



## Sources of Evidence-of-Learning: Learning and assessment in the era of big data

Bill Cope & Mary Kalantzis

To cite this article: Bill Cope & Mary Kalantzis (2015) Sources of Evidence-of-Learning: Learning and assessment in the era of big data, Open Review of Educational Research, 2:1, 194-217, DOI: [10.1080/23265507.2015.1074869](https://doi.org/10.1080/23265507.2015.1074869)

To link to this article: <http://dx.doi.org/10.1080/23265507.2015.1074869>



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



Published online: 12 Sep 2015.



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



Article views: 2



View related articles [↗](#)



View Crossmark data [↗](#)

# Sources of Evidence-of-Learning: Learning and assessment in the era of big data

BILL COPE & MARY KALANTZIS

*University of Illinois, Champaign, IL*

## Abstract

*This article sets out to explore a shift in the sources of evidence-of-learning in the era of networked computing. One of the key features of recent developments has been popularly characterized as ‘big data’. We begin by examining, in general terms, the frame of reference of contemporary debates on machine intelligence and the role of machines in supporting and extending human intelligence. We go on to explore three kinds of application of computers to the task of providing evidence-of-learning to students and teachers: (1) the mechanization of tests—for instance, computer adaptive testing, and automated essay grading; (2) data mining of unstructured data—for instance, the texts of student interaction with digital artifacts, textual interactions with each other, and body sensors; (3) the design and analysis of mechanisms for the collection and analysis of structured data embedded within the learning process—for instance, in learning management systems, intelligent tutors, and simulations. A consequence of each and all of these developments is the potential to record and analyze the ‘big data’ that is generated. The article presents both an optimistic view of what may be possible as these technologies and pedagogies evolve, while offering cautionary warnings about associated dangers.*

Keywords: big data, assessment, machine learning, learning analytics, educational data mining

## Introduction

Since the development of the institution of modern education, the primary site for gathering evidence-of-learning has been the test. The developments that we describe in this article transform the traditional artifact of the test in contradictory ways. On the one hand, computer-based, web-enabled testing mechanizes traditional assessment, embodying its logic, intensifying its historical functions, and extending its social effects. On the other hand, when curriculum content and student learning activities are delivered through the medium of networked computers, there arise new potentials to mine the unstructured ‘data exhaust’ emanating from these activities. It is also possible to design-in

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

This is an Open Access article distributed under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

processes for the collection of structured data. These latter developments suggest a shift in the primary site for the source of evidence-of-learning from the test to the learning process itself. Perhaps even, some day in the not too distant future, we will no longer need tests because the data that we have drawn from the learning process is itself sufficient evidence.

This article has several different agendas. It aims to be descriptive and analytical, developing a typology of educational data in the era of big data. It also represents a kind of advocacy, proposing that we explore affordances in emerging educational technologies and new ways of collecting and analyzing evidence-of-learning that might advance longstanding educational objectives. How might these technologies help us address ‘wicked problems’ where we educators have perennially failed, such as to reduce the levels of educational inequality, provide appropriate and engaging experiences for diverse populations of learners, and offer educational experiences which align in a more relevant way to people’s contemporary working, public and everyday cultural lives? This is also a moment in education when novel questions are being raised in novel sites and modes of pedagogical practice. These include, to name just a few trends, the rise of home and online schooling, massively open online courses (MOOCs), and the informal and semi-formal learning now more pervasively embedded in software devices and computer-mediated workplace training. However, we also want to avoid the implication that computers are the next cure-all in education—where, to use Cuban’s (2001) words, we often discover that that are ‘oversold and underused’. And indeed, the processes of applying ‘big data’ to education may raise a whole new series of problems, intensifying old problems or creating new ones. Or they may be used simply to replicate old pedagogies and unchanged educational and social outcomes.

### **Machine Intelligence Meets Human Intelligence**

‘Big data’, announces the title of a book written by two authors from the Oxford Internet Institute, represents ‘a revolution that will transform how we live, work and think’ (Mayer-Schönberger & Cukier, 2013). Sciences of direct observation are now being complemented or even at times replacing sciences where observations are mediated by data—in astronomy or genomic medicine, for instance. In our everyday lives, data is pervasively collected from our every computer-mediated interaction including our web searches and the content of our emails, in order to serve us targeted advertising. Recommendation systems use historical sales data to suggest that people who must surely be like you because they purchased the same thing as you, purchased something else that you might also want. Our phones ‘know’ our every contact and every geolocation, and such information may prove of interest to the police. These big data can be put to a myriad of uses. In the words of the subtitle of a White House report, we can use big data for the purposes of ‘seizing opportunities’ and ‘preserving values’ (Podesta, Pritzker, Moniz, Holdern, & Zients, 2014). The ‘preserving values’ part directs our attention to the dangers of big data, including discriminatory profiling, ‘filter bubbles’ and compromised privacy.

What then, is the nature of big data? What in general terms are the potentials of machine intelligence and machine-augmented human intelligence? And what are the implications for education?

*Informationalizing the Social*

The phrase ‘big data’ captures the idea that 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 geolocated. By ‘informationalized’ we mean, our social interactions are created and transmitted through digital information platforms, which (and this is the decisive factor) 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 (Kalantzis-Cope, 2010). We users care for the recording less than they do, at least in the longer term; in fact we mostly do not need the recordings beyond the moment of communicative interchange. This is how 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 one million customer transactions per day. Google processes 20 petabytes of data per day. Two hundred and fifty 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 *et al.*, 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.

*Big Data in Education*

The issues raised by the use of computers in education are by no means new. The project of using computers-in-education is now five decades old, beginning perhaps 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 foreshadowed 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, pp. 99, 109–110). The book has a future-inspiring photograph of such a machine—not yet an electronic but mechanical, like the adding machines of the time.

More than half a century after PLATO, two developments in computing 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 significance of ‘cloud computing’ (Erl, Puttini, & 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), one aspect of which is massively integrated social media. This turns data that was previously socially inscrutable, into socially scrutable 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 (DiCerbo & Behrens, 2014; Piety, 2013; West, 2012). It also represents a challenge. Though there is a mass of data, it is untidy, inconsistent and often hard to read.

