



Interpreting Evidence-of-Learning: Educational research in the era of big data

Bill Cope & Mary Kalantzis

To cite this article: Bill Cope & Mary Kalantzis (2015) Interpreting Evidence-of-Learning: Educational research in the era of big data, Open Review of Educational Research, 2:1, 218-239, DOI: [10.1080/23265507.2015.1074870](https://doi.org/10.1080/23265507.2015.1074870)

To link to this article: <http://dx.doi.org/10.1080/23265507.2015.1074870>



© 2015 The Author(s). Published by Taylor & Francis.



Published online: 12 Sep 2015.



Submit your article to this journal [↗](#)



Article views: 1



View related articles [↗](#)



View Crossmark data [↗](#)

Interpreting Evidence-of-Learning: Educational research in the era of big data

BILL COPE & MARY KALANTZIS

University of Illinois, Champaign, IL

Abstract

In this article, we argue that big data can offer new opportunities and roles for educational researchers. 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. We explore the methodological consequences of these developments for research methods.

Keywords: big data, educational research, research methodologies, data science, data publishing, research ethics

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 & Behrens, 2014; Piety, 2013; West, 2012). These, in turn, are aspects of the broader social phenomenon of big data (Mayer-Schönberger & Cukier, 2013; Podesta, Pritzker, Moniz, Holdern, & Zients, 2014).

A measure of the newness of this phenomenon is American Educational Research Association's (AERA's) *Handbook of Complementary Methods in Education Research* (Green, Camilli, Elmore, & 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 article, 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 laboratory 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 can collect 30 terabytes of data per night, covering the whole sky in four days. Computers comparing patterns in these digital images are 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 100 times more genetic indicators than their host; the microbic mix can be as much as 50% 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).

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 & Cukier, 2013), and even evolving into a new paradigm of science (Hey, Tansley, & Tolle, 2009).

In this article, 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.

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, & Means, 2012, p. ix). West (2012, p. 1) directs our attention to “real-time” assessment [with its] ... potential for improved research, evaluation, and accountability through data mining, data analytics, and web dashboards’. Behrens and DiCerbo (2013, p. 9) 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 teaching and learning activities become increasingly digital, the need for separate assessment activities should be brought increasingly into question’. Chung (2013, p. 3) 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’. DiCerbo and Behrens (2014, p. 8) 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’.

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 & Siemens, 2014; Castro, Vellido, Nebot, & Mugica, 2007; Siemens & 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 *et al.*, 2012; Knight, Shum, & Littleton, 2013; Mislevy, Behrens, DiCerbo, & Levy, 2012; Siemens & Baker, 2013; West, 2012).

These new fields build upon older subdisciplines 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 & Wiliam, 1998; Pellegrino, Chudowsky, & Glaser, 2001); technology-mediated psychometrics (Chang, 2015; Rupp, Nugent, & Nelson, 2012); self-regulated learning and metacognition (Bielaczyc, Pirolli, & Brown, 1995; Bransford, Brown, & Cocking, 2000; Kay, 2001; Kay, Kleitman, & Azevedo, 2013; Lane, 2012; Magnifico, Olmanson, & Cope, 2013; Shepard, 2008; Winne & Baker, 2013); and analyses of complex performance and holistic disciplinary practice (Greeno, 1998; Mislevy, Steinberg, Breyer, Almond, & Johnson, 2002).

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 is 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, & Whitaker, 2013).

Even more challenging than the bigness, the sources of evidence are varied—including computer-adaptive and diagnostic tests, automated essay scoring, learning games, social interaction analyses, affect meters, body sensors, intelligent tutors, simulations, semantic mapping, and learning management system data. Not only does each of these sources 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 10 methodological extensions of, or supplements to, traditional educational research methods when big data has come to school. They are ordered in an argument that starts with the nature of the data, and moves towards challenging questions of causality and generalizability in the context of learner differences. We will focus for our analysis mainly on what we call ‘semantically legible’ structured data where datapoints have been designed into the learning environment, and less on other data types such as computerized tests and the mining of unstructured data.

