimental method may be just right for big data. This is not to say that the numbers will speak to truth in statistical models. We will have to use non-statistical linguistic reasoning to address spurious correlations, which will be all the more frequent now that there is so much data. Indeed, even when statistical rules have been applied which appear sufficiently rigorous to warrant attribution of cause, we still need to be careful. Refiguring the methodological circumstances around quantitative medical science, Ioannidis concludes by way of salutary warning, ‘most published research findings are false’ (Ioannidis 2005).

Moreover, we can with equal validity apply other forms of causal reasoning to big educational data. In everyday life, few of our generalizations are statistically derived or even derivable. Most causes are definitively learned quite directly and easily. ‘A child can infer that shaking a toy can produce a rattling sound because it is the child’s hand, governed solely by the child’s volition, that brings about the shaking of the toy and the subsequent rattling sound.’ We are also given causes from the collective wisdom of linguistic inputs. ‘The glass broke because you pushed it’ (Pearl 2009: 253).

When we reach semantically legible datapoints, we can see the microdynamics of learning this directly and clearly. Neither student, nor teacher, nor researcher needs controls or randomization to make meaningful generalizations. In the case of data generated incidental to learning, it is possible to drill down into every constituent datapoint in order to explore precisely what happened to produce a particular outcome,
for an individual student or for a category of student identified—a student who is an outlier, or in a certain demographic category, or who made a particular mistake, or who otherwise excelled. In other words, we can look into data to determine the micro-dynamics and aggregated macro-dynamics of causation in ways that were not practicable in the past.

In big data, the causal chains can be made practically visible. The data are self-describing. The causal mechanisms can be explained in language, as revealingly in a single instance as many, and a single instance may be sufficient to make a certain kind of case.

10. Opening a Window on Variation

Because the statistical solution to the fundamental problem of causal inference estimates an average effect for a population of participants or units, it tells us nothing about the causal effect for specific participants or subgroups of participants’ (Schneider et al. 2007: 19). Big data can open a window on variation, supplementing and extending causal arguments in traditional experimental research. These are limited in their conclusions to gross effects, undifferentiated ‘what works’ pronouncements, and linear models of causal inference. Uniformity is posited. Averages are normalized. Research subjects are homogenized. This renders a certain kind of research outcome, but not others that might be equally helpful to us. Of course, experimental research can have subgroup analyses, though this requires an increase in sample size and the complexity of the research analysis, and then the subgroup must still be considered uniform.

Big data offers the potential to fill out the bigger picture with finely grained detail: the differences between individuals and similarities within subgroups, multiple causality, contributing factors, contingencies, non-linear pathways, causes and effects that are mutually influential, and emergent patterns. The divergences and outliers are at least as important as the median and the norm. Researchers are beginning to focus on learner differences as a critical factor in computer-mediated learning environments (Khajah, Wing, Lindsey, and Mozer 2014; Lee, Liu, and Popovic 2014; Snow, Varner, Russell, and McNamara 2014).

Big data analyses, designed to detect and interpret variation are also essential as we analyze learning environments whose intrinsic mechanism and advertised virtue is divergence—variously named as adaptive or personalized learning (Conati and Kardan 2013; Koedinger, Brunskill, Baker, and McLaughlin 2013; McNamara 2012; McNamara and Graesser 2012; Wolf 2010). Standardized research interventions demand fidelity or strict uniformity of implementation. However, in computer-mediated learning environments, recursive, dynamic, recalibrating systems are the new norm. Adaptive and personalized learning environments are unstandardized by design. The data they generate are dynamic because they are built to be self-adjusting systems. They are difference engines. Fortunately, the same systems that collect this data for the purpose of flexible adaptivity, can record, track and provide analytics which account for variation.
New Designs for Research Infrastructure

Not only does big data invite new research practices. These practices require the creation of new research infrastructures. Following are several infrastructure requirements among many that may be needed to further the mission of educational data sciences.

Publishing Data
Historically, the primary artifact documenting research activities and disseminating research results in the social sciences has been the scholarly journal article which synthesizes results but does not provide re-analyzable source data. In the era of print and print-emulating PDF, it was not feasible to provide re-manipulable source datasets. However, the journal system is currently in a phase of radical transformation as a consequence of the rise of online publishing, particularly in the natural and medical sciences (Cope and Kalantzis 2014).

One key aspect of the general transformation of the journal system in the context of online publication, is the possibility of publishing related datasets alongside journal articles (Hey, Tansley, and Tolle 2009: xxiv). The American Educational Research Association has had data sharing and data access policies in place since 2006. Since 2011, AERA has begun to develop a relationship with the Inter-university Consortium for Political and Social Research to publish datasets generated in educational research. More recently, the largest of the commercial journal publishers, Elsevier, has undertaken to make article-related datasets available on a no charge and open access basis, even if articles themselves still need to be purchased or accessed via subscription. Meanwhile, we have witnessed rapid growth in institutional and disciplinary repositories, also including significant datasets (Lynch 2008; Shreeves 2013).

