Education 2.0: Artificial Intelligence and the End of the Test

Bill Cope and Mary Kalantzis
College of Education, University of Illinois
billcope@illinois.edu

1 Introduction: Getting Past Education 1.0

We are on the cusp of a series of socio-technical revolutions. On one count, after the industrial revolutions of steam, electricity, and digitization, the next is Industry 4.0, a revolution in which artificial intelligence will be central (Schwab & Klaus, 2017). On another count, focusing now on internet communications, after the read-only web, then the social web, this is Web 3.0, or what web founder Tim Berners-Lee calls the semantic web.1 In this proposal web data is structured around ontologies and tagged and structured in such a way that supplementary meaning is added to natural language, images and other media (Cope, Kalantzis & Magee, 2011).

Schools have barely been touched by these changes. Even though we now find computers in classrooms, and learners accessing their knowledge and doing their work on digital devices, the social relationships of learning have remained much the same. In this paper, we’re going to look at one critically important aspect of education, the test. We are going to focus on this particularly because the test the primary measure of educational outcomes, learner knowledge and progress, and teacher, school and system effectiveness. Tests also influence curriculum, the tail wagging the proverbial dog.

Students are now doing tests online—but as knowledge artifacts they have changed little; they are still mostly the standardized, select response, right-and-wrong answer tests invented over a century ago. They are still doing essays, assignments and projects that they hand in for the teacher to mark—now perhaps

1 https://www.bloomberg.com/news/articles/2007-04-09/q-and-a-with-tim-berners-lee-businessweek-business-news-stock-market-and-financial-advice.
Artificial Intelligence and the End of the Test

with an upload into a learning management system, but the processes for the evaluation of project work have not changed. If these are marked by automated essay assessment technologies (Burstein, & Chodorow, 2003; Shermis, 2014; Vojak et al., 2011; Cope et al., 2011) as is increasingly the case for high stakes supply response tests, it is to grade and rank in the same ways that human graders have done, rather than to offer meaningful feedback.

Meanwhile, students are still reading textbooks and listening to lectures—increasingly these are e-textbooks and video lectures—but they are still consumers of knowledge more than active creators of knowledge. And even as passive consumers, we don’t even have a clear idea of what they are reading and viewing, and what they are making of the knowledge they are encountering. The teacher is still a cog in a content transmission and testing system, with little scope to design or modify curriculum. Meanwhile, teachers and students are also still having classroom discussions—now perhaps in web discussion boards as well as orally—but these data remain ephemeral and not part of what is assessed.

So, although we have computers in schools, the key education artifacts are a century and older: the textbook, the lecture, and ephemeral classroom discussion. Then, when we want to know what learners have come to know, we ignore what they have done already and give them a test. The technologies have changed, and new opportunities are opened out. But as yet, we have mostly availed ourselves of them. As knowledge and learning ecologies, we haven’t even made much progress beyond Education 1.0, the education of the industrial revolution.

How might things be different? How might artificial intelligence be part of the change? What might be the shape of Education 2.0?

2 Artificial Intelligence: Defining a Pivot Point in Education

We want to argue in this paper that artificial intelligence will be a pivot point towards an Education 2.0. But first, we want to define what we mean by this phrase (Cope & Kalantzis, 2015c).

Perhaps the most famous measure of machine intelligence is the Turing Test in which a computer and a person is each hidden behind a screen, and another person is asking them both 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 artificial intelligence (Turing, 1950).
The response of language philosopher John Searle was to set up the Turing test in 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 computer’s answer is correct and in this sense is indistinguishable from the competent human, does not mean that it understands Chinese (Searle, 1980).

Rather than these sorts of test of mimicry and deception, we want to suggest a different definition of AI. Computers are cognitive prostheses of an entirely different order from human intelligence. They are incredibly smart because they can do things that it would not be practicable or even sensible for humans to do. These things are dumb to the extent that they are limited to memory retrieval and calculation. Data are converted to number followed programmatically by algorithmic deduction. Computers can retain large amounts of trivial data and quickly do huge numbers of calculations which no human in their right mind would attempt—so in this sense only, they are smarter than humans.

In other words, it is no virtue of a computer to be smart like a human. It is the computer’s virtue to be smart in a way that no human ever can be, or would ever want to be.