1. *Multi-scalar Data Collection*

Evidence of learning has always been gleaned in a series of feedback loops at different scales, from the micro-dynamics of instruction, to assessment, to research. In the micro-dynamics 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 at the finest level of granularity, the traditional instruction/assessment distinction is blurred (Armour-Thomas & Gordon, 2013). 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 in fact that they soon become too many to interpret without computer assistance. This kind of evidence 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 involve either or both a machine response to learner action or machine-mediated human response, thereby harnessing both collective human intelligence and artificial intelligence.

Second, the distinction between formative and summative assessment is blurred. Semantically legible datapoints that are ‘designed in’ can serve traditional formative purposes (Black & 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?

Third, the distinction between assessment and research data collection processes is also blurred. The main difference between assessment and research data 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 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 are 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 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.

As a consequence, 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 the localized return of self-describing data in the form of feedback, 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, & 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 & DiCerbo, 2013; DiCerbo & Behrens, 2014).

A key opportunity arises if we focus our evidentiary work on the knowledge artifacts that learners create in digital media (Berland, Baker, & Blikstein, 2014). 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.

We also need to know more than individualized, ‘mentalist’ (Dixon-Román & Gergen, 2013) or purely individual-cognitive 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 do not 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.

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 $n = \text{all}$. There are no possibilities of sample error or sample bias. The population being researched might be all the users of a piece of software, or all teachers and classrooms, or all the students in a school or district. So, we may not need to sample at all, or worry about generalizability of the sample. 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 $n = \text{all}$ we can create A/B differences in the software and the create A/B groups of users for each variant. Then we need to be sure that each of A and B are big enough to support the conclusions we hope to reach (Tomkin & 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 *et al.*, 2012). Single cases can be grounded in quantitative as well as qualitative data. And, of course, as soon as $n = 1$ and $n = \text{all}$ 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 cannot change your research questions half way through. You cannot 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 & Hansmann, 2009). This is 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,¹ 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 $n = \text{all}$, 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, & 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 & Sluijsmans, 2010). Web 2.0 technologies have demonstrated the effectiveness of non-expert reputational and recommendation systems (Farmer & 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, & Bagley, 2013). These findings are corroborated in many studies of peer assessment (Labutov, Luu, Joachims, & Lipson, 2014; Piech *et al.*, 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 of perspectives (technically, permissions and roles in a software system), 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. Today, this domain is called ‘machine learning’. 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, Kline, Cope, McCarthey, & Kalantzis, 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 (Barnes & Stamper, 2008; Chaudhri, Gunning, Lane, & Roschelle, 2013; Woolf, 2010; Woolf, Lane, Chaudhri, & 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*

A paradox of big data is that it is not just empirical-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. Other data scientists have argued that processes of ‘data smashing’ will produce results that emerge directly from the data (Chattopadhyay & Lipson, 2014). 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 *et al.*, 2009). In this case, the theory was a form of reasoning, played through in mathematical models for sure, but not simply generated from empirical data. Moreover, 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. In fact, theory is written through the data universe when data models are written into the labels applied to data fields, database structures, XML markup schemas or domain ontologies (Cope, Kalantzis, & Magee, 2011). Theory, in other words, not only frames research question and the search for salient evidence, it is written into the most granular moments of data collection.

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 and certain quasi-experimental methods are strong enough to demonstrate cause (Pearl, 2009, p. 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, & Shavelson, 2007, pp. 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 $n = \text{all}$ 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 a causal black box, or the variables are at least limited to the few that are manageable. Often, and most straightforwardly for this genre of research, a single, gross cause is associated with a single gross effect, without providing causal explanation of what happened inside the box (Stern *et al.*, 2012, p. 7; Winne, 2006).

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 micro-dynamics 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 micro-dynamics 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 *et al.*, 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 $n = \text{all}$, we do not 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 the 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.

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, p. 253).

When we reach semantically legible datapoints, we can see the micro-dynamics of learning this directly and clearly (Winne & Baker, 2013). Neither student, nor teacher, nor researcher needs a control 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. And in the case of structured data, 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, p. 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 it is the subgroup that must 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, & Mozer, 2014; Lee, Liu, & Popovic, 2014; Snow, Varner, Russell, & 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 & Kardan, 2013; Koedinger, Brunskill, Baker, & McLaughlin, 2013; McNamara & 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 & Kalantzis, 2014).