This development promises to offer enormous benefits to researchers, including replication and extension studies, and meta-analyses which can dive deeper into subsidiary analyses than current meta-analyses, which can in practice go little further than to aggregate reports of gross effects (Glass 2006; Hattie 2009). The benefits of publishing data are already evident in the natural sciences—and most obviously in fields such as astrophysics and bioinformatics—where open data are yielding insights that might not otherwise have been gained. In a scientific version of what Lawrence Lessig calls ‘remix culture’ (Lessig 2008), school students are finding supernova and citizen scientists are protein folding, using published open access data.

Critical questions arise for the kinds of educational data science that we have been describing in this paper, for which new or expanded research infrastructures are required. The key challenge here is the collection, curation, and analysis of vast quantities of disparate data. Information scientist Allen Renear outlines the requirements as follows. All aspects of the environment need to be instrumented to ensure that as much data as possible is collected without relying on a prior determination of what is important and what is ‘noise’. The management of vast quantities of varied data requires highly specialized data curation, including standard and specialized metadata to support discoverability, usefulness, and reliability. Transformations to create higher levels of meaningful data from lower levels of raw data need to be documented and auditable—these transformation must generate computer-processable documentation of data provenance and workflow to ensure results are sound and reproducible. Conversion
processes are required to support integration with additional data and fusion with complementary data, as well as support for interoperating varied analysis tools—these conversion processes must also meet current high standards of data provenance and workflow documentation. Data need to be regularly enhanced and corrected to ensure continuing value and reliability—also managed through a system of auditable authenticity and version control. They must be monitored to comply with current standards for privacy and security regulations and policies, with additional compliance beyond requirements in order to ensure that all community stakeholders are comfortable, and will continue to be comfortable, with data collection and use (Nichols, Twidale, and Cunningham 2012; Renear and Palmer 2009; Wickett, Sacchi, Dubin, and Renear 2012).

One current initiative, in which the authors of the paper have been involved, is the National Data Service (http://www.nationaldataservice.org) whose development is being led by the National Center for Supercomputing Applications at the University of Illinois. The charter of NDS is to develop a framework consisting of an ‘international federation of data providers, data aggregators, community-specific federations, publishers, and cyber infrastructure providers’ that ‘builds on the data archiving and sharing efforts underway within specific communities and links them together with a common set of tools’. This will include a data repository or ‘hub’ where data and related instrumentation can optionally be stored with a specialized focus on very large, noisy datasets, and a metadata service linking datasets that are stored in institutional or publisher repositories, thus building federated linkages across disparate educational data repositories.

**Data Standards Development**

In the broader context of big data we have witnessed in recent times the emergence of standards and processes for aligning datasets whose data models diverge but whose empirical reference points overlap. A key challenge for the integration of data from the different types of computer-mediated learning environments, and also the replication, extension and meta-analysis of educational research data, is the commensurability of data models.

To address this challenge, a number of important standards have been developed, each reflecting the specific needs of groups of stakeholders. A 2009 initiative of the US Department of Education, the Common Education Data Standards (CEDS) supports interoperability of educational data across states and levels of education, through a data dictionary and mappings into localized data schemas (http://ceds.ed.gov). An initiative of Microsoft in 1998, the Schools Interoperability Framework (SIF) was created to develop a standard that enhanced interoperability between education applications (http://sifassociation.org). Founded in 1997, the IMS Global Learning Consortium has taken a leading role in the development of data standards for educational software environments, including the Learning Tools Interoperability Standard, now adopted by all leading learning management system and web software developers, including the partners to this proposal (http://www.imsglobal.org/).

These standards, however, are not designed to support the fine-grained data generated by computer-mediated learning environments. In recognition of this need, a working group within the IMS Consortium has since 2013 been working towards development is the ‘Caliper’ standard, designed specifically to address the challenge of comparing, correlating and contrasting learning activities across platforms, assessing the nature and

**Developing a Shared Vocabulary**
A key challenge in the era of readily accessible data is incommensurable data models. This is particularly vexing in the context of semantic overlap (Cope, Kalantzis, and Magee 2011)—when for instance, a learning environment collect structured data based on variant data models, perhaps completing a simulation, rating a written report against a rubric, and taking a quiz, all in the same topic area.

The widespread adoption of standards will undoubtedly offer a good deal of assistance in support of integrative and comparative big data analyses. However, the problem remains of historical or new datasets created in non-conforming data models, including the ‘big data’ generated in computer-mediated learning environments. Nor do standards, without additional support, provide semantically stable relations with other online data models such as web standards in the areas of geographic indicators, demographic classifications, discipline schemas, and topic descriptors.

One solution would be to support the distributed development by educational researchers, of an educational data dictionary. Such a development has been recommended by Woolf (Woolf 2010: 65); as has a notion of ‘open learning analytics’ that ties together data emerging from a number of platforms (Siemens, Gasevic, Haythornthwaite, Dawson, Shum, Ferguson, Duval, Verbert, and Baker 2011).