Here’s a case in point. In our research and development work at the University of Illinois we have developed an analytics tool called Scholar: VisiLearn to track and document student performance, ‘as-you-go’ and in three areas, knowledge (intellectual quality), focus (persistence), and help (collaboration). The visualization in Figure 1 is drawn from an analysis of the work of 87 students in our University of Illinois educational technologies class. Over an 8 week course the Analytics worked its way over 2,115,998 datapoints and offered 8,793 pieces of meaningful machine feedback and machine-supported human feedback. This visualization was never more than a few hours old, and every student had access to a visualization of their own progress towards course objectives.

Any teacher would be out of their mind to attempt any such thing. However, when the Analytics presents this information to the teacher, they gain insight into individual learners and the progress of the whole class that would have in the past been very hard to see. And for the learner, there is rich and detailed feedback that supports their learning as well as incremental progress data that tells them how well they are doing.

So computers do not in any helpful sense mimic human intelligence; at best they supplement human intelligence by doing things that humans never could or would or should do. Humans dig better with a shovel and a bulldozer (digging prostheses) than by hand; they form an embodied partnership with their digging prostheses. So, too, computers can serve as cognitive prostheses,
extensions of our thinking whose processes are little like our thinking but that we can use to supplement our thinking.

So, how might artificial intelligence defined in these terms take us in the direction of an Education 2.0? And specifically, how can it be applied to the challenge of learning what students have learned? How might it revolutionize assessment? How might it also transform curriculum and teaching?

3 What’s Wrong with Traditional Tests—Profoundly

Instinctively, learners know what is wrong with tests. But generation after generation, we have resigned ourselves to their inevitability. Here are the main problems, first in summary and in contrast with AI-supported, embedded assessments, followed by a more detailed analysis (table 1).
The measure of what we learn is long term memory. The traditional test checks what you can remember until the moment it is administered, and that you are free to forget the day after. This may have been appropriate to industrial-era society where information and tools of analysis were not readily at hand. But now these are readily available, in the cognitive prostheses that are ubiquitous, networked, digital devices.

The cognitive range measured in traditional tests is narrow. Remembering a fact or calculating a correct answer by correct application of a procedure are not only anachronistic cognitive skills (Dixon-Román, & Gergen, 2013). They are too narrow in fact for today’s world where the most valuable kinds of thinking have qualities that might be described as holistic, imaginative, emotionally sensitive, and a host of other such epistemic and productive virtues.

Traditional select response tests (e.g. multiple choice) in their nature throw up false positives and negatives. A false positive in such tests occurs in the case of

---

2 For a defense of this conception of learning, see: Kirschner, Paul A., John Sweller and Richard E. Clark. 2006. “Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching.” Educational Psychologist 41(2): 75–86.

---

| Traditional Tests                                      | AI-supported, Embedded Assessments |
|--------------------------------------------------------|-----------------------------------|
| Measure long-term memory                               | Assess higher-order thinking      |
| Address a narrow cognitive range: facts and procedures  | Address complex epistemic performance |
| A peculiar test logic, unlike other places of knowledge activity | Offer a broad range of data types and data points, authentic to knowledge work |
| Limited sampling                                       | Big Data: n=all                   |
| Disturbing experiences                                 | Embedded assessment is the learner’s friend |
| A linear process: backward looking and judgmental by nature | Recursive processes: prospective and constructive by nature |
| Individualized, isolating                              | Assess collaborative as well as individual intelligence |
| Insist on inequality                                   | Mastery learning, where every learner can succeed |

---

Table 1 Changing Assessment Paradigms
an answer you accidentally get right, even though you don’t understand the underlying principles, and a false negative when you get an answer wrong for a trivial reason. These data distortions are systematically built into select response assessments, because distractor items are designed to be nearly right (Haladyna, 2004). They are trick answers, right-looking answers to tempt you to give the wrong answer, and possibly for the right reasons, or reasons that make sense in terms of fuzzy logic. Conversely, if select response assessments are a game of trickery, you can play the game to get the right answer just by learning the tricks, such as the process of elimination where you successfully guess the right answer. In other words, false positives and negatives are endemic to the design of select response assessments. As knowledge artifacts, these are strange things, unparalleled elsewhere in learning and life—and so in a fundamental sense lack what in assessment theory is called “construct validity.”