A 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

et al., 2009, p. xxiv). The AERA 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 (2008) calls ‘remix culture’, 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 article, 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, & Cunningham, 2012; Renear & Palmer, 2009; Wickett, Sacchi, Dubin, & Renear, 2012).

One current initiative, in which the authors of this article have been involved, is the National Data Service (NDS) (<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 meta-data service linking datasets that are stored in institutional or publisher repositories, thus building federated linkages across disparate educational data repositories. Another initiative, specifically addressing educational data is the Pittsburgh Science of Learning Center DataShop (Koedinger *et al.*, 2010)

Data Standards Development

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.

Even 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 cannot 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. 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.

One possible response to this challenge is to develop data standards and processes for aligning datasets whose data models diverge but whose empirical reference points overlap. A number of 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 quality of learner interactions, and evaluating learner performance outcomes (IMS Global Learning Consortium, 2013).

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 *et al.*, 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 (2010, p. 65); as has a notion of ‘open learning analytics’ that ties together data emerging from a number of platforms (Siemens *et al.*, 2011).

An educational data dictionary would serve two purposes. At a description layer, it would be a way to be clear about 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 *et al.*, 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, Morris, Tscholl, Tracy, & 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 often simply found or harvested historical data. It is not a necessarily 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, & Hancock, 2014). Facebook users did indeed become alarmed, for what they regarded to be the unconscionable manipulation of their emotions. The institutional review board (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 (Daries *et al.*, 2014). 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 & Cukier, 2013, pp. 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 article 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 moves 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 article, 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?

Disclosure statement

No potential conflict of interest was reported by the authors.

Note

1. 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>

References

- Anderson, C. (2008, July 16). The end of theory: The data deluge makes the scientific method obsolete. *Wired*.
- Armour-Thomas, E., & Gordon, E. W. (2013). *Toward an understanding of assessment as a dynamic component of pedagogy*. Princeton, NJ: The Gordon Commission.

- Baker, R. S. J. D., & Siemens, G. (2014). Educational data mining and learning analytics. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences*. New York, NY: Cambridge University Press.
- Barnes, T., & Stamper, J. (2008). Toward automatic hint generation for logic proof tutoring using historical student data. In B. P. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems 9th International Conference* (pp. 373–382), Montreal.
- Behrens, J. T., & DiCerbo, K. E. (2013). Technological implications for assessment ecosystems. In E. W. Gordon (Ed.), *The Gordon commission on the future of assessment in education: Technical report* (pp. 101–122). Princeton, NJ: The Gordon Commission.
- Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19, 205–220.
- Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, 13, 221–252.
- Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Washington, DC: Office of Educational Technology, US Department of Education.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7–74.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience and school*, NRC Commission on Behavioral and Social Sciences and Education (Eds.). Washington, DC: National Academies Press.
- Castro, F., Vellido, A., Nebot, A., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. In *Studies in Computational Intelligence* (pp. 183–221). Berlin: Springer-Verlag.
- Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.
- Chang, H. H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, 80, 1–20.
- Chattopadhyay, I., & Lipson, H. (2014). Data smashing: Uncovering lurking order in data. *Journal of The Royal Society Interface*, 11, 20140826–20140826.
- Chaudhri, V. K., Gunning, D., Lane, H. C., & Roschelle, J. (2013). Intelligent learning technologies: Applications of artificial intelligence to contemporary and emerging educational challenges. *AI Magazine*, 34, 10–12.
- Chinn, C. A. (2006). The microgenetic method: Current work and extensions to classroom research. In J. L. Green, G. Camilli, P. B. Elmore, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 439–456). Mahwah NJ: Lawrence Erlbaum.
- Chung, G. K. W. K. (2013). *Toward the relational management of educational measurement data*. Princeton, NJ: The Gordon Commission.
- Conati, C., & Kardan, S. (2013). Student modeling: Supporting personalized instruction, from problem solving to exploratory open-ended activities. *AI Magazine*, 34, 13–26.
- Cope, B. (2011). *Method for the creation, location and formatting of digital content*, USP Office. Champaign, IL: Common Ground Publishing.
- Cope, B., & Kalantzis, M. (2014). Changing knowledge ecologies and the transformation of the scholarly journal. In B. Cope & A. Phillips (Eds.), *The future of the academic journal (second edition)*. Oxford: Elsevier.
- Cope, B., Kalantzis, M., Abd-El-Khalick, F., & Bagley, E. (2013). Science in writing: Learning scientific argument in principle and practice. *e-Learning and Digital Media*, 10, 420–441.
- Cope, B., Kalantzis, M., & Magee, L. (2011). *Towards a semantic web: Connecting knowledge in academic research*. Cambridge: Woodhead Publishing.
- Daries, J. P., Reich, J., Waldo, J., Young, E. M., Whittinghill, J., Seaton, D. T., Ho, A. D., & Chuang, I. (2014). Privacy, anonymity, and big data in the social sciences. *Queue*, 12, 1–12.
- DiCerbo, K. E., & Behrens, J. T. (2014). *Impacts of the digital ocean on education*. London: Pearson.
- Dixon-Román, E. J., & Gergen, K. J. (2013). *Epistemology in measurement: Paradigms and practices*. Princeton, NJ: The Gordon Commission.