So, to define our key phrase: 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.

### *Old Research Questions and New*

Big data may be able to help us to address 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. Moreover, the very socio-technical conditions that have made big data possible, are also 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.

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 4498 studies and involving four 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, pp. 220–221). Warschauer and Matuchniak (2010) similarly conclude that technology use in school has not been proven to improve student outcomes, though different kinds of pedagogical applications of technology do. More recently, in a review of technology integration of schools, Davies and West (2014, p. 841) conclude that although ‘students ... use technology to gather, organize, analyze, and report information, ... this has not dramatically improved student performance on standardized tests’. 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.

Specific developments in new fields of educational innovation also offer both challenges and opportunities for gathering evidence-of-learning. 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 *et al.*, 2014). We have witnessed the emergence of massive, open, free educational offerings, such as MOOCs (DeBoer, Ho, Stump, & Breslow, 2014; Peters & Britez, 2008). These phenomena are intensified by the processes of blended and ubiquitous learning (Cope & Kalantzis, 2009), and the blurring of pedagogies for in-person and remote learning interactions, such as the ‘flipped classroom’ (Bishop & Verleger, 2013). These developments are widely regarded to be transformative, or potentially transformative.

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, McCarthy, Vojak, & Kline, 2011; DiCerbo & Behrens, 2014; Mislevy, Almond, & Lukas, 2004; Quellmalz & Pellegrino, 2009); and adaptive, personalized or differentiated instruction which calibrates learning to individual needs (Conati & Kardan, 2013; Shute & Zapata-Rivera, 2012; Walkington, 2013; Wolf, 2010). Such models and processes of instructional delivery are variously labeled ‘constructivist’, ‘connectivist’ or ‘reflexive’ (Kalantzis & 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.

### *Machine-augmented Learning and Machine Learning*

To what extent and in what ways can machines support learning? Since the beginning of computing, questions have been raised about the intelligence of the machines and their

potential role of supporting human learning. Can they be helpful to learners because they can be smart like a teacher? Can they give feedback the way a teacher would? These questions require a response that addresses the larger question of the nature of ‘artificial intelligence’ (AI). And can computers learn from their own data, in order to help human learning? This question addresses a phenomenon now framed as ‘machine learning’.

The question of AI was famously posed in the form of a test by one of the founders of digital computing, Alan Turing. In his proposed test, a computer and a person is each hidden behind a screen, and another person is asking them questions via a teletype machine so the source of the answers is indistinguishable. If the person asking the questions cannot tell the difference between a human and a machine response to a question, then the machine may be taken to exhibit AI.

Digital computers can be constructed, and indeed have been constructed ... that ... can in fact mimic the actions of a human computer very closely. ... If one wants to make a machine mimic the behaviour of the human computer in some complex operation one has to ask him how it is done, and then translate the answer into the form of an instruction table. Constructing instruction tables is usually described as ‘programming.’ To ‘programme a machine to carry out the operation A’ means to put the appropriate instruction table into the machine so that it will do A. (Turing, 1950)

John Searle’s response is that this is not AI at all, or at least not something that the proponents of what he labels a ‘strong AI’ hypothesis would claim. Computers cannot think, he says. He sets out to disprove the Turing thesis with a hypothetical Chinese room. Behind the screen is a person who knows Chinese and a computer that can give the correct answer to the meaning of the Chinese character by using look-up tables. Just because the answer is correct, does not mean that the computer understands Chinese (Searle, 1980). His conclusion: ‘in the literal, real, observer-independent sense in which humans compute, mechanical computers do not compute. They go through a set of transitions in electronic states that we can interpret computationally’ (Searle, 2014).

To Searle, Stevan Harnad responds that computers set out to do no more than simulate or model patterns in the world in a way that records and represents, in increasingly smart ways, human knowledge and pattern-making (Harnad, 1989). Daniel Dennett says that Searle over-interprets and so misunderstands the Turing test, which is not about human-like thinking, but thinking which can be made to seem human-like, hence the aura of façade in all AI programs (Dennett, 1998). On our iPhones, Siri seems smart because, if our question is sufficiently predictable, she seems to recognize what we are asking her. She can look things up on the internet to give you an answer, and she can calculate things. Siri is in fact a recording of the voice of a woman named Susan Bennett who lives in Atlanta, Georgia. So the intelligence of our devices is no more than what Turing calls an ‘imitation game’, a piece of trickery by means of which we anthropomorphize the machine. How smart is the machine? Siri is smart enough to have a recording of a huge amount of human data at her finger tips per medium of the internet—the address of a restaurant, what the weather is forecast to be at 3:00 p.m., the year of the French Revolution. She has been programmed to answer many kinds of question and to look up many kinds of things and make numerical calculations. She has also been programmed to give

funny answers to insults. This is how our phones pass the Turing test every day if the test is seen to be no more than an imitation game. Computers fail this test to the extent that passing the test is a failure of credulity.

Computers in education, can be this smart: they can assess whether an answer to a question is right or wrong, because they have been programmed to ‘know’ the answer (for instance, computerized, select response tests). They can anticipate a range of sequences of learning and feedback which have been programmed into them (for instance, in an intelligent tutor). They can grade a text by comparing it with other texts that have already been graded by humans (automated essay scoring). They are alternately not very smart and very smart. They are not very smart in the sense that they can only repeat answers and sequences that have already been programmed into them by humans. But they are smart to the extent that they record and deliver already-recorded information and patterns of human action.

In this regard, computers are merely an extension of the knowledge architectures of writing. They make them more efficient by mechanizing parts of the reading path. Instead of looking up an answer to check whether it is correct (via tables of contents, indexes or lists of answers to questions in a book, printed upside down to help you resist the temptation of looking up the answer prematurely), they look it up for you—as long as the question and answer have been written up by humans in advance. Just as a thousand people could read a book, a thousand people can have an answer asked of them by the computer, and their answer checked. The machine is smart because it has been programmed by a human, in much the same way that a book is smart because it has been written by a human. Books are smart, and computers are smart in the same way—only mechanically more efficient. The smartest thing about computers, perhaps, is their ‘imitation game’, the way they have been designed to seem smart. It is not just the anthropomorphized interfaces (Siri’s voice or responsive avatars), but the text on the screen that says things to the learner, or tells them with a nicely presented alert when their answers are right or wrong.