Traditional tests are based on limited sampling and highly mediated inferences. How could a few hours at the end of a course be enough to sample what a learner has learned? Then there is a leap of inference, characterized by Pellegrino et al. as the assessment triangle: observation (the test that prompts student answers) <> interpretation (the test score) <> cognition (an inference about thinking based on the score) (Pellegrino, Chudowsky, & Glaser, 2001). This is a large leap from such a small sample, and as if something as complex, varied and multifaceted as cognition could be reduced to a number. This applies equally to the other canonical from of assessment, supply response assessments, or traditional essays.

Existentially, tests are disturbing experiences. Students mostly dread tests. What if there are unexpected questions, or if I have studied the wrong things? What if on the day, I can’t remember something? The dread arises not just when the stakes are high, but because they mostly are running blind, not knowing for
sure what will be in the test. Then, you don’t know how well you have been doing until it is too late to do anything about it. And you can’t learn from the test in a measurable way because it comes at the end. Tests are mostly summative (of learning: retrospective and judgmental in their orientation), and rarely formative (for learning: prospective and constructive in their focus). They are for the managers of educational systems more than they are for learners and their teachers (Ryan, & Shepard, 2008 & Linn, 2013).

Test logic is linear. Students learn the work then do the test, and after that if they pass, they can move on in a linear way to the next step in their learning or their life. There are no feedback loops—unless you have to repeat a course, and that is hardly a positive experience.

Test logic is isolating and individualized. Tests measure the memory and procedural capacities of individual brains. The social is excluded. No looking things up! No cheating! Knowledge, however is in its nature social, in workplaces for instance, and community life where we rely on readily accessible knowledge resources and the power of collaborations. This focus in tests on an individual’s thinking is unlike any other parts of knowledge and the intrinsically social environments in which knowledge is put to work (Cope & Kalantzis, 2015a).

Tests insist on inequality. Lastly, and perhaps the most egregious of the flaws of traditional tests, is that they insist on inequality. Children are placed into a Grade 3 literacy class because it is assumed they will all be able to learn to read and write at about that level. Then we want to insist on unequal outcomes.
and a defined point of measurement. Aspiring doctors have to get incredibly high scores to get into medical school. Then we insist on tests that differentiate them across a distribution curve. We insist that there must always be inequality, and in classical testing theory (e.g. item response theory (Mislevy, 2006)) we adjust our tests and their statistical calibrations in order to differentiate degrees of knowing. So here is a huge contradiction: to start by assuming everyone in a class is capable, then at the end to insist that only a few can be really smart, defined against the rest who are mediocre or dull. This culture of enforced inequality begins in Education 1.0 with intelligence testing, where Henry Goddard was by the 1920s able to differentiate across a statistical distribution people who were idiots, imbeciles, morons, average, above average, gifted, and genius (Goddard, 1920).

4 How Artificial Intelligence Opens Up a New Assessment Paradigm and Education 2.0

Let’s take each one of these eight characteristics of traditional tests and see how artificial intelligence can fundamentally change the assessment paradigm. This is an emerging area of educational research and development, called “learning analytics” (Behrens, & DiCerbo, 2013; Siemens, 2013; Siemens, Ryan, & Baker, 2013). With our colleagues in computer science at the University...
of Illinois, we have been developing and trialing these technologies in education settings from Grade 4 to higher education.\textsuperscript{3} Our aim has been to liberate learning from the shackles of traditional testing and to end the distinction between instruction and assessment—where no worthwhile instruction occurs without embedded feedback processes, and where there is no assessment that is not meaningful to learning. We have already created experimental versions of all the things we mention below.

*The measure of learning is higher order thinking.* This is an era in which we have wondrous cognitive prostheses. In our purses and in our pockets we have a massive encyclopedia elaborating on every significant fact, a map of the world with its every street, a calculator, and a myriad of other look-up and calculation apps. Instead of factual memory and correct application of procedures—we have ubiquitous computing machines to do that for us now—what we should be measuring is how well we use these memory-supporting and analysis-enhancing technologies. Today, the capacities we should be measuring are knowledge navigation and critical discernment that what distinguishes the true from the “fake” in available knowledge resources. The answers that are often matters of careful judgment and well informed perspective, and not simply, unequivocally “correct.” Some AI-supported assessment processes:

- Rubric-based peer-, self-, teacher- assessment of knowledge syntheses and objects (for instance projects, reports, designs), where the computer manages a complex peer-to-peer social interactions.\textsuperscript{4}

\textsuperscript{3} CGScholar research and development, supported by research grants from the US Department of Education, Institute of Education Sciences: “The Assess-as-You-Go Writing Assistant” (R305A090394); “Assessing Complex Performance” (R305B100008); “u-Learn.net: An Anywhere/Anytime Formative Assessment and Learning Feedback Environment” (EDIES-10-C-0018); “The Learning Element” (ED-IES-10-C-0021); and “InfoWriter: A Student Feedback and Formative Assessment Environment” (ED-IES-13-C-0099). Bill and Melinda Gates Foundation: “Scholar Literacy Courseware.” National Science Foundation: “Assessing ‘Complex Epistemic Performance’ in Online Learning Environments” (Award 1629561).

\textsuperscript{4} A selection of reports from our research group: Haniya, Samaa, Matthew Montebello, Bill Cope and Richard Tapping. 2018. “Promoting Critical Clinical Thinking through E-Learning,” in *Proceedings of the 10th International Conference on Education and New Learning Technologies* (*EduLearn18*). Palma de Mallorca ES. Montebello, Matthew, Bill Cope, Mary Kalantzis, Duane Searsmith, Tabassum Amina, Anastasia Olga Tzirides, Naichen Zhao, Min Chen and Samaa Haniya. 2018b. “Critical Thinking through a Reflexive Platform,” in *Proceedings of the 17th IEEE International Conference on Information Technology Based Higher Education and Training* (*ITEHT 2018*). Olhao PT. Montebello, Matthew, Petrilson Pinheiro, Bill Cope, Mary Kalantzis, Tabassum Amina, Duane Searsmith and Dungyun Cao. 2018c. “The Impact of the Peer Review Process Evolution on Learner Performance in E-Learning Environments,” in *Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S 2018)*. London
Machine feedback on the quality of feedback, comparing rubric criterion to response, and training data where previous reviewees have rated review quality.

The cognitive range that we want to measuring today is broad and deep: complex epistemic performance. We might want to measure critical, creative and design thinking (Ennis, 1996). We might want to measure the complex epistemic performance that underlies disciplinary practice: computational, scientific, clinical, or historical and other knowledge tradition or methodology. Or we might want to assess deep epistemological repertoires: thinking that is evidentiary/empirical, conceptual/theoretical, analytical/critical, and applied/creative (Cope & Kalantzis, 2015b). Some AI-supported assessment processes:

- Crowdsourcing of criterion-referenced peer assessment of that pushes learners in the direction of disciplinary reflection and metacognition. Cope, Kalantzis, Abd-El-Khalick, & Bagley (2013)
- Coded annotations, supported by machine learning where users train the system to recognize higher order thinking.\(^5\)
- Ontology-referenced maps that prompt knowledge creators and reviewers to add a second layer of meaning to text, image and data; this is direct support to learners, as well as machine learning training data. Olmanson et al. (2016)

We need to broaden the range of data types and data points for assessment. The dominance of select response assessments is based on the ease of their mechanization (Kalantzis & Cope, 2012). It has for some time been easy and cheap to mark item-based tests with a computer, starting with the notorious

---

\(^5\) A sample of technical explorations by members if our research group: Santu, Shubhra Kanti Karmaker, Chase Geigle, Duncan Ferguson, William Cope, Mary Kalantzis, Duane Searsmith and Chengxiang Zhai. 2018 (in review). “Sofsat: Towards a Set-Like Operator Based Framework for Semantic Analysis of Text.” Paper presented at the SIGKDD Explorations. Kuzi, Saar, William Cope, Duncan Ferguson, Chase Geigle and ChengXiang Zhai. 2018 (in review). “Automatic Assessment of Complex Assignments Using Topic Models.” Geigle, Chase. 2018. “Towards High Quality, Scalable Education: Techniques in Automated Assessment and Probabilistic Behavior Modeling.” Ph.D., Department of Computer Science, University of Illinois, Urbana IL.
“bubble tests.” Today, supply response tests (e.g. essays, short textual answers) can also be graded by computers easily and cheaply, but the purpose is the same, to judge students with grades. However, these two assessment technologies could be pushed in a more helpful direction for teachers and learners. Some AI-supported assessment processes:
- Select response assessments and quizzes that give students a second chance to answer, with an explanation.
- Computer adaptive and diagnostic select response tests that recalibrate to learner knowledge and offer specific, actionable feedback on areas of strength and weakness. Chang (2015)