- Farmer, F. R., & Glass, B. (2010). *Web reputation systems*. Sebastapol, CA: O'Reilly.
- Glass, G. V. (2006). Meta-analysis: The quantitative synthesis of research findings. In J. L. Green, G. Camilli, P. B. Elmore, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 427–438). Mahwah, NJ: Lawrence Erlbaum.
- Green, J. L., Camilli, G., Elmore, P. B., & Elmore, P. B. (2006). *Handbook of complementary methods in education research*. Mahwah, NJ: Lawrence Erlbaum.
- Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, *53*, 5–26.
- Hattie, J. A.C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Heath, J. (2014). Contemporary privacy theory contributions to learning analytics. *Journal of Learning Analytics*, *1*, 140–149.
- Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond WA: Microsoft Research.
- IMS Global Learning Consortium. (2013). Learning measurement for analytics. Whitepaper. Lake Mary, FL: IMS.
- Kay, J. (2001). Learner Control. *User Modeling and User-Adapted Interaction* *11*, 111–127.
- Kay, J., Kleitman, S., & Azevedo, R. (2013). Empowering teachers to design learning resources with metacognitive interface elements. In R. Luckin, S. Puntambekar, P. Goodyear, B. L. Grabowski, J. Underwood, & N. Winters (Eds.), *Handbook of design in educational technology*. London: Routledge.
- Khajah, M. M., Wing, R. M., Lindsey, R. V., & Mozer, M. C. (2014). *Integrating latent-factor and knowledge-tracing models to predict individual differences in learning* (pp. 99–106). Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), Indianapolis, IN.
- Knight, S., Shum, S. B., & Littleton, K. (2013). Epistemology, pedagogy, assessment and learning analytics. In *Third Conference on Learning Analytics and Knowledge (LAK 2013)* (pp. 75–84). Leuven: ACM.
- Koedinger, K. R., Baker, R. S. J. D., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC datashop. In *Handbook of educational data mining*. Boca Raton, FL: CRC Press.
- Koedinger, K. R., Brunskill, E., Baker, R. S. J. D., & McLaughlin, E. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine*, *34*, 27–41.
- Kramera, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*, 8788–8790.
- Labutov, I., Luu, K., Joachims, T., & Lipson, H. (2014). *Peer mediated testing*. Paper presented at the Data Mining for Educational Assessment and Feedback (ASSESS 2014), New York City.
- Lane, H. C. (2012). Cognitive models of learning. In N. Seel (Ed.) *Encyclopedia of the sciences of learning* (vol. 1, pp. 608–610). Heidelberg: Springer.
- Laurillard, D. (2012). *Teaching as a design science: Building pedagogical patterns for learning and technology*. London: Routledge.
- Lee, S. J., Liu, Y.-E., & Popovic, Z. (2014). *Learning individual behavior in an educational game: A data-driven approach* (pp. 114–121). Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), Indianapolis, IN.
- Lessig, L. (2008). *Remix: Making art and commerce thrive in the hybrid economy*. New York, NY: Penguin Press.
- Lynch, C. (2008). How do your data grow? *Nature*, *455*, 28–29.
- Magnifico, A., Olmanson, J., & Cope, B. (2013). New pedagogies of motivation: Reconstructing and repositioning motivational constructs in new media-supported learning. *e-Learning and Digital Media*, *10*, 484–512.