Computers can also be intelligent in another way. Not only repeating things they have been programmed to say, to a limited degree they can also learn. This aspect of AI is called ‘machine learning’. This means that the machine can learn from the data that it is presented. In fact, this is a process of finding statistical patterns in meanings that can be represented using numerical tokens. In supervised machine learning, computer interaction (an action sequence in an intelligent tutor, the words typed in an essay) that has been attributed a value by a human (an affective judgment such as ‘engagement’, or a grade) is compared statistically with a new pattern of interaction or text. If similar, the human judgment is attributed by the machine to the new activity or text. Unsupervised machine learning presents the person with statistically significant patterns, and asks them to attribute a meaning. The machine may then record the conclusion, thereby adding to its apparent smarts. But the computer is still essentially no more than a recording machine and a calculating machine. However, as we will see in the sections that follow, these mere recording and calculating machines might be very helpful to learners to the extent that they record and calculate data types that have been construed by humans to be evidence-of-learning.

The typology that follows in the next sections of this article are designed to capture the principal educational data types: (1) mechanized tests; (2) minable unstructured data; (3)



structured data where learning analytics tools have been designed into the learning activities. 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. The experience of a whole day in a school with one-to-one computer access may draw learners past quite a few of these data types.

### The Mechanization of Tests

Traditionally, tests have taken two main forms: select-response tests, and supply response tests. Select response tests can be mechanized and enhanced with a new generation of survey-psychometric systems. Supply response assessments can be mechanized with automated essay scoring using natural language processing technologies.

#### *Survey-psychometric Systems*

Frequently located within or alongside learning management systems, selected response assessments rely upon long-established traditions of what we would term survey psychometrics. We use this phrase to describe a method of asking symptomatic survey questions aimed to elicit the extent of a student's knowledge of a topic. If we ask you a certain number of questions about a topic, and you get a certain number of them right, we can come to the conclusion that you understand this topic. The general 'assessment argument' (Pellegrino, Chudowsky, & 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 (CATs) tailor the test to the trait level of the person taking the test. They differentiate subset areas of knowledge within a test (Chang, 2012, 2015). 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 like Knewton, who are working with publishers to embed 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 *et al.*, 2013a). 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, & Shapira, 2014).

#### *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 & Chodorow, 2003; Chung & Baker, 2003; Cotos & Pendar, 2007; Shermis, 2014). Thus far, these

technologies have been less successful in providing meaningful feedback to writers beyond the mechanics of spelling in grammar (McNamara, Graesser, McCarthy, & Cai, 2014; Vojak, Kline, Cope, McCarthy, & Kalantzis, 2011; Warschauer & Grimes, 2008), although increasingly sophisticated technologies for formative assessment of writing are in development (Cope & Kalantzis, 2013; Roscoe & McNamara, 2013). 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) (Landauer, McNamara, Dennis, & Kintsch, 2007; McNamara *et al.*, 2014; Vojak *et al.*, 2011). Other areas of development include analyses of conceptual structures or topic models in written texts (Paul & Girju, 2010), argument (Ascaniis, 2012), and sentiment analysis, in discussion forums for instance (Wen, Yang, & Rose, 2014).

### **Unstructured Data and Educational Data Mining**

To mechanize the test is to make more efficient the canonical processes for eliciting evidence-of-learning. This may well generate big data in the form of answers by large numbers of students and to innumerable questions, or huge corpora of student texts which can be used via machine learning techniques to improve automated essay scoring. However, the remaining two of our three data types are endemic to the regime of big data: the analysis of unstructured data where there is no predefined data model; and the generation of structured data embedded within learning.

Students work in a learning management system; they write their assignments in an online writing and assessment space; they play an educational game; they interact with each other in discussion boards. Along the way, they leave trace data that is recorded in log files: keystrokes, edit histories, time stamps, clickstreams, web navigation paths, social network interactions, or eye tracking and body movement data. The subdiscipline of educational data mining has emerged, one of whose primary aims is to mine this ‘data exhaust’, with a view to separating the signal from the noise (Baker & Siemens, 2014; Castro, Vellido, Nebot, & Mugica 2007; Siemens & Baker, 2013). It aims to ask questions such as, what patterns of action visible in the data predict relative success or failure?

Here we will take several potential sources of unstructured data: learning games, social interaction analyses, affect meters, and body sensors.

#### *Learning Games*

This computer game 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, because now 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 today reach an audience larger than Hollywood. The question educators have been attempting to address is how can we bring

the medium and levels of engagement of games to education through the ‘gamification’ of learning (Gee, 2005; Magnifico, Olmanson, & Cope, 2013)? 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). From the perspective of educational data mining, each one of these principles reflects a possible moment of machine-mediated reflexive data collection; each is a recordable teaching moment—either in the form of the unstructured data generated, or records of learning that might also be structured into the game. 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 (Gee, 2003, pp. 203–210). 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 *et al.*, 2014).

### *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 *et al.*, 2014; Wise, Zhao, & Hausknecht 2013). Natural language processing methods can be used to parse the content of interactions (Xu, Murray, Woolf, & 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, & Cannavacciuolo, 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 & Vatrappu, 2011; Perera, Kay, Koprinska, Yacef, & Zaiane, 2009). The processes of collecting and analyzing such data have been termed ‘social learning analytics’ (Ferguson & Shum, 2012).

### *Affect Meters*

Motivation and affect are key factors in learning (Magnifico *et al.*, 2013). Log files can 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, & Benitez-Guerrero, 2013). In the context of social learning, web reputation technologies (Farmer & 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, & Henkel, 2003). Computer-mediated learning environments can monitor student sentiments with affect meters of one kind or another, collecting structured as well as unstructured data: emotive-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, & Graesser, 2010; Chung, 2013; D’Mello, 2013; Fancsali, Ritter, Stamper, & Berman, 2014; Winne & Baker, 2013; Wixon *et al.*, 2014).