Changing the focus of sampling to big data: n=all. When students are working in computer mediated environments—reading text, watching videos, engaging in classroom discussions, writing and offering peer reviews on projects, and reviewing the reviews, we are able to assess everything they do. Here is the paradox: assessment is now everywhere, so by comparison the limited sampling of tests becomes quite inadequate. Moreover, all assessment is formative (constructive, actionable feedback), and summative assessment is no more than a retrospective view of the learning territory that has been covered as evidenced in formative assessment data. Some AI-supported assessment processes:
- “Big data” analytics, where the size of the data is related to the scope of data collection and the granularity of datapoints. Cope, & Kalantzis (2016)

Embedded assessment is the learner’s friend. Machine, peer and teacher formative assessments come at a time when they can be helpful to learners. Progress data can tell students what they have achieved in a course or unit of work, and what they still need to do to meet curriculum and teacher objectives. Some AI-supported assessment processes:
- Developing a culture of mutual help with peer and machine offering feedback at semantically legible datapoints—i.e. every assessment datapoint can make manifest sense to the student.
- Overall progress visualizations: clear learning objectives, transparent progress data.

Assessment logic is recursive. This means that learning is characterized by feedback loops where a learner can act on feedback, seek further feedback, and act on it again, to the extent that is necessary for their learning. Some AI-supported assessment processes:
- Incremental feedback and data transparency allow a student to keep working until they meet a detailed learning objective and overall course objectives.

Intelligence is collaborative. Cheating only happens when learning is measured as isolated memory recall and correct answers using procedures. When
knowledge is acknowledged to be collaborative, the collaborations can be recorded and included in the assessment process. Students learn by giving feedback as much as by receiving it. In fact giving feedback against the criteria of a rubric prompts students to think in disciplinary and metacognitive terms. These social sources of feedback, moreover, are multifaceted (different kinds of datapoint), and multi-perspectival (peer, teacher, self, machine). Some AI-supported assessment processes:

- Measuring individual contributions to collaborative work in shared digital spaces. Montebello et al. (2018a)
- Rating the helpfulness of feedback, using reputation measurement methods now ubiquitous on the web.
- Machine moderation of peer ratings, recalibration for inter-rater reliability.

*Every student can succeed!* Half a century ago, Benjamin Bloom conceived the notion of mastery learning, or the notion that every student in a given class can achieve mastery, perhaps with additional time and support (Bloom, 1968). Today’s computer-mediated learning environments can achieve this, albeit by mechanisms that Bloom could never have imagined. These processes are personalized to the extent that assessment is not at a fixed moment in time, but a record of progress towards mastery which may take some students longer than others. The key is data transparency for learners and teachers. For the teacher: here is a data visualization showing that a particular student needs additional support. For the learner: here is a data visualization that shows what you have done so far in your journey to achieve mastery as defined by the teacher or the curriculum, and this precisely, is what you still need to do to achieve mastery. Some AI-supported assessment processes:

- Data transparency for students: clear learning objectives and incremental progress visualizations showing towards those objectives.
- Data transparency for teachers: class progress visualizations, showing effectiveness of instruction, just-in-time data identifying students who need additional support.

And How an Artificial Intelligence Assessment Paradigm Opens up a New Pedagogical Paradigm

Education 1.0 was focused on individual intelligence, memorization of facts, and correct application of procedures. The teacher and the textbook were at the center of a knowledge transmission process in a hub-and-spoke model. In part, these characteristics were determined by the available knowledge measurement technologies: traditional tests.
In this paper, we have outlined the ways in which artificial intelligence might be able to assess knowledge in new ways, with embedded assessments that are always contributing in an incremental way to the learning process. Part of this involves managing a great deal of social complexity that was not possible in the communications architecture of the traditional classroom—managing the peer review process for instance, or tracking contributions to online classroom discussions. These coordinating functions are managed by artificial intelligence in our narrow definition. So is the incremental collection of large amounts of granular data, the range of data types for collection and analysis, and the presentation of these data to learners and teachers in the form of visual syntheses. The scope of this collection and analysis would not be feasible for teachers without computer support. Finally, machine learning processes can, with human training, begin make sense of these data patterns. With these tools, we might be able to say that we have arrived at a new kind of education, Education 2.0, where the emphases are on recursive feedback, higher order thinking, and collaborative intelligence (Cope, & Kalantzis, 2017).