- Maroulis, S., Guimerà, R., Petry, H., Stringer, M. J., Gomez, L. M., Amaral, L. A. N., & Wilensky, U. (2010). Complex systems view of educational policy research. *Science*, 330, 38–39.
- Martin, R. C. (2009). *Agile software development, principles, patterns, and practices*. Upper Saddle River, NJ: Prentice Hall.
- Martinez-Garcia, A., Morris, S., Tscholl, M., Tracy, F., & Carmichael, P. (2012). Case-based learning, pedagogical innovation, and semantic web technologies. *IEEE Transactions on Learning Technologies*, 5, 104–116.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.
- McNamara, D. S., & Graesser, A. C. (2012). Coh-metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum-Denecke (Eds.) *Applied natural language processing: Identification, investigation and resolution* (pp. 188–205). Hershey, PA: IGI Global.
- Mislevy, R. J. (2013). *Four metaphors we need to understand assessment*. Princeton, NJ: The Gordon Commission.
- Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, 4, 11–48.
- Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (2002). Making sense of data from complex assessments. *Applied Measurement in Education*, 14, 363–389.
- Monroy, C., Rangel, V. S., & Whitaker, R. (2013). STEMscopes: Contextualizing learning analytics in a K-12 Science curriculum. In *Third Conference on Learning Analytics and Knowledge (LAK 2013)* (pp. 210–219). Leuven: ACM.
- National Research Council. (2014). *Proposed revisions to the common rule for the protection of human subjects in the behavioral and social sciences*. Washington, DC: The National Academies Press.
- Nichols, D. M., Twidale, M. B., & Cunningham, S. J.. (2012). Metadatapedia: *A proposal for aggregating metadata on data archiving*. Paper presented at the iConference 2012, Toronto, ON.
- O'Reilly, T. (2005). *What is Web 2.0? Design patterns and business models for the next generation of software*. Retrieved from: <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- Pearl, J. (2009). *Causality: Models, reasoning and inference*. Cambridge: Cambridge University Press.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.
- Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). *Tuned models of peer assessment in MOOCs*. Paper presented at the *Educational Data Mining 2013*, Memphis, TN.
- Piety, P. J. (2013). *Assessing the big data movement*. New York, NY: Teachers College Press.
- Podesta, J., Pritzker, P., Moniz, E., Holdern, J., & Zients, J. (2014). *Big data: Seizing opportunities, preserving values*. Washington, DC: Executive Office of the President. Retrieved from: https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf
- Renear, A. H., & Palmer, C. L. (2009). Strategic reading, ontologies, and the future of scientific publishing. *Science*, 325, 828–832.
- Rupp, A. A., Nugent, R., & Nelson, B. (2012). Evidence-centered design for diagnostic assessment within digital learning environments: Integrating modern psychometrics and educational data mining. *Journal of Educational Data Mining*, 4, 1–10.
- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. Washington, DC: American Educational Research Association.
- Schoenfeld, A. H. (2006). Design experiments. In J. L. Green, G. Camilli, P. B. Elmore, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 193–205). Mahwah, NJ: Lawrence Erlbaum.
- Shaw, J. (2014, September-October). Why “big data” is a big deal: Information science promises to change the world. *Harvard Magazine*.
- Shepard, L. (2008). Formative assessment: Caveat emperor. In C. A. Dwyer (Ed.), *The future of assessment* (pp. 279–304). Mahwah, NJ: Lawrence Erlbaum.