### *Body Sensors*

Body sensors can also be 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 *et al.*, 2010; Grafsgaard, Wiggins, Boyer, Wiebe, & Lester, 2014; Schneider & Pea, 2014; Vatrapu, Reimann, Bull, & Johnson, 2013). Student movement not connected with screen presentations include wearable technologies such as bracelets (Woolf, 2010, p. 19), radio-frequency identification (RFID) chips in student identification (ID) cards (Kravets, 2012), group interactions in multi-tabletop environments (Martinez-Maldonado, Yacef, & 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 & Johnson-Glenberg, 2013).

### **Structured Data, Embedded Data Collection and Learning Analytics**

Both computerized testing and educational data mining are primarily designed to provide evidence-of-learning after the fact. However, computer mediated learning environments can also have, embedded within them, structured data collection. This has become one of the main areas of interest for another subdiscipline of education, learning analytics (Bienkowski, Feng, & Means, 2012; Knight, Shum, & Littleton, 2013; Mislevy, Behrens, Dicerbo, & Levy, 2012; Siemens & Baker, 2013; West, 2012).

Embedding mechanisms to collect evidence-of-learning within computer mediated learning environments has a number of advantages. The data is self-describing, less open than mining of unstructured data to conjectures based on correlations of statistically identifiable patterns activity with judgments of ‘training data’ made by humans. It is semantically legible, which means that every constituent datapoint is explicable, for instance a qualitative comment by a peer against a review criterion, a language suggestion made by the machine, an annotation made by a peer, an answer to a question, or a step in a learning sequence in an intelligent tutor. Every datapoint has the potential to offer immediately actionable feedback, in other words formative assessment. When such modes of assessment are embedded into digital learning environments—learning management systems, intelligent tutors or writing environments, for instance—there is the potential to generate huge amounts of structured data offering evidence of learning for individuals and cohorts, in any moment as well as progress views over time.

Intelligent tutors, simulations, semantic mapping tools and learning management systems are all capable of incorporating embedded and formative assessment, and consequentially capturing structured data for evidence-of-learning.

### *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, & Graesser, 2013; Chaudhri *et al.*, 2013b; Vanlehn, 2006). 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, & McLaughlin, 2013; Woolf, 2010). 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 (Graesser, VanLehn, Rosé, Jordan, & Harter, 2001).

### *Simulations*

Simulations share a close family resemblance with games, however 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, little or no competitive scoring and open-ended outcomes (Sauvé, Renaud, Kaufman, & 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, & 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 & Dede, 2010).

### *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 & 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, 1978; Ausubel, Novak, & 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 *et al.*, 2013; Cañas *et al.*, 2004; Chang, Sung, & Lee, 2003; Kao, Chen, & Sun, 2010; Liu, 2002; Su & Wang, 2010; Tzeng, 2005). The authors of this article have developed one such tool, InfoWriter, in which learners during their writing highlight and diagram the ideas that the writing represents (Olmanson *et al.*, submitted for publication). It is also possible to machine-generate semantic maps from text using natural language processing methods (Girju, Badulescu, & Moldovan, 2006).

### *Analytics Tools in Learning Management Systems*

Comprehensive learning management systems—for instance Blackboard, Desire2Learn, Canvas and Coursera—have the potential to generate large amounts of both structured and unstructured data. 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 click-stream 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.

### **Changing the Sources of Evidence-of-Learning**

Computer-mediated learning environments have the potential to generate huge amounts of data, from varied data sources, and representing different data types. It is 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. We want to conclude now by highlighting some shifts that are relevant to the field of education in the era of big data: the sites of data collection, the size of the datapoints, the objects of measurement, the role of the machine, and the responsibilities of the data analyst.

### *The Sites of Data Collection*

As we have seen, the collection of evidence-of-learning can now be embedded within learning—to the extent even that we may have such a comprehensive data source that we no longer need conventional tests. Even the latest extensions of traditional testing processes—select and supply-response assessments—can be embedded within learning, so becoming a more constructive part of the learning process than when their view is purely retrospective. Select response assessments can and sometimes do also provide immediate answers. By this means, answering the question is at once a data collection point and a learning experience.

These embedded datapoints might be small—a tiny piece of feedback for a learner, or a decision point about the next step in learning for an individual student in a

personalized learning environment. Or they may be aggregated to a higher level for the purposes of learner profiling when teacher and student are working on the co-design of individual learning programs. Or at a whole class, cohort, specific demographic, or school level, they may be to generate data for the purposes of administrative accountability.

In this regard, learning-embedded assessment may spell the practical end of the distinction between formative and summative assessment (Cope & Kalantzis, *in press*). Formative assessment is assessment during and *for* learning, providing feedback to learners and their teachers which enhances their learning. Summative assessment is retrospective assessment *of* learning, typically a test at the end of a unit of work, a period of time, or a component of a program. This distinction was first named by Michael Scriven in 1967 to describe educational evaluation, then applied by Benjamin Bloom to assessment of learning (Airasian, Bloom, & Carroll, 1971; Bloom, 1968). The subsequent literature on formative assessment has consistently argued for its effectiveness (Baker, 2007; Bass & Glaser, 2004; Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006; OECD Centre for Educational Research and Innovation, 2005; Pellegrino *et al.*, 2001; Shepard, 2008; Wiliam, 2011). There have also been frequent laments that formative assessment has been neglected in the face of the rise of standardized, summative assessments as an instrument of institutional accountability (Armour-Thomas & Gordon, 2013; Gorin, 2013; Kaestle, 2013; Ryan & Shepard, 2008).