Testing is a big part of education, and it is not just a cliché to say that testing drives the system. If we change the tests, we might be able to change the system.
Table 2 Changing Educational Paradigms

| Education 1.0                                      | Education 2.0                                                                 |
|---------------------------------------------------|------------------------------------------------------------------------------|
| Teacher-centered                                   | Learner as agent, participant                                                 |
| Learner as knowledge consumer                     | Learner as knowledge producer                                                |
| Knowledge transmission and replication             | Knowledge as discoverable, navigation, critical discernment                   |
| Long term memory                                   | Devices as “cognitive prostheses”—social memory and immediate calculation    |
| Knowledge as fact, correctly executable theorem, definition | Knowledge as judgment, argumentation, reasoning                               |
| Cognitive focus                                    | Focus on knowledge representations, “works” (ergative)                        |
| Individual minds                                   | Social, dialogical minds                                                      |
| Long cycle feedback, retrospective and judgmental (summative assessment) | Short cycle feedback, prospective and constructive (reflexivity, recursive feedback, formative assessment) |

References

Behrens, John T. and Kristen E. DiCerbo. (2013). “Technological Implications for Assessment Ecosystems.” Pp. 101–22 in The Gordon Commission on the Future of Assessment in Education: Technical Report, edited by E.W. Gordon. Princeton NJ: The Gordon Commission.

Block, James H., ed. (1971). Mastery Learning: Theory and Practice. New York: Holt Rinehart & Winston.

Bloom, Benjamin S. (1968). “Learning for Mastery.” Evaluation Comment 1(2):1–2.

Burstein, Jill and Martin Chodorow. (2003). “Directions in Automated Essay Scoring.” in Handbook of Applied Linguistics, edited by R. Kaplan. New York: Oxford University Press.

Chang, Hua-Hua. (2015). “Psychometrics Behind Computerized Adaptive Testing.” Psychometrika 80(1):1–20. doi: 10.1007/s11336-014-9401-5.

Cope, Bill and Mary Kalantzis. (2017). “Conceptualizing E-Learning.” Pp. 1–45 in E-Learning Ecologies, edited by B. Cope and M. Kalantzis. New York: Routledge.

Cope, Bill and Mary Kalantzis. (2016). “Big Data Comes to School: Implications for Learning, Assessment and Research.” AERA Open 2(2):1–19.
Cope, Bill and Mary Kalantzis. (2015a). “Assessment and Pedagogy in the Era of Machine-Mediated Learning.” Pp. 350–74 in Education as Social Construction: Contributions to Theory, Research, and Practice, edited by T. Dragonas, K.J. Gergen, S. McNamee and E. Tseliou. Chagrin Falls OH: Worldshare Books.

Cope, Bill and Mary Kalantzis. (2015b). “The Things You Do to Know: An Introduction to the Pedagogy of Multiliteracies.” Pp. 1–36 in A Pedagogy of Multiliteracies: Learning by Design, edited by B. Cope and M. Kalantzis. London: Palgrave.

Cope, Bill and Mary Kalantzis. (2015c). “Sources of Evidence-of-Learning: Learning and Assessment in the Era of Big Data.” Open Review of Educational Research 2(1): 194–217. doi: http://dx.doi.org/10.1080/23265507.2015.1074869.

Cope, Bill, Mary Kalantzis and Liam Magee. (2011). Towards a Semantic Web: Connecting Knowledge in Academic Research. Cambridge UK: Elsevier.

Cope, Bill, Mary Kalantzis, Sarah J. McCarthy, Colleen Vojak and Sonia Kline. Ibid. “Technology-Mediated Writing Assessments: Paradigms and Principles.” 79–96.

Cope, Bill, Mary Kalantzis, Fouad Abd-El-Khalick and Elizabeth Bagley. (2013). “Science in Writing: Learning Scientific Argument in Principle and Practice.” e-Learning and Digital Media 10(4):420–41.

Dixon-Román, Ezekiel J. and Kenneth J. Gergen. (2013). “Epistemology in Measurement: Paradigms and Practices.” Vol. Princeton NJ: The Gordon Commission.