An educational data dictionary would serve two purposes. At a description layer, it would be a way to be clear that what we mean by data descriptors, and to map synonyms where different schemas or data models gave different labels for semantically commensurable things. In other words, it would support data consistency via definitions and synonym alignment in schema-to-schema crosswalks. At a functional layer, using ontology-matching technologies, it is possible to facilitate data interoperability, integration and transfer. This might occur either automatically or via queries where a particular alignment remains ambiguous and the machine transliterations require further human training (Cope 2011; Cope, Kalantzis, and Magee 2011).

Ideally, it would be members of the educational research community who build out the content of such a dictionary in a wiki-like environment, facilitating the interaction of standards-based curation and formal data modeling (ontologies), and field-initiated informal discussion of educational terminology for tagging (‘folksonomies’) (Martinez-Garcia, Simon Morris, Tscholl, Tracy, and Carmichael 2012; Wichowski 2009). For each dictionary term, there might be a natural language definition, formal schema synonyms, such as in CEDS, IMS and SIF, and term relations—parent/child, sibling, part/kind, causal chains, etc. Export/import facilities with ontology building and automated reasoning technologies would facilitate big data analyses and learning analytics research. Infrastructure along these lines is necessary if we are to make progress in educational data science.
Research Ethics Reconfigurations

Big data in the natural sciences do not present the ethical requirements of consent and privacy that arise in the social and medical sciences. In traditional research models, clear relationships of consent could be established in agreements preceding any research activity. Big data, however, is found or harvested data. It is not a product of carefully prefigured research design. Conversely, big data research also frequently involves deeper intervention into social practices. This research (such as the A/B research) is itself a form of social engineering. Consent to be involved has casually been given in user agreements at the point of account sign up, and no further consent is requested or provided for subsequent research using these data.

For this reason, Kramera and colleagues saw nothing exceptional about their Facebook study in which 700,000 Facebook users were split into A and B groups who and then fed different mixes of positive and negative posts. “Experimental Evidence of Massive-scale Emotional Contagion Through Social Networks” was the alarming title of the subsequent paper in which they report on the results of this experiment (Kramera, Guillory, and Hancock 2014). Facebook users did indeed become alarmed, for what they regarded to be the unconscionable manipulation of their emotions. The IRB approval from Cornell University relied on consent via the research subjects’ Facebook user agreement—in which among other things, users consent to be watched and analyzed, to receive push messages, to be manipulated by the arrangement of posts by Facebook in the feed, and to cede unrestricted ownership of intellectual property to Facebook.

Recognizing that traditional consent protocols are impractical in the context of big data, the Committee on Revisions to the Common Rule for the Protection of Human Subjects in Research in the Behavioral and Social Sciences has recommended a new category of ‘excused’ research (National Research Council 2014). This is Recommendation 5.5: ‘Investigators using non-research private information (e.g., student school or health records) need to adhere to the conditions for use set forth by the information provider and prepare a data protection plan consonant with these conditions.’ This is how the Facebook research was justified. And here is Recommendation 2.3: ‘New forms of large-scale data should be included as not human-subjects research if all information is publicly available to anyone (including for purchase), if persons providing or producing the information have no reasonable belief that their private behaviors or interactions are revealed by the data, and if investigators have no interaction or intervention with individuals.’ The ‘not-human-subjects’ semantics reads counter-intuitively. Also, although secure anonymization is one of the promises of big data, if the un-named data provided on an individual is complete enough, the identity of that individual can be inferred using big data analytics, too.

In both areas of concern—manipulation and privacy—big data presents big challenges. The data are not neutral. Both collection and analysis protocols are often purpose-designed for social engineering. Sometimes the effects are contrary to the interests of individuals, for instance in cases where profiling has a discriminatory effect. Predictive analytics can be used to raise your insurance premium, increase your chance of arrest, or pre-determine your place in learning track (Heath 2014; Mayer-Schönberger and Cukier 2013: 151, 160; Podesta et al. 2014). Such recursive data-subject relationships are an intrinsic feature of big data. However, if big data are not to become big brother,
users need to be recruited as co-collectors, co-analyzers, co-researchers—equal parties in the data-driven decisions that may today be made over their own lives.

Moving Forward with Educational Research in the Era of Big Data

The incidental recording of actions and interactions in computer-mediated learning environments opens out new sources of educational data, and new challenges for researchers. The emerging research and development frameworks that we have described in this paper are deeply embedded in real-time learning. As a consequence, research is more tightly integrated into development. The cycles of feedback are more frequent. And learning science is moving faster. This is both inevitable and necessary when the learning environments themselves are so dynamic and fluid. Research becomes integral to the processes of self-adjustment and continuous, iterative development that characterizes these environments.

In this paper, we have attempted to demonstrate the ways in which educational data science generates different kinds of data, requiring new kinds of evidentiary reasoning that might work as a supplement to, and in some circumstances as a substitute for, traditional education research methods. Might we also venture to suggest that in time, in the context of big educational data, some of the old, heavy duty, industrial-strength research methodologies might go the way of the gold standard?

References


Mislevy, Robert J. 2013. "Four Metaphors We Need to Understand Assessment." The Gordon Commission, Princeton NJ.


125-134 in *LAK '13 Proceedings of the Third International Conference on Learning Analytics and Knowledge* New York NY.