However, the modes of embedded assessment enabled by computer-mediated learning, may reverse this imbalance (Behrens & DiCerbo, 2013; Knight, Shum, & Littleton, 2013; Pea, 2014). Indeed, it is conceivable that summative assessments could be abandoned, and even the distinction between formative and summative assessment. In a situation where data collection has been embedded within the learner's workspace, 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. Learning analytic and data mining processes can be used to produce progress generalizations at different levels of granularity. At the same time, it may always be possible to drill down to specific programs, learners, all the way to every and any of the datapoints upon which these generalizations are based.

Under these conditions of learning, all our assessment would be formative, and summative assessment simply a retrospective perspective on the same data. Such learning environments, where the distinctions between instruction and assessment are so blurred (Armour-Thomas & Gordon, 2013), might require that we move away from the old assessment terminology, with all its connotative baggage. Perhaps a notion of 'reflexive pedagogy' might replace the traditional instruction/assessment dualism. And, instead of formative and summative assessment as different collection modes, designed differently for different purposes, we may 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 a prospective/constructive rather than retrospective/judgmental, though the learning objective remains the same (Cope & Kalantzis, 2013).

*The Size of the Datapoints*

Just as significant as the bigness of big educational data is the smallness of its constituent datapoints. Indeed, this 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 is 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.

*The Object of Measurement*

The evidentiary focus of traditional assessment, and also many of the new generation of learning analytics tools, is that elusive object, cognition—the ‘theta’ of latent cognitive traits in item response theory (Mislevy, 2013), or the ‘g’ of intelligence in IQ tests. Classical testing logic runs along these lines: cognition developed in learning => observation in a test => interpretation of the test results as evidence of cognition (Pellegrino *et al.*, 2001). The test was a separate object, located after learning and supporting a retrospective interpretation. However, when the focus is on knowledge artifacts, we have direct observation of disciplinary knowledge practice as-it-happens. Knowledge is assessable in the form of its representation in the artifacts of disciplinary practice and the processes of their construction (Knight *et al.*, 2013). Now we have the basis for a less mediated interpretation of learning.

As a consequence, in the era of digital we do not need to be so conjectural in our evidentiary arguments. We do not need to look for anything latent when we have captured so much evidence in readily analyzable form about the concrete products of complex knowledge work, as well as a record of all the steps undertaken in the creation of these products. The focus of our attention to evidence-of-learning in the era of machine-mediated learning can now turn to authentic knowledge artifacts, and the running record that learners create in their practice of the discipline as they create these artifacts. Our focus for analysis, in other words, is not on things that students can think, but the knowledge representations that they make. These artifacts constitute evidence of complex epistemic performance—a report on a 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 documented invention or design of an object, a worked mathematical or statistical example, a field study report, or executable computer code with user stories. 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, and navigation paths through sources, to name a few. In these ways, the range of collectable data surrounding the knowledge work is hugely expanded.

Raising our evidentiary expectations, how then can educators come to conclusions about dimensions of learning as complex as mastery of disciplinary practices, complex epistemic performances, collaborative knowledge work and multimodal knowledge representations (Behrens & DiCerbo, 2013; Berland, Baker, & Blikstein, 2014; DiCerbo &



Behrens, 2014; Winne, 2014)? The answer, we suggest, lies in a shift to richer data environments and more sophisticated analytical tools, many of which can be pre-emptively designed into the learning environment itself, a process of ‘evidence-centered design’ (Mislevy *et al.*, 2012; Rupp, Nugent, & Nelson, 2012).

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 inferences to tangible knowledge representations. We might name as a shift in focus from the cognitive to the ‘artifactual’. Where the cognitive can be no more than putative knowledge, the artifactual is a concretely represented knowledge and its antecedent knowledge processes. We call this ‘ergative’ or work-oriented pedagogy and assessment, in contrast to constructivist or cognitively-oriented pedagogy and assessment (Cope & Kalantzis, *in press*).

### *The Role of the Machine*

Big data is not simply machine generated, despite the apparent intelligence expressed through the anthropomorphized user interfaces. Having dismissed the Turing test, Searle concedes to the helpful pragmatics of a ‘weak AI’ theses. The computer is an artifice of human communication, an extension of older textual architectures, and significant extension nevertheless, if not the qualitative leap that its most enthusiastic advocates would have us believe. It is a cognitive prosthesis, an extension of the social mind, and as such part of a continuous history that begins with writing and later becomes books and schools. To give an example, in the era of big data, human intelligence can be magnified through the collection and calculation of ‘crowdsourced’ (Surowiecki, 2004) human judgments. 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). The machines seem intelligent, and they are to the extent that they have collected and calculated the significance of a lot of human intelligence, just as books and libraries and teachers have in the past—except using more data than human teachers or learners could ever have done. Computers are only smart to the extent that they are machines that record and externalize human thinking.

### *The Responsibilities of the Data Analyst*

Everyone is a data analyst now. The teacher accesses the data to know their learners and recalibrate their teaching. In this evidentiary environment, the teacher can be, should be,

positioned as researcher. This may demand of them a new kind of data literacy. As these platforms and environments all generate large amounts of data, what follows are expectations that teachers become data literate (Twidale, Blake, & Gant, 2013), in support of processes of evidence-based decision-making (Mandinach & Gummer, 2013). However, as much as anything it is the responsibility of software engineers and user interface designers to create environments where it is not necessary to understand obscure statistical formulae because salient information about learning is presented in elegant visualizations, and because it is always possible to drill down to retrace specific learning sequences (Worsley & Blikstein, 2014). This data can also 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. Moreover, this is the same embedded data a researcher can use. 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.

### Conclusions

Big data has come to school. Or at least if the hardware has not yet been delivered, it is on its way and the rudiments of the software can be seen in the range of new educational technologies and pedagogical practices described in this article. However, we have a long way to go to develop software and pedagogies 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<sup>1</sup> 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 has been a big learning journey, and we have barely begun, but that is another story (Cope & Kalantzis, 2013).

There are also potential dangers in the world of big data. Some educational technologies may intensify and ossify old didactic pedagogies. Testing technologies may further institutionalize the mechanisms of social inequality written into norm-based assessment, no matter how efficient they now are (Cope & Kalantzis, *in press*). We also 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 (Podesta *et al.*, 2014).