Ennis, Robert H. (1996). Critical Thinking. Upper Saddle River NJ: Prentice Hall.

Goddard, Henry H. (1920). Human Efficiency and Levels of Intelligence. Princeton NJ: Princeton University Press.

Haladyna, Thomas M. (2004). Developing and Validating Multiple-Choice Test Items. Mahwah NJ: Lawrence Erlbaum Associates.

Kalantzis, Mary and Bill Cope. (2012). New Learning: Elements of a Science of Education (Edn 2). Cambridge UK: Cambridge University Press. Chapter 10.

McMichael, Maureen A., Matthew C. Allender, Duncan Ferguson, Bill Cope, Mary Kalantzis, Matthew Montebello and Duane Searsmith. (2018). “Use of a Novel Learning Management System for Teaching Critical Thinking to First Year Veterinary Students.” In Preparation.

Mislevy, Robert J. (2006). “Cognitive Psychology and Educational Assessment.” Pp. 257–305 in Educational Measurement, edited by R.L. Brennan. New York: Praeger.

Montebello, Matthew, Bill Cope, Mary Kalantzis, Duane Searsmith, Tabassum Amina, Anastasia Olga Tzirides, Naichen Zhao, Min Chen and Samaa Haniya. (2018a). “Deepening E-Learning through Social Collaborative Intelligence.” in Proceedings of the 48th IEEE Annual Frontiers in Education (FIE) Conference. San Jose CA.

Montebello, Matthew, Petrilson Pinheiro, Bill Cope, Mary Kalantzis, Tabassum Amina, Duane Searsmith and Dungyun Cao. (2018c). “The Impact of the Peer Review Process Evolution on Learner Performance in E-Learning Environments.” in Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S 2018). London UK.
Olmanson, Justin, Katrina Kennett, Sarah J. McCarthey, Duane Searsmith, Bill Cope and Mary Kalantzis. (2016). “Visualizing Revision: Leveraging Student-Generated between-Draft Diagramming Data in Support of Academic Writing Development.” Technology, Knowledge and Learning 21(1):99–123.

Pellegrino, James W., Naomi Chudowsky and Robert Glaser, eds. (2001). Knowing What Students Know: The Science and Design of Educational Assessment. Washington DC: National Academies Press.

Piety, Phillip J. (2013). Assessing the Big Data Movement. New York: Teachers College Press.

Pinheiro, Petrilson. (2018). “Text Revision Practices in an E-Learning Environment: Fostering the Learning by Design Perspective.” Innovation in Language Learning and Teaching. doi: https://doi.org/10.1080/17501229.2018.1482902.

Ryan, Katherine E. and Lorrie A. Shepard, eds. (2008). The Future of Test-Based Accountability. New York: Routledge. Linn, Robert L. 2013. “Test-Based Accountability.” Vol. Princeton NJ: The Gordon Commission.

Schwab, Klaus. (2017). The Fourth Industrial Revolution: Currency. https://www.bloomberg.com/news/articles/2007-04-09/q-and-a-with-tim-berners-lee-businessweek-business-news-stock-market-and-financial-advice.

Searle, John R. (1980). “Minds, Brains, and Programs.” Behavioral and Brain Sciences (3):417–57.

Siemens, George. (2013). “Learning Analytics: The Emergence of a Discipline.” American Behavioral Scientist 57(10):1380–400. doi: 10.1177/0002764213498851.

Siemens, George and Ryan S J.d. Baker. (2013). “Learning Analytics and Educational Data Mining: Towards Communication and Collaboration.” Pp. 252–54 in Second Conference on Learning Analytics and Knowledge (lak 2012). Vancouver BC: ACM.

Shermis, Mark D. (2014). “State-of-the-Art Automated Essay Scoring: Competition, Results, and Future Directions from a United States Demonstration.” Assessing Writing 20:53–76. doi: 10.1016/j.asw.2013.04.001.

Turing, Alan M. (1950). “Computing Machinery and Intelligence.” Mind 59:433–60.

Vojak, Colleen, Sonia Kline, Bill Cope, Sarah J. McCarthey and Mary Kalantzis. (2011). “New Spaces and Old Places: An Analysis of Writing Assessment Software.” Computers and Composition 28(2):97–111.