- Shreeves, S. L. (2013). The role of repositories in the future of the journal. In B. Cope & A. Phillips (Eds.), *The future of the academic journal*. Cambridge: Woodhead.
- Siemens, G., & Baker, R. S. J. D. (2013). Learning analytics and educational data mining: Towards communication and collaboration. In *Second Conference on Learning Analytics and Knowledge (LAK 2012)* (pp. 252–254). Vancouver, BC: ACM.
- Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., Duval, E., Verbert, K., & Baker, R. S. J. D. (2011). *Open learning analytics: An integrated and modularized platform*. Society for Learning Analytics Research.
- Snow, E., Varner, L., Russell, D., & McNamara, D. (2014). *Who's in control?: Categorizing nuanced patterns of behaviors within a game-based intelligent tutoring system* (pp. 185–192). Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), Indianapolis, IN.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. London: Department for International Development.
- Stober, T., & Hansmann, U. (2009). *Agile software development: Best practices for large software development projects*. Berlin: Springer.
- Strijbos, J.-W., & Sluismans, D. (2010). Unravelling peer assessment: Methodological, functional and conceptual developments. *Learning and Instruction*, 20, 265–269.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. New York, NY: Doubleday.
- Tomkin, J. H., & Charlevoix, D. (2014). *Do professors matter?: Using an a/b test to evaluate the impact of instructor involvement on MOOC student outcomes*. Proceedings of the First ACM Conference on Learning @ Scale, Atlanta, GA.
- Vempaty, A., Varshney, L. R., & Varshney, P. K. (2013). *Reliable classification by unreliable crowds* (pp. 5558–5562). Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver.
- Vojak, C., Kline, S., Cope, B., McCarthey, S., & Kalantzis, M. (2011). New spaces and old places: An analysis of writing assessment software. *Computers and Composition*, 28, 97–111.
- Weiner, N. (1948/1965). *Cybernetics, or the control and communication in the animal and the machine*. Cambridge, MA: MIT Press.
- West, D. M. (2012). *Big data for education: Data mining, data analytics, and web dashboards*. Washington, DC: Brookings Institution.
- Wichowski, A. (2009). Survival of the fittest tag: Folksonomies, findability, and the evolution of information organization. *First Monday*, 14. Retrieved from: <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2447/2175>
- Wickett, K. M., Sacchi, S., Dubin, D., & Renear, A. H. (2012). *Identifying content and levels of representation in scientific data*. Baltimore, MD: American Society for Information Science and Technology.
- Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- Winne, P. H. (2006). How software technologies can improve research on learning and bolster school reform. *Educational Psychologist*, 41, 5–17.
- Winne, P. H., & Baker, R. S. J. D. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, 5, 1–8.
- Wolf, M. A. (2010). *Innovate to educate: System [re]design for personalized learning, a report from the 2010 symposium*. Washington, DC: Software and Information Industry Association.
- Woolf, B. P. (2010). *A roadmap for education technology*. Düsseldorf: GROE.
- Woolf, B. P., Lane, H. C., Chaudhri, V. K., & Kolodner, J. L. (2013). AI grand challenges for education. *AI Magazine*, 34, 66–84.

Notes on Contributors

Bill Cope is Professor in the Department of Educational Policy Studies at the University of Illinois, Urbana-Champaign. He is Principal Investigator in a series of projects funded by the Institute of Educational Sciences in the US Department of Education and the Bill and Melinda Gates Foundation researching and developing multimodal writing and assessment spaces. From 2010–2013 he was Chair of the Journals Publication Committee of the American Educational Research Association. Recent books include *The Future of the Academic Journal* (with Angus Phillips, eds), Elsevier, Oxford, 2009 (second edition 2014), and *Towards a Semantic Web: Connecting Knowledge in Academic Research* (with Kalantzis and Magee), Elsevier, Oxford, 2010. Email: billcope@illinois.edu

Mary Kalantzis is Dean of the College of Education at the University of Illinois, Urbana-Champaign. She was formerly Dean of the Faculty of Education, Language and Community Services at RMIT University in Melbourne, Australia, and President of the Australian Council of Deans of Education. With Bill Cope, she is co-author of *New Learning: Elements of a Science of Education*, Cambridge University Press, 2008 (second edition 2012); *Literacies*, Cambridge University Press, 2012 (second edition 2015); and *Making Sense: A Grammar of Multimodality*, Cambridge University Press (forthcoming). Also with Bill Cope, she is co-editor of *Ubiquitous Learning*, University of Illinois Press, 2009. Email: kalantzi@illinois.edu