### Disclosure statement

No potential conflict of interest was reported by the author(s).

### 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

- Airasian, P. W., Bloom, B. S., & Carroll, J. B. (1971). *Mastery learning: Theory and practice*. New York: Holt Rinehart & Winston.
- Aleven, V., Beal, C. R., & Graesser, A. C. (2013). Introduction to the special issue on advanced learning technologies. *Journal of Educational Psychology, 105*, 929–931.
- Armour-Thomas, E., & Gordon, E. W. (2013). *Toward an understanding of assessment as a dynamic component of pedagogy*. Princeton, NJ: The Gordon Commission.
- Ascaniis, S. D. (2012). Criteria for designing and evaluating argument diagramming tools from the point of view of argumentation theory. In N. Pinkwart & B. M. McLaren (Eds.), *Educational technologies for teaching argumentation skills*. Sharjah: Bentham Science Publishers.
- Ausubel, D. (1963). *The psychology of meaningful verbal learning*. New York: Grune & Stratton.
- Ausubel, D. (1978). In defense of advance organizers: A reply to the critics. *Review of Educational Research, 48*, 251–257.
- Ausubel, D., Novak, J., & Hanesian, H. (1978). *Educational psychology: A cognitive view*. New York: Holt, Rinehart & Winston.
- Baker, E. L. (2007). *Moving to the next generation system design: Integrating cognition, assessment, and learning*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California.
- Baker, R. S. J. D., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*, 223–241.
- 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.
- Bass, K. M., & Glaser, R. (2004). *Developing assessments to inform teaching and learning*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.
- 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, Matthew, Ryan S. Baker, & Paulo Blikstein. (2014). Educational Data Mining and Learning Analytics: Applications to Constructionist Research. *Technology, Knowledge and Learning, 19*, 205–220.
- 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.
- Bishop, J., & Verleger, M. (2013). *The flipped classroom: A survey of the research*. Atlanta, GA: American Society for Engineering Education.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7–74.
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 1–12.
- Blumschein, P., Hung, W., Jonassen, D., & Strobel, J. (2009). *Model-based approaches to learning: Using systems models and simulations to improve understanding and problem solving in complex domains*. Rotterdam: Sense.

- Bredeweg, B., Liem, J., Beek, W., Linnebank, F., Gracia, J., Lozano, E., ... Mioduser, D. (2013). DynaLearn – An intelligent learning environment for learning conceptual knowledge. *AI Magazine*, 34, 46–65.
- Bull, S., & Vatrappu, R. (2011). *Supporting collaborative interaction with open learner models: Existing approaches and open questions* (pp. 761–765). Paper presented at the 9th International Conference on Computer-Supported Collaborative Learning 2011, Hong Kong.
- Burstein, J., & Chodorow, M. (2003). Directions in automated essay scoring. In R. Kaplan (Ed.), *Handbook of applied linguistics*. New York, NY: Oxford University Press.
- Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., Gómez, G., ... Carvajal, R. (2004). *CMAP-TOOLS: A knowledge modeling and sharing environment. in concept maps: Theory, methodology, technology*. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Proceedings of the First International Conference on Concept Mapping*, Pamplona, Spain.
- Castro, F., Vellido, A., Nebot, À., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. In *Studies in Computational Intelligence* (pp. 183–221). Berlin: Springer-Verlag.
- Chang, H.-H. (2012). Making computerized adaptive testing diagnostic tools for schools. In R. W. Lissitz and H. Jiao (Eds.), *Computers and their impact on state assessment: Recent history and predictions for the future* (pp. 195–226). Charlotte, NC: Information Age Publishing.
- Chang, H.-H. (2015). Psychometrics behind computerized adaptive testing. *Psychometrika*, 80(1), 1–20.
- Chang, K. E., Sung, Y. T., & Lee, C. L. (2003). Web-based collaborative inquiry learning. *Journal of Computer Assisted Learning*, 19, 56–69.
- Chaudhri, V. K., Cheng, B. H., Overholtzer, A., Roschelle, J., Spaulding, A., Clark, P., ... Gunning, D. (2013a). ‘Inquire biology’: A textbook that answers questions. *AI Magazine*, 34, 55–72.
- Chaudhri, V. K., Gunning, D., Lane, H. C., & Roschelle, J. (2013b). Intelligent learning technologies: Applications of artificial intelligence to contemporary and emerging educational challenges. *AI Magazine*, 34, 10–12.
- Chung, G. K. W. K. (2013). *Toward the relational management of educational measurement data*. Princeton, NJ: The Gordon Commission.
- Chung, G. K. W. K., & Baker, E. L. (2003). Issues in the reliability and validity of automated scoring of constructed responses. In M. D. Shermis & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary assessment* (pp. 23–40). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education*, 42, 309–328.
- 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., & Kalantzis, M. (2009). Ubiquitous learning: An agenda for educational transformation. In B. Cope & M. Kalantzis (Eds.), *Ubiquitous learning*. Champaign, IL: University of Illinois Press.
- Cope, B., & Kalantzis, M. (2013). Towards a new learning: The ‘Scholar’ social knowledge workspace, in theory and practice. *e-Learning and Digital Media*, 10, 334–358.
- Cope, B., & Kalantzis, M. (in press). Assessment and pedagogy in the era of machine-mediated learning. In T. Dragonas, K. J. Gergen, & S. McNamee (Eds.), *Education as social construction: Contributions to theory, research, and practice*. Chagrin Falls, OH: Worldshare Books.
- Cope, B., Kalantzis, M., McCarthey, S., Vojak, C., & Kline, S. (2011). Technology-mediated writing assessments: Paradigms and principles. *Computers and Composition*, 28, 79–96.
- Cotos, E., & Pendar, N. (2007). Automated diagnostic writing tests: Why? How? In *Technology for Second Language Learning Conference* (pp. 1–15). Ames, IA: Iowa State University.
- Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.
- D’Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology*, 105, 1082–1099.

- D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., ... Graesser, A. (2010). *A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning*. Paper presented at the Intelligent Tutoring Systems: 10th International Conference, ITS 2010, Pittsburgh, PA, USA, June 14–18, 2010, In V. Alevan, J. Kay, & J. Mostow (eds), *Proceedings, Part I*. Berlin: Springer.
- Davies, R. S., & West, R. E. (2014). Technology integration in schools. in J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (eds), *Handbook of research on educational communications and technology* (pp. 841–853). New York: Springer.
- DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing 'Course': Reconceptualizing educational variables for massive open online courses. *Educational Researcher*, 43, 74–84.
- Dennett, D. C. (1998). *Brainchildren: Essays on designing minds*. Cambridge, MA: MIT Press.
- DiCerbo, K. E., & Behrens, J. T. (2014). *Impacts of the digital ocean on education*. London: Pearson.
- Erl, T., Puttini, R., & Mahmood, Z. (2013). *Cloud computing: Concepts, technology and architecture*. Upper Saddle River, NJ: Prentice Hall.
- Espelage, D. L., Holt, M. K., & Henkel, R. R. (2003). Examination of peer-group contextual effects on aggression during early adolescence. *Child Development*, 74, 205–220.
- Fancsali, S. E., Ritter, S., Stamper, J. C., & Berman, S. (2014). Personalization, non-cognitive factors, and grain-size for measurement and analysis in intelligent tutoring systems. *Proceedings of the 2nd Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium* (pp. 124–135). Pittsburgh, PA: US Army Research Laboratory.
- Farmer, F. R., & Glass, B. (2010). *Web reputation systems*. Sebastapol, CA: O'Reilly.
- Ferguson, R., & Shum, S. B. (2012). Social learning analytics: Five approaches. In *Second Conference on Learning Analytics and Knowledge (LAK 2012)* (pp. 23–33). Vancouver, BC: ACM.
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.
- Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. London: Routledge.
- Gee, J. P. (2005). *Why video games are good for your soul: Pleasure and learning*. Melbourne: Common Ground.
- Girju, R., Badulescu, A., & Moldovan, D. (2006). Automatic discovery of part-whole relations. *Computational Linguistics*, 32, 83–135.
- Gorin, J. S. (2013). *Assessment as evidential reasoning*. Princeton, NJ: The Gordon Commission.
- Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, (Winter), 39–51.
- Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2014). *Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue* (pp. 122–129). Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), Indianapolis, IN.
- Harnad, S. (1989). Minds, machines and searle. *Journal of Theoretical and Experimental Artificial Intelligence*, 1, 5–25.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Kaestle, C. (2013). *Testing policy in the United States: A historical perspective*. Princeton, NJ: The Gordon Commission.
- Kalantzis, M., & Cope, B. (2012). *New learning: Elements of a science of education*. Cambridge: Cambridge University Press.
- Kalantzis-Cope, P. (2010). Whose property? Mapping intellectual property rights, contextualizing digital technology and framing social justice. In P. Kalantzis-Cope & K. Gherab-Martin (Eds.), *Emerging digital spaces in contemporary society: Properties of technology* (pp. 131–144). New York, NY: Palgrave Macmillan.
- Kao, G. Y.-M., Chen, K.-C., & Sun, C.-T. (2010). Using an e-learning system with integrated concept maps to improve conceptual understanding. *International Journal of Instructional Media*, 37, 151–161.

- 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., 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.
- Kravets, D. (2012, September 7). Tracking school children with RFID tags? It's all about the Benjamins. *Wired*.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. New York, NY: Routledge.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., ... Van Alstyne, M. (2009). Computational social science. *Science*, 323, 721–723.
- Liddo, A. D., Shum, S. B., Quinto, I., Bachler, M., & Cannavacciuolo, L. (2011). Discourse-centric learning analytics. In *Conference on Learning Analytics and Knowledge (LAK 2011)* (pp. 23–33). Banff, AB: ACM.
- Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*, 42, 445–452.
- Liu, E. Z. F. (2002). Incomas: An item bank based and networked concept map assessment system. *International Journal of Instructional Media*, 29, 325–335.
- 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.
- Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42, 30–37.
- Martinez-Maldonado, R., Yacef, K., & Kay, J. (2013). *Data mining in the classroom: Discovering groups' strategies at a multi-tabletop environment* (pp. 121–128). Paper presented at the International Conference on Educational Data Mining.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York, NY: Houghton Mifflin Harcourt.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-matrix*. New York, NY: Cambridge University Press.
- Mislevy, R. J. (2013). Postmodern test theory. In E. W. Gordon (Ed.), *The Gordon commission on the future of assessment in education: Technical report* (pp. 79–100). Princeton, NJ: The Gordon Commission.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). *A brief introduction to evidence-centered design*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California.
- 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., Oranje, A., Bauer, M. I., von Davier, A., Hao, J., Corrigan, S., ... John, M. (2014). *Psychometric considerations in game-based assessment*. Redwood City CA: GlassLab.
- Molnar, A., Rice, J. K., Huerta, L., Shafer, S. R., Barbour, M. K., Miron, G., ... Horvitz, B. (2014). *Virtual schools in the U.S. 2014: Politics performance policy*. Boulder, CO: National Education Policy Center.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199–218.
- Novak, J. D., & Cañas, A. J. (2008). *The theory underlying concept maps and how to construct and use them* (Technical Report IHMC CmapTools 2006-01 Rev 01-2008). Florida Institute for Human and Machine Cognition.
- 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>

- OECD Centre for Educational Research and Innovation. (2005). *Formative assessment: Improving learning in secondary classrooms*. Paris: Organization for Economic Cooperation and Development (OECD).
- Olmanson, J., Kennett, K., McCarthey, S., Searsmith, D., Cope, B., & Kalantzis, M. (submitted for publication). Visualizing revision: Between-draft diagramming in academic writing.
- Paul, M., & Girju, R. (2010). *A two-dimensional topic-aspect model for discovering multi-faceted topics*. Paper presented at the 24th AAAI Conference on Artificial Intelligence, Atlanta, GA.
- Pea, R. (2014). *The learning analytics workgroup: A report on building the field of learning analytics for personalized learning at scale*. Stanford, CA: Stanford University.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.
- Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, 21, 759–772.
- Peters, M. A., & Britez, R. G. (2008). *Open education and education for openness*. Rotterdam: Sense.
- 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](https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf)
- Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and testing. *Science*, 323, 75–79.
- Rebolledo-Mendez, G., Du Boulay, B., Luckin, R., & Benitez-Guerrero, E. I. (2013). Mining data from interactions with a motivational aware tutoring system using data visualization. *Journal of Educational Data Mining*, 5, 72–103.
- Roscoe, R. D., & McNamara, D. S. (2013). Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, 105, 1010–1025.
- 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.
- Ryan, K. E., & Shepard, L. A. (2008). *The future of test-based accountability*. New York, NY: Routledge.
- Sauvé, L., Renaud, L., Kaufman, D., & Marquis, J.-S. (2007). Distinguishing between games and simulations: A systematic review. *Educational Technology & Society* 10, 247–256.
- Schneider, B. & Pea, R. (2014). *The effect of mutual gaze perception on students' verbal coordination* (pp. 138–144). Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), Indianapolis, IN.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3, 417–457.
- Searle, J. R. (2014, October 9). What your computer can't know. *New York Review of Books*.
- Segal, A., Katzir, Z., Gal, Y. (Kobi), Shani, G., & Shapira, B. (2014). *EduRank: A collaborative filtering approach to personalization in e-learning* (pp. 68–75). Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), Indianapolis, IN.
- Shepard, L. (2008). Formative assessment: Caveat emperor. In C. A. Dwyer (Ed.), *The Future of Assessment* (pp. 279–304). Mahawah NJ: Lawrence Erlbaum.
- Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing*, 20, 53–76.
- Shute, V., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach & A. Lesgold (Eds.), *Adaptive technologies for training and education*. New York, NY: Cambridge University Press.
- Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2. Retrieved from: [http://www.ingedewaard.net/papers/connectivism/2005\\_siemens\\_ALearningTheoryForTheDigitalAge.pdf](http://www.ingedewaard.net/papers/connectivism/2005_siemens_ALearningTheoryForTheDigitalAge.pdf)

- 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.
- Skinner, B. F. (1954 (1960)). The science of learning and the art of teaching. In A. A. Lumsdaine & R. Glaser (Eds.), *Teaching Machines and Programmed Learning* (pp. 99–113). Washington, DC: National Education Association.
- Speck, J., Gualtieri, E., Naik, G., Nguyen, T., Cheung, K., Alexander, L., & Fenske, D. (2014). *ForumDash: Analyzing online discussion forums* (pp. 4–5). Proceedings of the First ACM Conference on Learning @ Scale, Atlanta, GA.
- Strijbos, J.-W. (2011). Assessment of (computer-supported) collaborative learning. *IEEE Transactions on Learning Technologies*, 4, 59–73.
- Su, C. Y., & Wang, T. I. (2010). Construction and analysis of educational assessments using knowledge maps with weight appraisal of concepts. *Computers & Education*, 55, 1300–1311.
- 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.
- Swan, M. (2012). Sensor mania! The internet of things, wearable computing, objective metrics, and the quantified self 2.0. *Journal Sensor and Actuator Networks*, 1, 217–253.
- Tergan, S.-O. (2005). Digital concept maps for managing knowledge and information. In S.-O. Tergan and T. Keller (Eds.), *Knowledge and Information Visualization* (pp. 185–204). Berlin: Springer-Verlag.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433–460.
- Twidale, M. B., Blake, C., & Gant, J. (2013). *Towards a data literate citizenry* (pp. 247–257). Paper presented at the iConference 2013.
- Tzeng, J. Y. (2005). Developing a computer-based customizable self-contained concept mapping for Taiwanese history education. In P. Kommers & G. Richards (Eds.), *Proceedings of World conference on educational multimedia, hypermedia and telecommunications* (pp. 4105–4111). Chesapeake, VA: AACE.
- Vanlehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16, 227–265.
- Vatrapu, R., Reimann, P., Bull, S., & Johnson, M. (2013). *An eye-tracking study of notational, informational, and emotional aspects of learning analytics representations* (pp. 125–134). Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13), New York, NY.
- Vojak, C., Kline, S., Cope, B., McCarthy, S., & Kalantzis, M. (2011). New spaces and old places: An analysis of writing assessment software. *Computers and Composition*, 28, 97–111.
- Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology*, 105, 932–945.
- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3, 22–36.
- Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education*, 34, 179–225.
- Waters, J. K. (2014). Adaptive learning: Are we there yet? *Technological Horizons in Education*, 41. Retrieved from: <http://thejournal.com/articles/2014/05/14/adaptive-learning-are-we-there-yet.aspx>
- Wen, M., Yang, D., & Rose, C. (2014). *Sentiment analysis in MOOC discussion forums: What does it tell us?* (pp. 130–137). Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), Indianapolis, IN.
- West, D. M. (2012). *Big data for education: Data mining, data analytics, and web dashboards*. Washington, DC: Brookings Institution.
- Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, 9, 229–237.



- 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.
- Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. In *Third Conference on Learning Analytics and Knowledge (LAK 2013)* (pp. 48–56). Leuven: ACM.
- Wixon, M., Arroyo, I., Muldner, K., Burleson, W., Rai, D., & Woolf, B. (2014). *The opportunities and limitations of scaling up sensor-free affect detection* (pp. 145–152). Proceedings of the 7th International Conference on Educational Data Mining (EDM 2014), Indianapolis IN.
- 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.
- Worsley, M., & Blikstein, P. (2014). *Deciphering the practices and affordances of different reasoning strategies through multimodal learning analytics*. Paper presented at the Third Multimodal Learning Analytics Workshop, Istanbul, Turkey.
- Xu, X., Murray, T., Woolf, B. P., & Smith, D. (2013). *Mining social deliberation in online communication: If you were me and I were you*. Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013), Memphis, TN.

### 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](mailto: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](mailto:kalantzi@illinois.edu)